

**Ministère de l'Agriculture**

**Ecole Nationale Supérieure Agronomique de Montpellier**

**THESE**

**Présentée pour obtenir le grade de**

**Docteur en Sciences de l'Ecole Nationale Supérieure Agronomique de Montpellier**

**Ecole doctorale : Biologie, Diversité et Adaptation des Plantes Cultivées**

**Option : Amélioration des Plantes**

**DÉTERMINISME GÉNÉTIQUE DE QUELQUES COMPOSÉS BIOCHIMIQUES DE LA  
GRAINE DE CAFÉ VERT IMPLIQUÉS DANS LA QUALITÉ À LA TASSE.**

**ETUDE D'UN CROISEMENT INTERSPECIFIQUE ENTRE  
*COFFEA PSEUDOZANGUEBARIAE* ET *C. LIBERICA* VAR. *DEWEVREI***

par

**Chin-Long KY**

Soutenance prévue le 7 avril 2000 devant le jury d'examen composé de :

M. CHARCOSSET A.	Directeur de Recherche à l'INRA	Rapporteur
M. JAY M.	Professeur à l'Université de Lyon I	Rapporteur
M. CHARRIER A.	Professeur à l'ENSAM	Examineur
M. GUYOT B.	Ingénieur de Recherche au CIRAD	Examineur
Industriel café		Examineur
M. HAMON S.	Directeur de Recherche à l'IRD	Examineur
M. NOIROT M.	Directeur de Recherche à l'IRD	Directeur de Thèse

# Sommaire

<b>CHAPITRE 1 : INTRODUCTION .....</b>	<b>1</b>
<b>CHAPITRE 2 : ETUDE BIBLIOGRAPHIQUE .....</b>	<b>3</b>
2. 1. LES CAFÉIERS.....	3
2. 1. 1. <i>Taxonomie</i> .....	3
2. 1. 2. <i>Morphologie et mode de reproduction</i> .....	4
2. 1. 3. <i>Structure du génome des caféiers</i> .....	6
2. 1. 4. <i>Amélioration de C. canephora</i> .....	7
2. 1. 5. <i>En résumé</i> .....	9
2. 2. LES COMPOSÉS BIOCHIMIQUES DES GRAINS DE CAFÉ.....	10
2. 2. 1. <i>La composition biochimique des grains verts de café</i> .....	10
2. 2. 2. <i>Les principaux constituants biochimiques de l'arôme et du goût du café</i> .....	14
2. 2. 3. <i>Biosynthèse et génétique des composés biochimiques étudiés</i> .....	19
2. 2. 4. <i>En résumé</i> .....	22
2. 3. CARTOGRAPHIE GÉNÉTIQUE ET RECHERCHE DE QTLs .....	23
2. 3. 1. <i>Cartographie génétique</i> .....	23
2. 3. 2. <i>Recherche de QTLs</i> .....	27
2. 3. 4. <i>En résumé</i> .....	29
<b>CHAPITRE 3 : MATÉRIEL ET MÉTHODES .....</b>	<b>31</b>
3. 1. MATÉRIEL VÉGÉTAL .....	31
3. 1. 1. <i>Les espèces cultivées</i> .....	31
3. 1. 2. <i>Le croisement interspécifique : PSE x DEW</i> .....	32
3. 2. LES MÉTHODES D'ACQUISITION DES DONNÉES .....	34
3. 2. 1. <i>Extraction, séparation et analyse des composés biochimiques</i> .....	38
3. 2. 2. <i>Extraction et dosage de l'ADN</i> .....	40
3. 2. 3. <i>Technique de marquage par AFLP</i> .....	43
3. 3. LES MÉTHODES D'ANALYSE DES DONNÉES .....	43
3. 3. 1. <i>Etude de la diversité</i> .....	43
3. 3. 2. <i>Approche quantitative du déterminisme génétique</i> .....	44
3. 3. 3. <i>Cartographie de la descendance BCDEW</i> .....	45
3. 3. 4. <i>Recherche de QTLs</i> .....	48

6. 2. 1 Carte de liaison génétique.....	76
6. 2. 2. Distorsions de ségrégation.....	77
6. 3. EN RÉSUMÉ.....	79
<b>CHAPITRE 7 : HÉRÉDITÉ DES CARACTÈRES D'INTÉRÊT : IDENTIFICATION DE QTLs .....</b>	<b>80</b>
7. 1. RÉSULTATS.....	81
7. 1. 1. Présentation de l'ensemble des QTLs détectés.....	81
7. 1. 2. Interprétation pour les caractères présentant plusieurs QTLs.....	83
7. 1. 3. Identification des QTLs expliquant les variations de teneur en caféine.....	86
7. 2. DISCUSSION.....	88
7. 2. 1. Rappel des principaux résultats.....	89
7. 2. 2. Les limites de l'identification des QTLs.....	89
7. 2. 3. La précision sur les effets des QTLs.....	93
7. 2. 4. Les QTLs identifiés pour les acides chlorogéniques et la caféine.....	93
7. 3. EN RÉSUMÉ.....	95
<b>CHAPITRE 8 : CONCLUSIONS ET PERSPECTIVES.....</b>	<b>96</b>
8. 1. CONCLUSIONS.....	96
8. 1. 1. Relations entre composés biochimiques.....	97
8. 1. 2. Déterminisme génétique des composés biochimiques.....	97
8. 1. 3. Cartographie génétique.....	97
8. 1. 4. Recherche de QTLs.....	98
8. 3. PERSPECTIVES.....	98
8. 3. 1. Evaluations biochimiques.....	99
8. 3. 2. Cartographie de différents croisements interspécifiques.....	99
8. 3. 3. Clonage et expression de gène d'intérêt.....	100
<b>CHAPITRE 9: RÉFÉRENCES BIBLIOGRAPHIQUES.....</b>	<b>101</b>
<b>ANNEXES</b>	

---

**Publications dans des revues scientifiques à comité de lecture (Rang A) :**

- Ky C.-L., Noirot M. & Hamon S.** (1997) Comparison of five purification methods for chlorogenic acids in green coffee beans (*Coffea* sp.).  
**Journal of Agricultural and Food Chemistry** 45 : 786-790.
- Ky C.-L., Louarn J., Guyot B., Charrier A., Hamon S. & Noirot M.** (1999) Relations between and inheritance of chlorogenic acid content in an interspecific cross between *Coffea pseudozanguebariae* and *Coffea liberica* var 'dewevrei'.  
**Theoretical and Applied Genetics** 98 : 628-637.
- Ky C.-L., Doulbeau S., Guyot B., Charrier A., Hamon S., Louarn J. & Noirot M.** (1999) Inheritance of sucrose content in the interspecific cross : *Coffea pseudozanguebariae* x *Coffea liberica* 'dewevrei'.  
**Plant Breeding** (sous-presse).
- Ky C.-L., Barre P., Lorieux M., Trouslot P., Akaffou S., Louarn J., Charrier A., Hamon S. & Noirot M.** (2000) Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.).  
Theoretical and Applied Genetics (accepté pour publication).

---

**Compte-rendu de colloques et communications orales :**

- Ky C-L, Hamon S, Louarn J, Akaffou S, Guyot B, Chrestin H, Charrier A & Noirot M.** (1999) The use of genetic resources for coffee cup taste improvement.  
**SIBAC** (sous presse)
- Ky C.-L., Louarn J., Akaffou S., Chrestin H., Charrier A., Hamon S. & Noirot M** (1999) Qualitative improvement of coffee seeds by genetical and physiological approaches.  
**Proc. Coll. ASIC 18** (sous-presse).

---

**Posters :**

- Ky C.-L., Hamon S., Louarn J., Akaffou S., Guyot B., Chrestin H., Charrier A. & Noirot M.** (1999) The use of genetic resources for coffee cup taste improvement. **SIBAC**, mai, Londrina, Brésil.
- Ky C.-L., Barre P., Lorieux M., Trouslot P., Akaffou S., Louarn J., Charrier A., Hamon S., Noirot M.** (2000). **Plant and Animal Genome VIII**, Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.), San-Diego, CA, USA, 9-12 janvier.

## Liste des figures

**Figure 2.1.1.** : Répartition géographique des principales espèces du genre *Coffea* sous-genre *Coffea*.

**Figure 2.1.2.** : Coupes transversales présentant les différents types de fruits de l'espèce *Coffea liberica* var. *dewevrei*.

**Figure 2.2.1.** : Composition biochimique des grains verts de *Coffea arabica* (en % MS) (Smith 1985).

**Figure 2.2.2.** : Composition biochimique des grains verts de *Coffea canephora* (en % MS) (Smith 1985).

**Figure 2.2.3.** : Structures chimiques des principaux isomères d'acides chlorogéniques rencontrés

**Figure 2.2.4.** : Nombre de composés aromatiques identifiés dans l'arôme café (Flament, 1991).

**Figure 2.2.5.** : Chaîne de biosynthèse conduisant à la formation des acides chlorogéniques chez les caféiers.

**Figure 2.3.1.** : Origine biologique des distorsions de ségrégation observées chez les végétaux, en relation avec leur stade de développement. D'après Gadish et Zamir (1986).

**Figure 2.3.2.** : Valeur seuil du *LOD score* (ordonnée) en fonction de l'espacement moyen entre les marqueurs (abscisse) pour différentes tailles du génome (exprimées en nombre de chromosomes). D'après Lander et Botstein (1989).

**Figure 3.1.** : Schéma illustrant le principe de la technique AFLP.

**Figure 4.1.** : Relation entre les variables % CQA et teneur DQA chez *C. arabica*.

**Figure 4.2.** : Relation entre teneur en 4,5-DQA et teneur en 3,5-DQA chez *C. arabica*.

**Figure 4.3.** : Relation entre les variables %CGA1 et teneur en FQA chez *C. arabica*.

**Figure 4.4.** : Distribution bi-modale de la teneur en 3-CQA chez *C. canephora*.

**Figure 4.5.** : Relation entre les variables % 3-CQA et % 3,5-DQA chez *C. canephora*.

**Figure 4.6.** : Relation entre les variables % 5-FQA et % CGA1 chez *C. canephora*.

**Figure 4.7.** : Relation entre les variables % 5-CQA et % 4,5-DQA chez *C. canephora*.

**Figure 5.1.** : Régression linéaire illustrant l'additivité de la teneur en saccharose, excepté pour les hybrides BCPSE.

**Figure 5.2.** : Régression linéaire illustrant l'additivité de la teneur totale en acides chlorogéniques (CGA). \*\*\* : Très hautement significatif.

**Figure 5.3. a) à f)** : Relations entre les différents isomères au sein des 3 classes d'acide chlorogéniques.

**Figure 5.4. a) et b)** : Relations entre classes d'acides chlorogéniques.

**Figure 6.1.** : Autoradiogrammes illustrant la sélection de 6 marqueurs AFLP spécifiques de l'espèce *C. pseudozanguebariae*, absents des 7 génotypes de l'espèce *C. liberica* var. *dewevrei* et présents chez les 10 hybrides F1.

**Figure 6.2.** : Autoradiogramme montrant la ségrégation de 6 marqueurs AFLP chez 27 hybrides BCDEW (combinaison d'amorce *Eco* ACC / *Mse* CTT). Les flèches rouge localisent les bandes prises en considération pour la cartographie.

**Figure 6.3.** : carte de liaison génétique de la descendance BCDEW.

**Figure 6.4.** : Répartition des distances entre locus sur la carte génétique construite à partir de la descendance BCDEW.

**Figure 6.5.** : Distribution des 181 marqueurs cartographiés sur le descendance BCDEW en fonction des types de distorsion.

**Figure 6.6.** : Distorsions de ségrégation observées sur la carte génétique de la descendance BCDEW.

## Liste des tableaux

**Tableau 2.1.1.** : Classification des *Coffeae* d'après Leroy (1980) et Bridson (1987).

**Tableau 2.1.2.** : Moyennes des viabilités polliniques (en %) dans les combinaisons hybrides entre taxons Africains (Louarn, 1992).

**Tableau 2.2.1.** : Constituants biochimiques des grains verts impliqués dans la formation et notes organoleptiques associées à l'arôme café (Tressl, 1989 ; Flament, 1991 ; Shibamoto, 1991).

**Tableau 3.1.** : Origines spontanées et sub-spontanées éthiopiennes (36) et kenyanes (2) de l'espèce *Coffea arabica*.

**Tableau 3.2.** : Origines spontanées et sub-spontanées (38) de l'espèce *Coffea canephora* collectées dans cinq pays africains.

**Tableau 3.3.** : Schéma d'obtention et génotypes de la descendance de cartographie génétique.

**Tableau 3.4.** : Génotypes correspondant à l'étude de génétique quantitative (Chapitre 5) pour les acides chlorogéniques (CGA), le saccharose (SAC) et la trigonelline (TRI).

**Tableau 4.1.** : Moyennes et amplitudes (Min-Max) des variables biochimiques des accessions spontanées de *C. arabica* (exprimée en % MS).

**Tableau 4.2.** : Composantes principales de la diversité intraspécifique de *C. arabica* pour les acides chlorogéniques.

**Tableau 4.3.** : Moyennes et amplitudes (Min-Max) des variables biochimiques des accessions spontanées de *C. canephora* (en % MS).

**Tableau 4.4.** : Teneurs moyennes (exprimées en % MS) et résultats des ANOVA (valeurs de F et de p), des caractères biochimiques présentant des différences significatives entre les origines géographiques de *C. canephora*.

**Tableau 4.5.** : Teneurs moyennes (exprimées en % MS) et résultats (valeurs de F et de p) des ANOVA, des caractères biochimiques présentant des différences significatives entre les origines géographiques de *C. canephora*.

**Tableau 4.6.** : Teneurs moyennes (exprimées en % MS) et résultats (valeurs de F et de p) des ANOVA, des caractères biochimiques présentant des différences significatives entre les deux groupes CGA-1 et CGA-2 discriminés par leur teneur en 3-CQA de *C. canephora*.

**Tableau 4.7.** : Composantes principales de la diversité intraspécifique de *C. canephora* pour les acides chlorogéniques. Les coefficients de corrélation sont indiqués sans parenthèses (en gras : les corrélations significatives).

**Tableau 4.8.** : Différences interspécifiques entre les accessions spontanées *C. arabica* et *C. canephora*.

**Tableau 4.9.** : Comparaison des accessions spontanées *C. arabica* et *C. canephora* par le test de Kruskal & Wallis.

**Tableau 5.1.** : Variations des teneurs en saccharose (exprimée en % MS) chez 5 hybrides BCDEW et 2 hybrides BCPSE pour les années de récolte : 1995 et 1997.

**Tableau 5.2.** : Variations des teneurs en trigonelline (exprimée en % MS) chez 4 hybrides BCDEW et 3 hybrides BCPSE pour les deux années de récoltes : 1995 et 1997.

**Tableau 5.3.** : Analyses de variance, partiellement hiérarchisées, réalisées pour évaluer les différences entre deux années de récolte (effet fixe), pour 12 génotypes (effet aléatoire) et pour 4 arbres par génotype (effet aléatoire hiérarchisé dans les génotypes), pour les teneurs en acides chlorogéniques.

**Tableau 5.4.** : Teneur en saccharose (en % MS) chez les espèces parentales, les hybrides F1 et les hybrides backcross (BCDEW et BCPSE). Les différences inter-génotypes sont données par les résultats du test F (\*\*\*) : Très hautement significatif,  $p < 0,001$ .

## Liste des annexes

**Annexe 2.1.** : Principales caractéristiques agronomiques des caféiers (Le Pierrès, 1995 ; Louarn, 1992).

**Annexe 2.2.** : Chaîne de biosynthèse conduisant à la formation de la caféine.

**Annexe 3.1.** : Solutions courantes.

**Annexe 3.2.** : Origine des 7 géotypes PSE (Kenya) et des 13 géotypes DEW (République Centrafricaine) retenus pour l'étude de génétique quantitative (Chapitre 5).

**Annexe 3.3.** : Présentation des 55 variables utilisées pour la recherche des QTLs.

**Annexe 5.1.** : Profils d'élution du saccharose à partir d'extraits de grains verts à maturité des espèces *C. liberica* var. 'dewevrei' (DEW) et *C. pseudozanguebariae* (PSE).

**Annexe 5.2.** : Profils de co-élution de la trigonelline et de la caféine d'extraits de grains verts à maturité des espèces : *C. liberica* var. 'dewevrei' (DEW) et *C. pseudozanguebariae* (PSE) obtenus par HPLC.

**Annexe 5.3.** : Profils d'élution obtenus par HPLC des acides chlorogéniques d'extraits de grains verts à maturité de l'espèce *C. liberica* var. 'dewevrei' (DEW) et de l'espèce *C. pseudozanguebariae* (PSE)

**Annexe 5.4.** : Différences inter et intra-groupe (DEW, BCDEW, F1, BCPSE et PSE) révélées par les tests F et les contributions factorielles

**Annexe 5.5.** : Teneur en saccharose, en trigonelline et en acides chlorogéniques chez : 13 géotypes DEW et 7 géotypes PSE (% MS).

**Annexe 6.1.** : Nombre de marqueurs AFLP générés à partir de 30 combinaisons d'amorces différentes.

**Annexe 6.2.** : Locus (AFLP et RFLP) montrant des distorsions de ségrégation.

**Annexe 7.1.** : Principales caractéristiques des QTLs détectés pour un LOD score > 2,0 [localisation : groupe de liaison, position en cM à partir du premier marqueur du groupe, marqueurs encadrant le LOD maximum ; pourcentage de variance expliqué par chaque QTL (%) et par le modèle multi-QTL (% total)].

**Annexe A** : Ky C.-L., Noirot M. & Hamon S. (1997) Comparison of five purification methods for chlorogenic acids in green coffee beans (*Coffea* sp.). *Journal of Agricultural and Food Chemistry* 45 : 786-790.

**Annexe B** : Ky C.-L., Louarn J., Guyot B., Charrier A., Hamon S. & Noirot M. (1999) Relations between and inheritance of chlorogenic acid content in an interspecific cross between *Coffea pseudozanguebariae* and *Coffea liberica* var 'dewevrei'. *Theoretical and Applied Genetics* 98 : 628-637.

**Annexe C** : Ky C.-L., Doubeau S., Guyot B., Charrier A., Hamon S., Louarn J. & Noirot M. (1999) Inheritance of sucrose content in the interspecific cross : *Coffea pseudozanguebariae* x *Coffea liberica* 'dewevrei'. *Plant Breeding* (sous-presse).

**Annexe D** : Ky C.-L., Barre P., Lorieux M., Trouslot P., Akaffou S., Louarn J., Charrier A., Hamon S. & Noirot M. (2000) Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). *Theoretical and Applied Genetics* (accepté pour publication).

## Liste des abréviations et des sigles

**ACP** : analyse en composantes principales

**ADN** : acide désoxyribonucléique

**ADNc** : acide désoxyribonucléique complémentaire

**AFLP** : amplification fragment length polymorphism

**ARA** : *Coffea arabica*

**ARN** : acide ribonucléique

**ARNm** : acide ribonucléique messenger

**ATP** : adénosine triphosphate

**BC** : backcross

**BCDEW** : backcross 1 sur *C. liberica* var. *dewevrei*

**BCPSE** : backcross 1 sur *C. pseudozanguebariae*

**BET** : bromure d'éthidium

**CAN** : *Coffea canephora*

**CGA** : acides chlorogéniques

**CIRAD** : Centre de coopération Internationale en Recherche Agronomique pour le développement

**cM** : centimorgan

**CNRA** : Centre National de Recherche Agronomique

**CQA** : classe des acides caféylquiniques

**DEW** : *Coffea liberica* var. *dewevrei*

**dNTP** : désoxynucléotide triphosphate

**DQA** : classes des acides dicaféylquiniques

**F1** : hybrides de première génération

**FAO** : Food and Agricultural Organisation

**FQA** : classes des acide férulylquinique

**GF** : groupe de liaison simulé

**HPLC** : high performance liquid chromatography

**IPGRI** : International Plant Genetic Resources Institute

**IRD** : Institut de Recherche pour le Développement

**LOD** : logarithm of odds ratio

**MS** : matière sèche

**ORSTOM** : institut français de recherche scientifique pour le développement en coopération

**PCR** : polymerase chain reaction

**PSE** : *Coffea pseudozanguebariae*

**QTL** : quantitative trait loci

**r** : coefficient de corrélation

**R<sup>2</sup>** : coefficient de détermination

**RAPD** : random amplified polymorphic DNA

**RFLP** : restriction fragment length polymorphism

**TAE** : tris-acétate-EDTA voir annexe 3.1

**TB** : tris-borate-EDTA voir annexe 3.1

**TE** : tris-EDTA voir annexe 3.1

**UV** : ultra-violet

## **Chapitre 1 : Introduction**

---

## Chapitre 1 : Introduction

Environ six millions de tonnes de café sont produites annuellement dans plus de 70 pays de la zone inter-tropicale. Plus d'un milliard de tasses de café sont bues chaque jour par plusieurs centaines de millions de consommateurs des pays les plus industrialisés. Cette production est assurée par deux espèces de caféiers : l'une tétraploïde, *Coffea arabica*, et l'autre diploïde, *C. canephora* (nommée Robusta). La première espèce assure 68 % de la production et son café, plus aromatique et moins amer, possède deux à trois fois moins de caféine. Plus apprécié par les consommateurs, son prix d'achat est en moyenne deux fois plus élevé que celui du Robusta.

Ces deux espèces ont des exigences écologiques différentes : *C. arabica* est cultivée sur les plateaux tropicaux d'altitude (> 600 mètres) et les cinq premiers pays producteurs sont : le Brésil, la Colombie, le Mexique, l'Éthiopie et le Guatemala. *C. canephora* est cultivée dans les zones tropicales humides de basse altitude et les cinq grands pays producteurs sont : l'Indonésie, le Vietnam, le Brésil, l'Ouganda et la Côte d'Ivoire. En fait, c'est l'absence de conditions favorables à la culture de *C. arabica* qui détermine la caféiculture du Robusta. L'amélioration qualitative du goût à la tasse de ce dernier constitue donc un objectif de première importance pour les pays qui le produisent. Dans ce cadre, un programme de recherche est actuellement développé à l'IRD (Institut de recherche pour le développement) par l'unité de recherche « Génomique des caféiers ».

L'objectif de ce travail est l'étude du déterminisme génétique de l'accumulation des composés biochimiques majeurs impliqués dans la qualité du café. Ces composés biochimiques sont les acides chlorogéniques, la caféine, la trigonelline et le saccharose. Il s'agit ici : i) d'identifier leurs relations phénotypiques et ii) de localiser les *quantitative trait loci* (QTLs) impliqués dans leurs expressions. Le transfert des gènes d'intérêts, que ce soit par introgression ou par transformation génétique, nécessite de bonnes connaissances de la physiologie et de la génétique de la plante. La physiologie permet d'étudier les voies

métaboliques conduisant à la synthèse des composés biochimiques désirés (ou non) et de choisir entre suppression ou sur-expression d'une voie métabolique. La génétique permet d'évaluer le nombre de gènes impliqués et leur(s) localisation dans le génome. Dans le cas d'un gène, sa localisation par rapport au centromère, son indépendance ou sa liaison génétique avec d'autres gènes, sa présence dans une zone du génome sujette à des distorsions de ségrégation sont autant d'informations permettant de déterminer la faisabilité du transfert. Il s'agit donc ici d'un questionnement scientifique original, dans la mesure où le transfert des gènes et l'étude physio-génétique de leur régulation ne sont pas séparés : les gènes sont pris en compte en termes d'intégration dans le génome et ayant sa propre valeur adaptative.

Le but final étant d'améliorer la qualité à la tasse de *C. canephora* pour l'approcher de celle de *C. arabica*, le premier objectif de cette thèse a été d'évaluer la diversité biochimique des deux espèces cultivées et de les comparer. Le second objectif a été d'étudier le déterminisme génétique et les corrélations phénotypiques de ces composés dans un croisement entre deux espèces très différenciées pour ces caractères : *C. liberica dewevrei* (DEW) et *C. pseudozanguebariae* (PSE). Une première approche quantitative a été suivie par une recherche de QTL sur une carte génétique du croisement. Les distorsions de ségrégation observées dans ce croisement interspécifique ont aussi été analysées en relation avec les possibilités de transfert génétique.

Dans ce document, après une présentation bibliographique en trois volets et une description du matériel végétal et des méthodes utilisées, nous présenterons les résultats obtenus lors de ce travail. Nous aborderons la diversité biochimique des espèces cultivées (Chap. 4), puis la génétique quantitative des composés biochimiques (Chap. 5), la cartographie génétique du croisement PSE x DEW (Chap. 6) et la recherche de QTLs des caractères biochimiques, agronomiques et phénologiques (Chap. 7). Une conclusion générale, comportant des perspectives à ce travail, sera présentée dans la dernière partie (Chap. 8).

## **Chapitre 2 : Etude bibliographique**

---

## Chapitre 2 : Etude bibliographique

Cette étude bibliographique comporte trois parties : i) la biologie de la plante étudiée, le caféier ; ii) les composés biochimiques présents dans la graine et dans l'arôme du café ; iii) un rappel des méthodes de cartographie et de recherche de QTLs.

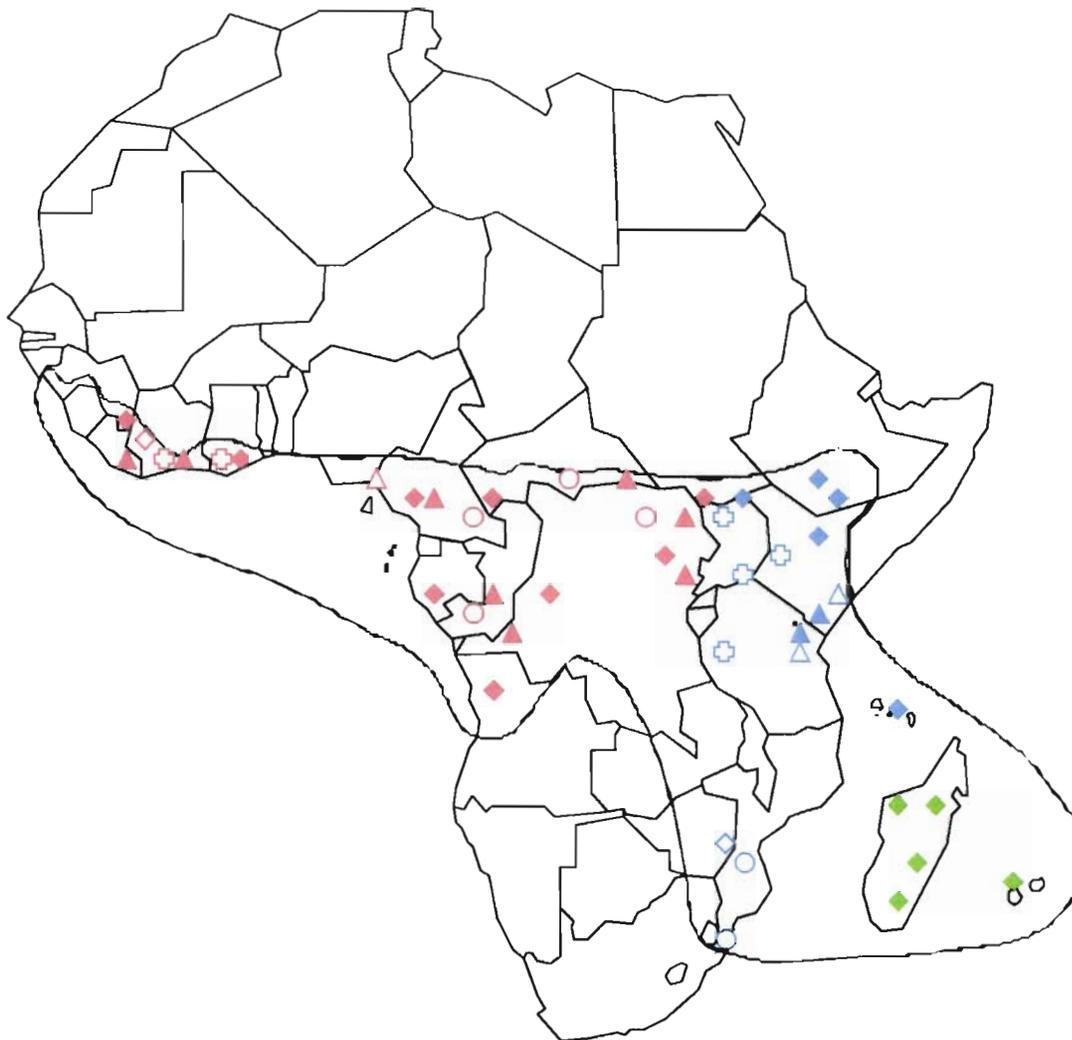
### 2. 1. Les caféiers

Cette partie présente la taxonomie, la morphologie, le mode de reproduction et la structure du génome propre aux caféiers. Nous nous intéresserons plus particulièrement à l'espèce cultivée de basse altitude *C. canephora* Pierre, que nous présenterons au travers des stratégies adoptées pour son amélioration.

#### 2. 1. 1. Taxonomie

Les caféiers sont des arbres ou arbustes endémiques des forêts intertropicales humides d'Afrique. Ils appartiennent à la famille des Rubiacées et sont caractérisés, entre autre, par la présence d'un sillon invaginé longitudinal sur la face ventrale des graines (placentation dite « cofféenne »). Les classifications taxonomiques les plus récentes reposent sur la description de la structure des inflorescences et répartissent les caféiers en deux genres : *Coffea* et *Psilanthus* (Leroy, 1980 ; Bridson et Verdcourt, 1988 ; Bridson 1994), chacun subdivisé en deux sous-genres (Tab. 2.1.1.).

Au sens strict, le terme de caféiers est réservé aux espèces du sous-genre *Coffea*. Ce sous-genre constitue un vaste complexe multispécifique regroupant plus de 80 taxons et dont l'aire de répartition naturelle s'étend de l'Afrique de l'ouest à Madagascar et aux îles Mascareignes (Fig. 2.1.1.). C'est d'ailleurs en fonction de cette répartition biogéographique que les caféiers ont été subdivisés en deux sections. La section « *Eucoffea* » recouvre les caféiers spontanés du continent africain. On distingue principalement : i) de l'ouest au centre de l'Afrique des espèces de basse altitude, dont *C. brevipes* Hiern, *C. canephora* Pierre,



OUEST & CENTRE	EST	MADAGASCAR
△ <i>C. brevipes</i>	◆ <i>C. arabica</i>	◆ > 50 taxons
◆ <i>C. canephora</i>	⊕ <i>C. eugenioides</i>	
○ <i>C. congensis</i>	▲ <i>C. pseudozanguebariae</i>	
◇ <i>C. humilis</i>	○ <i>C. racemosa</i>	
▲ <i>C. liberica</i>	◇ <i>C. salvatrix</i>	
⊕ <i>C. stenophylla</i>	△ <i>C. sessiliflora</i>	

Figure 2.1.1. : Répartition géographique des principales espèces du genre *Coffea* sous-genre *Coffea*.

*C. congensis* Froehner, *C. humilis* Chev., *C. liberica* Bull ex. Hiern, *C. stenophylla* Don. et ii) en Afrique de l'est des espèces adaptées à des altitudes variées, dont *C. arabica* Linné, *C. eugenoides* Moore, *C. pseudozanguebariae* Bridson, *C. racemosa* Lour., *C. salvatrix* Swynn., *C. sessiliflora* Bridson. La section des *Mascarocoffea* recouvre les caféiers spontanés de Madagascar, des îles Mascareignes et de l'archipel des Comores, où sont différenciés plus de 50 taxons.

Etant donnée la grande étendue du complexe multispécifique des caféiers, de nombreux travaux ont été entrepris afin de le structurer. Les études ont porté sur les enzymes (Berthaud *et al.*, 1989 ; Anthony, 1992), la diversité génétique des ADN chloroplastiques et ribosomiques (Cros, 1994 ; Cros *et al.*, 1997 ; Lashermes *et al.*, 1996a et c ; Lashermes *et al.*, 1997) sur quelques composés biochimiques (Rakotomalala, 1993) et les relations en croisement (Charrier, 1978 ; Louarn, 1992). Ce chapitre ne constitue pas une synthèse de ces études (que l'on pourra retrouver dans : Barre, 1997), mais résume les principales caractéristiques de ces plantes pérennes : morphologie, mode de reproduction, phénologie et structure du génome. Une récapitulation des espèces botaniques décrites est résumée en annexe 2.1.

**Tableau 2.1.1.** : Classification des *Coffeae* d'après Leroy (1980) et Bridson (1987)

Structure florale	Genre <i>Coffea</i> étamines et styles exerts	Genre <i>Psilanthus</i> étamines et styles inclus
Rameaux à croissance monopodiale Fleurs axillaires	Sous-genre <i>Coffea</i> (plus de 80 taxons)	Sous-genre <i>Psilanthus</i> (1 taxon)
Rameaux à croissance sympodiale Fleurs terminales	Sous-genre <i>Baracoffea</i> (7 taxons)	Sous-genre <i>Afrocoffea</i> (20 taxons)

## 2. 1. 2. Morphologie et mode de reproduction

Les caféiers sont des arbustes ou de petits arbres dont la hauteur varie de 1 à 2 mètres pour les espèces les moins développées (*C. humilis* et *C. heterocalyx*) à plus d'une vingtaine



**Photo 1** (Le Pierrès) : Floraison groupée de l'espèce *Coffea canephora* Pierre.



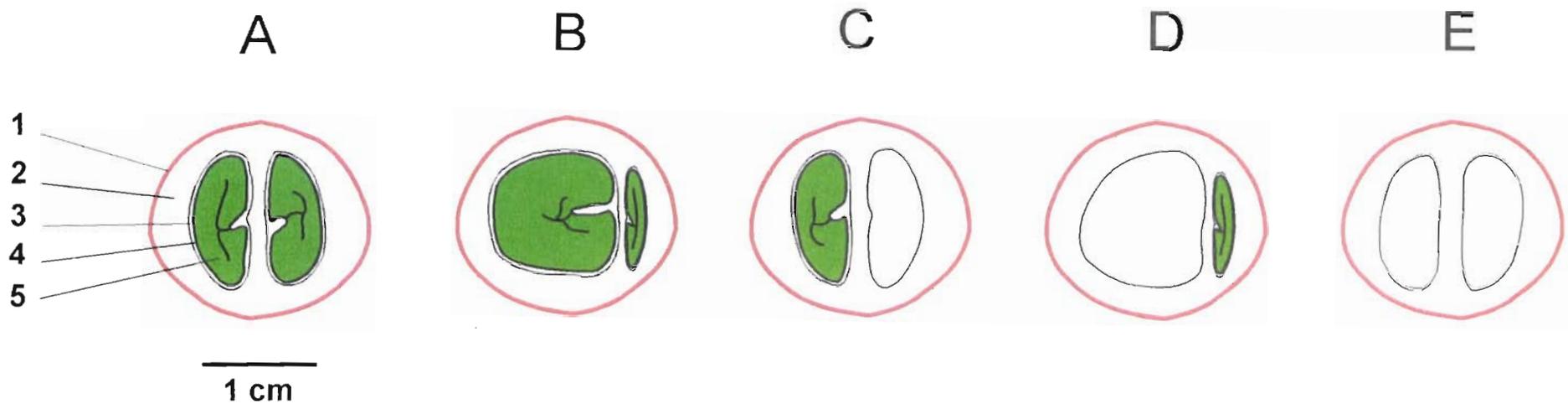
**Photo 2** (Charrier) : *Coffea canephora* portant simultanément des glomérules de fruits et de fleurs.

de mètres pour les plus grandes (*C. liberica*). L'arbre présente une architecture relativement régulière, possédant une tige unique, dressée (orthotrope) et régulièrement ramifiée sur tous ses nœuds. Les rameaux latéraux supra-axillaires restent obliques (plagiotropes plus ou moins ramifiées), relativement grêles et souples : ce sont les branches florifères. Les feuilles sont persistantes (exceptées en saison sèche pour *C. racemosa*) et montrent une grande diversité de forme (arrondies, allongées, lancéolées) et de taille (longueurs variant de 5-6 cm à plus de 20 cm) selon les espèces.

Les espèces et taxons de caféiers montrent une allogamie stricte (Krug et Carvalho, 1951). Le déterminisme génétique de cette auto-stérilité n'a été étudié que chez l'espèce *C. canephora*, où l'auto-incompatibilité est de type gamétophytique avec une série allélique de type S (Devreux *et al.*, 1959 ; Conagin et Mendes, 1981 ; Berthaud, 1986, Lashermes *et al.*, 1996b). Quelques espèces et taxons autogames font exception : *C. arabica*, *C. sp.* Moloundou (Anthony, 1992) et *C. heterocalyx* (Rakotomalala, 1993).

La différenciation des boutons floraux sur les nœuds des plagiotropes a lieu à la fin de la saison des pluies (d'août à octobre en Côte d'Ivoire). La pluviométrie doit être suffisante et bien répartie pour espérer un nombre important de fleurs l'année suivante. Ces initiations florales échelonnées n'évoluent pas durant la courte saison sèche (novembre à janvier). Ce sont les premières pluies de plus de 50 mm en février-mars qui conduisent à des floraisons groupées spectaculaires (Photo 1). Ces floraisons centrifuges (de la base vers les extrémités des plagiotropes) ont lieu entre le sixième (*C. racemosa*) et le dixième jour (*C. eugenioides*) après la pluie déclencheuse. Les inflorescences produites sont pluriflores (quatre fleurs et plus), pauciflores (autour de 2 à 3 fleurs) ou uniflores selon l'espèce. Les fleurs blanches et odorantes, dont les étamines et les stigmates sont exertes, se prêtent à une pollinisation à la fois anémophile et entomophile.

La durée du cycle de développement du fruit (durée du cycle de fructification) varie de onze semaines pour les espèces les plus précoces (*C. racemosa* et *C. sessiliflora*) à plus de dix mois pour les espèces les plus tardives (*C. liberica* et *C. congensis*). Le caféier porte alors fréquemment des glomérules de fruits et de fleurs (Photo 2). Les fruits, appelés drupes ou cerises, sont constitués d'un exocarpe coloré, d'un mésocarpe charnu (mucilage) et de deux graines accolées par leur face plane à invagination « cofféenne » (voir coupe transversale A, fig. 2.1.2.). Là aussi, une grande variété de forme, de taille et de couleur de fruit est observée parmi les espèces.



**Figure 2.1.2.** : Coupes transversales présentant les différents types de fruits de l'espèce *Coffea liberica* var. *dewevrei*. Selon la terminologie de De Reffye (1974), les fruits sont dits de type : (A) avec deux graines normales (2 loges pleines) ; (B) avec une graine arrondie appelée « caracoli » et une écaille (fruit n'ayant qu'un ovule fertile : 1 loge pleine) ; (C) avec une graine normale et une loge vide (2 loges, l'une pleine, l'autre vide) ; (D) avec une loge vide et une écaille et (E) avec deux loges vides. Les légendes correspondantes à la coupe transversale (A) sont : (1) exocarpe (pellicule rouge) ; (2) mésocarpe (mucilage) ; (3) endocarpe (parche) ; (4) spermoderme (pellicule argentée) et (5) albumen (fève). Les paramètres de fertilité sont : le coefficient de remplissage des fruits  $CREM = (2A + B + C) / (\text{Nb de fruits})$  ; le taux de fruits à une écaille  $FECA = (B + D) / (\text{Nb de fruits})$  ; le taux de loges vides  $TVID = (C + D + 2E) / [2(A + C + E) / B + D]$  et le taux de graines « caracoli »  $TCAR = B / (2A + B + C)$ .

Un défaut de pollinisation ou un avortement post-zygotique précoce conduit à l'existence de fruits à une graine : les caracolis, tandis qu'un avortement tardif de l'albumen mène à la présence d'une loge vide. La combinaison de ces événements explique les différentes catégories de fruits observées (Fig. 2.1.2.) (De Reffye, 1974).

L'organogenèse complète, depuis la formation du bois jusqu'à la maturité des fruits, couvre à peu près deux cycles phénologiques (environ deux années). Ceci explique l'alternance des bonnes et petites récoltes en combinaison avec les stress. En effet, les plagiotropes qui portent de nombreux glomérules de fruits ne peuvent produire, sauf fertilisation et irrigation copieuse, une longue pousse feuillée.

### 2. 1. 3. Structure du génome des caféiers

L'analyse des caryotypes sur pointes radicaires montre que tous les caféiers sont diploïdes avec  $2n = 2x = 22$  chromosomes, à l'exception de *C. arabica* (Bouharmont, 1959 ; Louarn, 1972) qui est amphidiploïde avec  $2n = 4x = 44$  chromosomes. Elle résulterait de l'hybridation de deux espèces progénitrices : *C. canephora* et *C. eugenioides* (Lashermes *et al.*, 1999).

Au niveau structural, les chromosomes sont de petite taille et difficilement identifiables en microscopie optique (Bouharmont, 1963). Cependant, quatre chromosomes se distinguent des autres par leur plus grande taille (Bouharmont, 1963).

Comme le montrent les relations génétiques en croisement, les espèces diploïdes ont un génome de base commun peu différencié (Carvalho et Monaco, 1967 ; Charrier, 1976 ; Louarn, 1992). Néanmoins deux grands groupes de caféiers diploïdes africains se distinguent : i) les caféiers d'Afrique centrale et occidentale et ii) les caféiers d'Afrique de l'est. La fertilité des hybrides F1 est relativement élevée à l'intérieur des groupes et faible entre groupes (Tab. 2.1.2.). Enfin, l'espèce *C. eugenioides* apparaît particulière avec un comportement en croisement similaire vis à vis des groupes précédents. Dans la quasi-totalité des cas, la stérilité des hybrides F1 est liée au nombre d'univalents présents dans les cellules-mères des grains de pollen, c'est-à-dire résultant de problèmes d'appariement chromosomique (Louarn, 1992).

La quantité d'ADN par noyau somatique varie de  $2C = 0,93$  à  $1,78$  pg chez les espèces diploïdes pour atteindre  $2,47$  pg chez l'espèce *C. arabica* (Cros *et al.*, 1995 ; Barre *et al.*, 1996). Ainsi, les plus petits génomes ont une taille comparable à celle du riz ( $2C = 1$  pg ; Bennett et Leitch, 1995). Mais surtout, cette variation de taille de génome du simple au double est à opposer à la constance du nombre chromosomique.

**Tableau 2.1.2.** : Moyennes des viabilités polliniques (en %) dans les combinaisons hybrides entre taxons Africains (Louarn, 1992). Les grains de pollen ont été colorés au carmin acétique (Louarn, 1976). Les codes pour les taxons sont pour l'Afrique de l'ouest : *C. canephora* (CAN), *C. congensis* (CON), *C. humilis* (HUM), *C. liberica dewevrei* (DEW), *C. liberica liberica* (LIB), *C. stenophylla* (STE) et pour l'Afrique de l'est : *C. eugenioides* (EUG), *C. pseudozanguebariae* (PSE), *C. racemosa* (RAC), *C. salvatrix* (SAL) et *C. sessiliflora* (SES).

Entre taxons Ouest Africain		Entre taxons Est Africain		Entre taxons Est et Ouest Africain	
CAN x CON	80	SES x RAC	75	SES x CAN	10
LIB x CON	68	PSE x SAL	72	SES x CON	11
CON x DEW	60	SES x PSE	56	SES x DEW	18
LIB x CAN	53	RAC x PSE	32	SES x STE	10
CAN x DEW	45	SES x EUG	19	RAC x CAN	< 5
LIB x DEW	44	PSE x EUG	16	RAC x CON	12
STE x DEW	42	EUG x SAL	14	RAC x DEW	16
LIB x STE	40			PSE x CAN	11
HUM x LIB	51			PSE x CON	8
HUM x STE	37			PSE x DEW	18
STE x CAN	38			PSE x LIB	28
HUM x CAN	32			EUG x CAN	12
				EUG x CON	15
				EUG x DEW	< 10
				EUG x LIB	< 5
				EUG x STE	< 5

Des travaux d'hybridation génomique *in situ* effectués sur les espèces *C. liberica dewevrei* et *C. pseudozanguebariae*, leurs hybrides F1 et leurs backcross 1 (1 cycle de rétrocroisement) réciproques ont montré que : i) les chromosomes des deux espèces sont facilement identifiables, ii) la quantité d'ADN des hybrides dépend du nombre de chromosomes de chacune des deux espèces présents dans son génome (Barre *et al.*, 1998b).

Une carte génétique a été développée sur une descendance d'environ 130 génotypes haploïdes-doublés de *C. canephora* à partir de 47 marqueurs RFLP et 100 marqueurs RAPD. Cette carte constituée de 15 groupes de liaison possède une taille de 1402 cM (Paillard *et al.*, 1996).

#### 2. 1. 4. Amélioration de *C. canephora*

*C. canephora* est vigoureuse et résistante (d'où son surnom de Robusta) à de nombreuses maladies, nématodes et insectes (Berthaud et Charrier, 1988 ; Carneiro, 1997 ; Cramer, 1957 ; Ferwerda, 1948). En revanche, ses grains contiennent deux à trois fois plus de caféine, son café est plus amer et l'arôme y est moins développé que chez *C. arabica*. Les principaux objectifs des programmes de recherche caféière ont été orientés vers l'amélioration du rendement, de la granulométrie, de l'adaptation aux conditions locales (adaptation à la sécheresse, ou au contraire aux zones inondables), mais aussi des qualités organoleptiques. Les schémas d'amélioration exploitent la diversité intraspécifique et/ ou la diversité interspécifique.

##### Amélioration intraspécifique

Jusque dans les années 1950, des sélections massales simples pour la productivité ont eu lieu dans les descendance du noyau fondateur, au Brésil (Paulino *et al.*, 1978), en Ouganda (Millot, 1974), en Afrique centrale (Capot, 1962) et en Afrique occidentale (Fressanges, 1954 ; Cordier, 1961). Le schéma de sélection reposait sur la sélection de génotype d'intérêt à partir de matériel local, de clones, de semences distribuées ou introduites (Ferwerda, 1969).

En 1958, l'Institut Français du Café et du Cacao (IFCC) a mis en place une stratégie d'amélioration commune à tous les pays francophones africains (Côte d'Ivoire, République Centrafricaine, Cameroun, Ouganda) et malgaches, producteurs de café issu de *C. canephora*. Le schéma de sélection a consisté à retenir dans les collections de caféiers locaux, ou introduits, des arbres à productivité élevée et forte granulométrie pour la multiplication végétative par bouturage (Braudeau *et al.*, 1962 ; Snoeck, 1968). Simultanément, en

opération-conjointe avec l'ORSTOM, des analyses génétiques de leur comportement en croisement (aptitudes générale et spécifique à la combinaison) ont été entreprises à Madagascar pour les caractères ci-dessus mentionnés.

L'existence d'un hétérosis chez les hybrides entre populations naturelles guinéennes et congolaises de *C. canephora* a permis de proposer un schéma de sélection récurrente réciproque (Comstock *et al.*, 1949 ; Berthaud, 1986 ; Leroy et Charrier, 1990 ; Leroy, 1993). Les résultats d'un premier cycle de croisement guinéen-congolais montrent un rendement de 2670 kg/ ha et une granulométrie de 14 g/ 100 graines, contre respectivement 2220 kg/ ha et 11 g/ 100 graines pour le clone 461 vulgarisé en Côte d'Ivoire (Leroy *et al.*, 1997). Ces croisements intraspécifiques guinéo-congolais constituent des croisements témoins pour les croisements interspécifiques.

Lorsque la diversité intraspécifique n'a pas permis l'amélioration génétique, la variabilité interspécifique a été prise en considération.

#### Amélioration interspécifique

De nombreux programmes d'amélioration génétique se sont intéressés aux hybrides interspécifiques avec *C. canephora* (Carvalho et Monaco, 1967 ; Charrier, 1976 ; Louarn, 1987 ; Yapo *et al.*, 1992). Au niveau diploïde, seuls les hybrides avec *C. congensis* (hybrides « Congusta ») sont normalement fertiles (Anthony et Le Pierrès, 1987 ; Berthaud et Charrier, 1988 ; Louarn, 1982 ; Yapo *et al.*, 1990) et ont été largement vulgarisés à Madagascar. Adaptées aux zones inondables, ils montrent une meilleure granulométrie et qualité à la tasse. Un programme d'introgession à partir de l'espèce *C. eugenioides* a commencé dans les années 70 à Madagascar (Louarn, 1976), mais n'a pas pu être mené à terme.

En fait, les deux plus importants programmes d'amélioration interspécifique de *C. canephora* ont concerné l'obtention d'hybrides F1, au niveau tétraploïde (Capot, 1972) et hexaploïde (Berthaud, 1977 ; Le Pierrès et Anthony, 1980 ; Le Pierrès, 1995) avec *C. arabica*. Ce sont les hybrides « Arabusta ». L'arbre recherché devait être adapté aux conditions de basse altitude et présenter des qualités organoleptiques et granulométriques intermédiaires. Au niveau tétraploïde, comme au niveau hexaploïde, les problèmes de fertilité liés à l'existence de tétravalents lors de la méiose (Grassias-Hubault, 1980 ; Le Pierrès, 1995) n'ont pas permis l'utilisation généralisée de ces formes d'hybrides dans la caféiculture.

### 2. 1. 5. En résumé

Toutes les espèces agronomiquement intéressantes sont réunies dans le genre *Coffea* L. sous-genre *Coffea*. Parmi les quatre-vingt taxons, seules *C. canephora* et *C. arabica* sont largement cultivées. Ces deux espèces ont fait l'objet de nombreux programmes de sélection depuis le début du siècle, en vue d'accroître la production (multipliée par cinq en un siècle) pour *C. canephora*, et la résistance aux problèmes phytosanitaires pour *C. arabica*. Les résultats sont encourageants, mais ont laissé les qualités gustatives à l'écart des programmes d'amélioration. Si le café produit par *C. arabica* constitue une référence gustative, il n'en est pas de même de celui produit par *C. canephora*.

Sur la base de ces nouveaux critères de sélection, l'amélioration s'oriente vers l'utilisation de la diversité des autres espèces de caféiers. Ceci nécessite une connaissance préalable du déterminisme génétique des caractères à introgresser et implique donc l'étude de grandes descendance en ségrégation. L'amélioration de cette plante pérenne est longue, en raison de la durée des cycles de sélection pour des arbres à observer pendant plusieurs années, et des surfaces agricoles à mettre en œuvre pour de tels programmes. Les espèces sauvages utilisées en croisement avec *C. canephora* et maintenues à la station IRD de Divo sont les suivantes :

- *C. pseudozanguebariae* pour l'absence de caféine des grains, le cycle court de fructification et l'adaptation à la sécheresse,
- *C. heterocalyx* pour l'autogamie,
- *C. eugenioides* pour sa parenté avec *C. arabica*,
- *C. liberica* pour sa fructification groupée et sa granulométrie.

## 2. 2. Les composés biochimiques des grains de café

Cette partie a pour objectif de donner un aperçu de la biochimie du café. Tout d'abord, une présentation des composés biochimiques des grains verts sera effectuée. Puis, un aperçu des composés caractéristiques de l'arôme café sera donné, en précisant les origines biochimiques de leurs formations et leurs notes organoleptiques. Après cette double présentation, un aperçu de la biosynthèse et de la génétique des composés retenus pour la suite de notre étude sera données.

### 2. 2. 1. La composition biochimique des grains verts de café

La plupart des composés biochimiques des grains verts sont à l'origine, après torréfaction, de l'arôme caractéristique du café ; c'est pourquoi on les dénomme souvent précurseurs d'arôme.

Les analyses les plus complètes se rapportent aux espèces cultivées : *C. arabica* et *C. canephora*. Les composés biochimiques rencontrés dans les grains verts sont par ordre d'importance (d'un point de vue quantitatif) : les glucides, les lipides, les protéines, les dérivés azotés non protéiques, les acides organiques et les substances minérales. Les figures 2.2.1. et 2.2.2. illustrent les données chiffrées relatives à ces deux espèces.

#### Les glucides

Les glucides représentent plus de 50 % de la matière sèche (% MS) des grains verts. La majeure partie de ces glucides est constituée de polysaccharides, dont la nature et la répartition exactes ne sont pas encore établies de façon définitive en raison des difficultés liées à leur séparation (Thaler et Arneht, 1967). Parmi les glucides de faible poids moléculaire (monosaccharides et oligosaccharides) (Trugo, 1985), le saccharose est le sucre libre majoritaire des grains verts (plus de 98% du total) (Clifford, 1985a). En moyenne, les grains verts de *C. arabica* contiennent deux fois plus de saccharose que ceux de *C. canephora*. La présence, à l'état de traces, d'autres sucres de faible poids moléculaire tels que le glucose, le fructose, le stachyose et le raffinose a aussi été observée (Kroplien 1971 ; Tressl *et al.*, 1982).

#### Les lipides

Les lipides représentent de 10 à 20 % de la matière sèche des grains verts de café. Ce sont des composés saponifiables (acides gras et dérivés) ou non (stérols, di et triterpènes, tocophérols) ou entrant dans la composition de la cire (Folstar, 1985). Parmi les composés

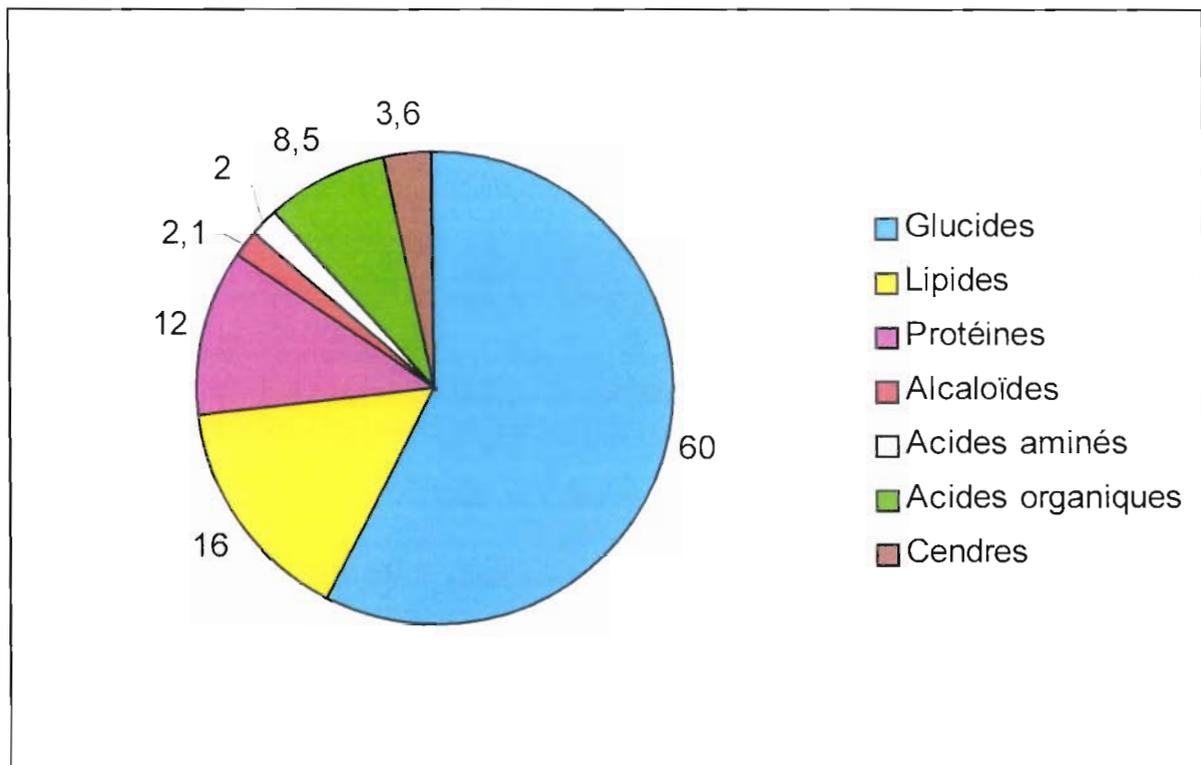


Figure 2.2.1. : Composition biochimique des grains verts de *Coffea arabica* (en % MS) (Smith 1985).

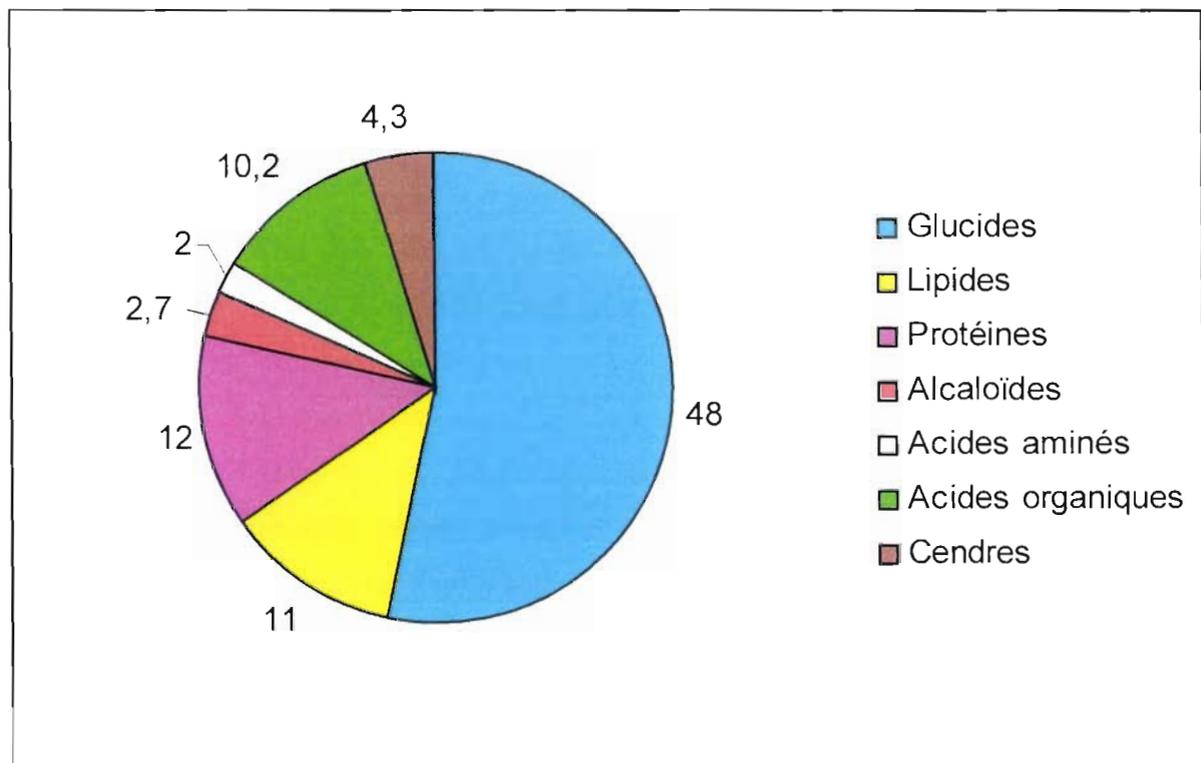


Figure 2.2.2. : Composition biochimique des grains verts de *Coffea canephora* (en % MS) (Smith 1985).

saponifiables, une grande proportion est formée de triacylglycérols. L'hydrolyse possible de leurs liaisons esters entre acides gras et glycérol par des acides ou des bases donnent des savons (sels de Na<sup>+</sup> ou K<sup>+</sup> des acides gras) (Speer *et al.*, 1993). A l'inverse, la deuxième catégorie, très minoritaire, comporte des composés à liaison insaponifiable. Elle comprend deux diterpènes majoritaires spécifiques du café, le cafestol et le kawhéal, largement étudiés pour leurs propriétés physiologiques particulières (De Ross *et al.*, 1997). D'autres hétérosides diterpéniques amers, caractéristiques des caféiers sauvages dépourvus de caféine, ont été découverts, comme le mascaroside de *C. vianneyi*, la camafarine de *C. buxifolia* et le mozambioside de *C. pseudozanguebariae* (Prewo *et al.*, 1990 ; Rakotomalala, 1993). Enfin, la fraction lipidique externe de la graine, extractible en traitant les grains non broyés au chloroforme, comporte essentiellement des cires, dont la composition reste très voisine de celle des lipides présents à l'intérieur des grains (Folstar *et al.*, 1977).

### Les protéines

Les teneurs en protéines des grains verts varient de 10 à 15 % MS. D'après Thaler (1975), un tiers des protéines des grains verts se trouve associé aux structures pariétales (notamment avec l'arabinogalactane). Comme dans tout matériel végétal, une multitude d'enzymes a été retrouvée. La plupart, non spécifiques des caféiers, possèdent des fonctions déjà bien connues. D'autres ont été particulièrement étudiées dans les grains verts de caféiers, comme l' $\alpha$ -galactosidase impliquée dans la germination (Shadaksharaswamy et Ramachandra, 1968), la malate déshydrogénase comme marqueur chimiotaxinomique (Payne *et al.*, 1973), les peroxydases comme indicateurs de la qualité des grains (Armorin *et al.*, 1977). La forte activité de la polyphénol oxydase est aussi connue comme bon indicateur de qualité (Armorin et Silva, 1968).

### Les dérivés azotés non protéiques

Les principaux dérivés azotés non protéiques sont : i) les alcaloïdes (caféine et trigonelline) et ii) les acides aminés libres.

### *Les alcaloïdes*

La caféine (triméthyl-1,3,7 xanthine) est l'alcaloïde majoritaire des grains verts de café. Composé très stable, il n'est pas dégradé lors de la torréfaction. Dans le café-boisson, il est connu pour être le stimulant principal du système nerveux central (Viani, 1985). A faible dose journalière (50 à 300 mg), ses effets physiologiques sont nombreux et se traduisent

essentiellement par une augmentation de l'attention, de l'acuité visuelle, des capacités de mémorisation, de l'activité locomotrice et des troubles du sommeil (insomnie). Quatre récepteurs de la caféine (antagoniste de l'adénosine) ont été identifiés. La fixation de la caféine sur l'un d'entre eux, le récepteur A1, entraîne une libération de la glutamine, qui est un neurotransmetteur excitateur du système nerveux. La caféine stimule aussi la libération de dopamine et d'acétylcholine (neurotransmetteurs excitateurs) dans le cortex préfrontal humain.

Elle confère aussi une résistance des plantes aux attaques d'éventuels pathogènes. Ainsi, son effet d'inhibiteur de croissance du champignon phytopathogène *Manocrosporium ambrosium* a été clairement démontrée chez les théiers (Savitri Kumar *et al.*, 1995).

La teneur en caféine dans le genre *Coffea* varie de 0 à 3,19 % MS (Anthony *et al.*, 1993). L'absence de caféine est une caractéristique assez générale des caféiers malgaches et exceptionnellement des caféiers est africains comme *C. pseudozanguebariae*. Les plus fortes teneurs en caféine ont été observées chez l'espèce *C. canephora* où la variation intraspécifique est comprise entre 1,2 à plus de 3,5 % MS (Charrier et Berthaud, 1975).

La trigonelline (1-méthylpyridine-3-carboxilate) est, après la caféine, l'alcaloïde le plus important chez les caféiers. A l'inverse de la caféine, la trigonelline est thermiquement instable et très peu stimulante. Lors de la torréfaction, la pyrolyse de la trigonelline conduit à la formation d'une vitamine, l'acide nicotinique (Czok, 1965 ; Stennert et Maier, 1994). Le café-boisson est une source importante en cette vitamine. Ainsi, la prise d'une tasse de café correspond à environ 20% de la ration journalière nécessaire à l'organisme humain.

Les teneurs concernant les deux espèces cultivées sont de l'ordre de 0,79 à 1,06 % MS chez l'espèce *C. arabica* et de 0,66 % à 0,68 % MS chez *C. canephora* (Stennert et Maier, 1993 et 1994). Des analyses sur d'autres espèces de caféiers ont eu lieu, mais pas de façon aussi extensive que pour la caféine. Chez *C. liberica*, la teneur en trigonelline varie de 1,0 à 1,4 % MS et chez un échantillon de *C. stenophylla* la valeur trouvée est de 1,89 % MS (Mazzafera, 1991).

Chez d'autres plantes, la trigonelline est par ailleurs un osmorégulateur du stress hydrique et un régulateur du cycle cellulaire dans les méristèmes racinaires (Evans et Tramontano, 1984).

### *Les acides aminés libres*

Ces derniers existent à l'état libre au côté des protéines. Certains d'entre eux constituent des marqueurs taxonomiques de certaines espèces ou variétés de *Coffea*, comme l'arginine, la phénylalanine et la tyrosine pour *C. canephora*, l'acide pipécolique pour *C. arabica* et le tryptophane pour *C. arabica* de Timor et du Cap vert (Correia *et al.*, 1994).

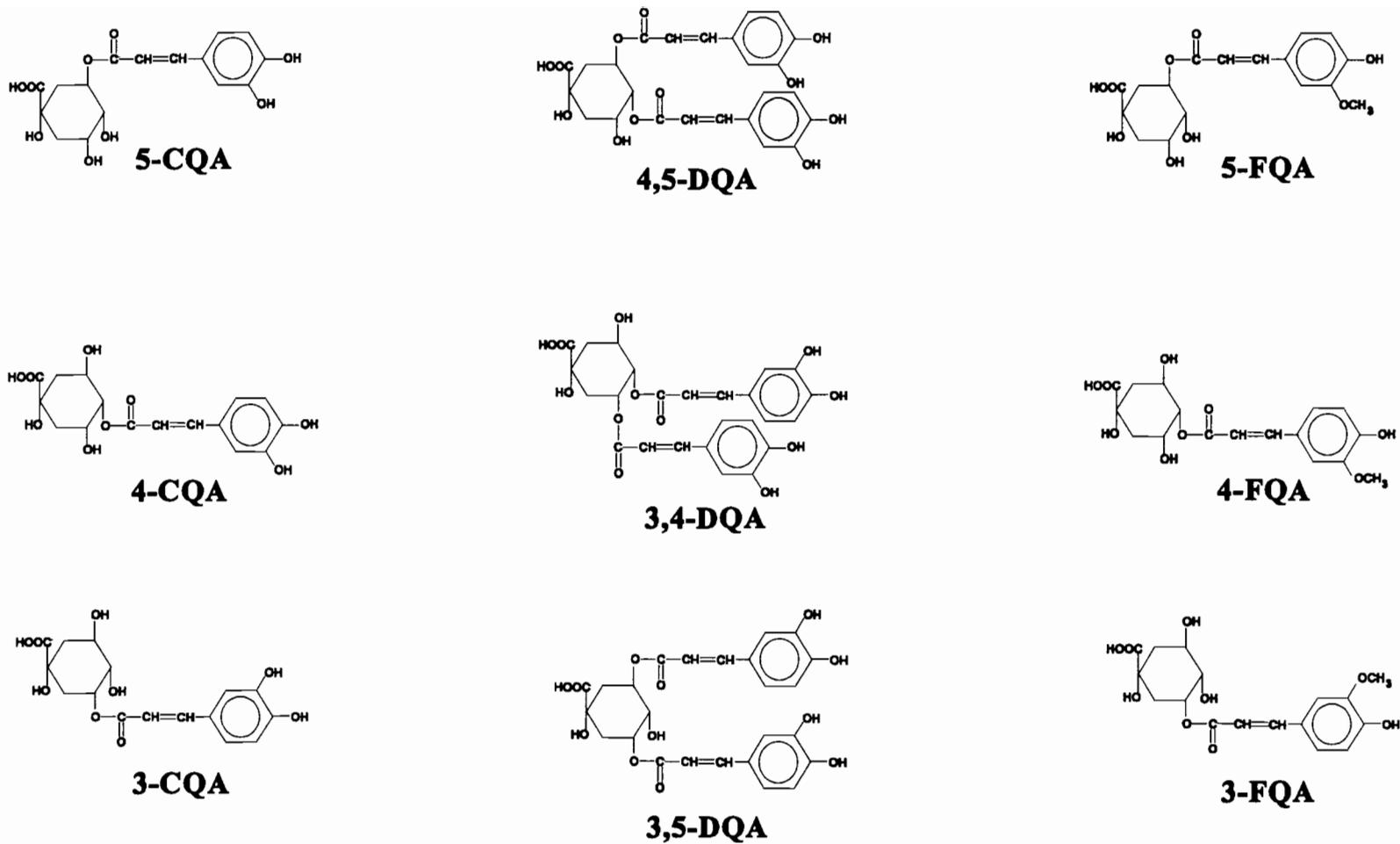
L'espèce *C. canephora* est plus riche en acides aminés libres que *C. arabica* (0,44% MS vs. 0,40% MS). Chez les deux espèces, les principaux acides aminés sont les acides glutamique et aspartique, l'alanine, la sérine. Elles diffèrent néanmoins par une teneur deux fois plus faible en acide glutamique et par une plus grande importance de la phénylalanine chez *C. canephora* (Tressl *et al.*, 1982 ; Arnold *et al.*, 1994 ; Arnold et Ludwig, 1996).

### Les acides organiques

Deux classes d'acides organiques sont rencontrées dans les grains verts de café : i) les acides aliphatiques formés au cours du métabolisme cellulaire (acide pyruvique, succinique, malique, lactique...) et ii) les acides phénols et dérivés (acides chlorogéniques essentiellement). Seule, cette dernière classe, particulièrement importante chez les caféiers, sera développée ici.

Les acides chlorogéniques (CGA) ou esters d'acides quiniques constituent les plus importants composés phénoliques du métabolisme secondaire dans le règne végétal. Les acides chlorogéniques sont des précurseurs de la lignine (Hahlbroch et Grisebach, 1979) . Ils participent au contrôle de la germination et de la croissance cellulaire. Ils sont impliqués également dans les mécanismes de défense contre les phytopathogènes (Echandi et Fernandez, 1962 ; Zuluaga *et al.*, 1971 ; Waites *et al.*, 1978 ; Macheix *et al.*, 1990). Les acides chlorogéniques sont particulièrement abondants chez les espèces cultivées de caféiers (Streuli, 1970 ; Armorin *et al.*, 1974 ; Maier, 1987), surtout au niveau des grains verts et de leurs dérivés : café torréfié, en poudre ou en boisson (Balyaya et Clifford, 1995 ; Clifford, 1985a). Ils ont été décrits pour la première fois par Robiquet et Boutron en 1837 (Sondheimer, 1964).

Chez les végétaux, ils sont communément subdivisés en six classes : i) les acides caféylquiniques (CQA), ii) les acides dicaféylquiniques (DQA), iii) les acides férulylquiniques (FQA), iv) les acides *p*-coumarylquiniques (*p*CoQA), v) les acides caféoylférulylquiniques (CFQA) et vi) les acides féruloyle caféylquiniques (FCQA). Chez les caféiers, les trois premières classes (CQA, DQA et FQA) représentent plus de 98 % de la



**Figure 2.2.3.** : Structures chimiques des principaux isomères d'acides chlorogéniques rencontrés chez les caféiers. Les classes d'acides chlorogéniques sont : les acides caféylquiniques (CQA), les acides dicaféylquiniques (DQA) et les acides férulylquiniques (FQA) (de gauche à droite). Chaque classe est composée de trois isomères (de haut en bas).

fraction totale en CGA (Clifford et Staniforth, 1977 ; Morishita et al., 1989). Chacune de ces classes est composée de trois isomères. Ces derniers diffèrent par le nombre et la nature du résidu acide hydroxycinnamique (acides caféiques ou acides féruliques) estérifiés sur les positions 3, 4 et/ ou 5 de l'acide quinique, comme le montre la figure 2.2.3. L'isomère acide caféyl 5-quinique (5-CQA) est le constituant majoritaire dans le café vert (Clifford, 1985b). Des composés voisins résultant d'une liaison entre l'unité acide caféique et un acide aminé comme la tyrosine ont été détecté à l'état de trace chez certaines variétés de culture (Clifford et al., 1989a ; Correia *et al.*, 1995).

La teneur en CGA varie considérablement d'une espèce à l'autre (Clifford et Jarvis, 1988 ; Clifford *et al.*, 1989b ; Anthony *et al.*, 1993 ; Rakotomalala *et al.*, 1993), de 0,14 % MS chez l'espèce *C. rhamnifolia* Bridson à 9,90 % MS chez *C. sessiliflora*. Le taux de CQA (CQA/CGA) varie de 61 % chez *C. sp. mouloundou* à 100 % chez *C. farafanganensis* Leroy. Chez *C. canephora*, les taux de DQA (DQA/CGA) et de FQA (FQA/CGA) atteignent respectivement 36 % et 12 %. Si l'on considère le continent Africain, un gradient de la teneur en acides chlorogéniques existe entre espèces d'Afrique de l'est et d'Afrique de l'ouest (Anthony *et al.*, 1993).

### Les éléments minéraux

Les substances minérales que l'on retrouve sous forme de cendres (après combustion totale des grains verts de café) représentent en moyenne 4 % MS. Le potassium est de loin le plus abondant, avec près de 1,7 % MS des grains (Poisson, 1977). A côté de cela, de nombreux oligoéléments sont décelés, comme par exemple le cuivre qui serait impliqué dans les mécanismes de résistance aux maladies (Quijano-Rico et Spettel, 1973).

### **2. 2. 2. Les principaux constituants biochimiques de l'arôme et du goût du café**

Différentes transformations technologiques post récolte (dépulpage par voie sèche ou par voie humide, triage et transport, torréfaction des grains) vont concourir à la formation de composés aromatiques à partir des précurseurs d'arôme des grains verts (De Maria *et al.*, 1994).

L'arôme du café torréfié est un mélange complexe de composés aromatiques appartenant à de nombreuses familles chimiques. On trouve aussi bien des composés hautement hydrophobes tels les hydrocarbures, que des composés hautement hydrophiles comme les alcools et les acides. Chaque composé volatil possède ses propres caractéristiques

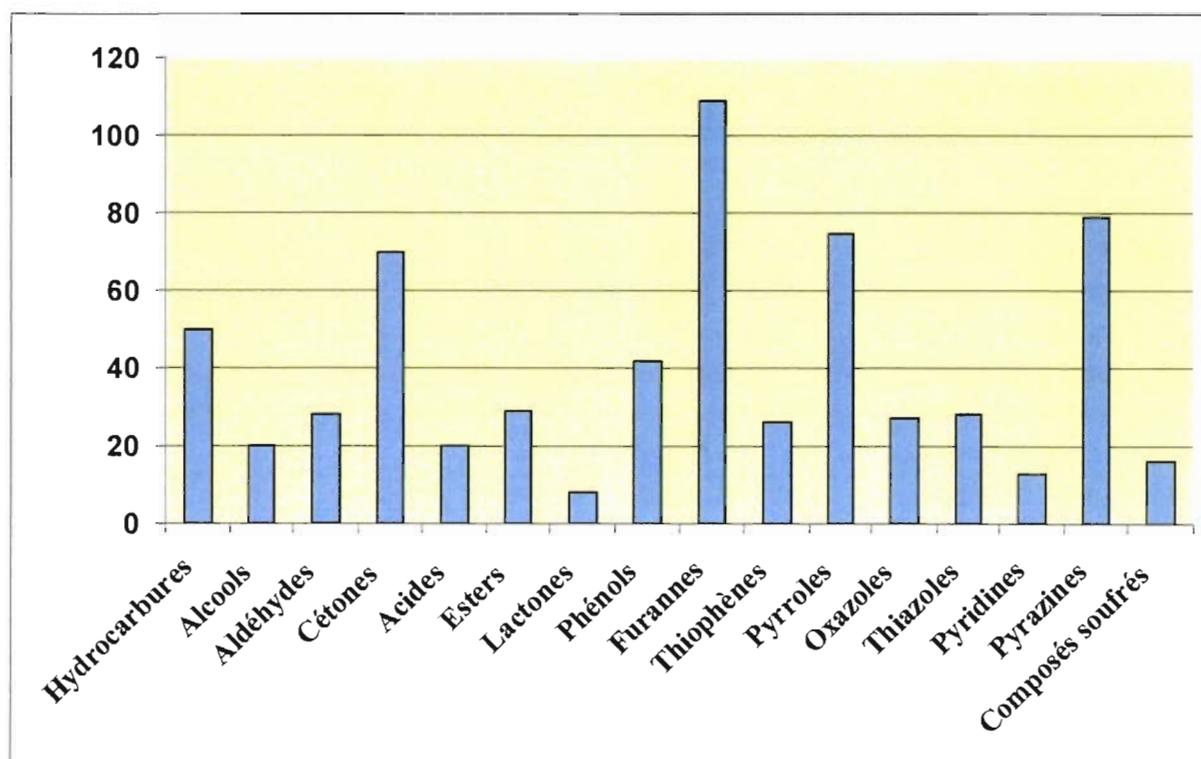


Figure 2.2.4. : Nombre de composés aromatiques identifiés dans l'arôme café (Flament, 1991).

olfactives, que nous percevons comme une note agréable ou désagréable. La concentration, la proportion et l'interaction d'un composé avec un (ou plusieurs) autre(s) affectent la qualité finale de l'arôme. Cette dernière n'est pas la résultante additive des effets de chacun des composés, mais des combinaisons synergiques ou antagonistes de ces effets.

Le but de ce paragraphe est de présenter brièvement les composés aromatiques (hydrocarbures, alcools, aldéhydes, cétones, composés soufrés, furannes, pyrazines, pyrroles et pyridines) (Fig. 2.2.4.), en relation avec leurs propriétés olfactives et leurs formations à partir des constituants biochimiques du café vert (Tab. 2.2.1.).

### Les composés non-hétérocycliques

Cette classe de composés comprend (par ordre de complexité de la fonction organique) : les hydrocarbures, les alcools, les aldéhydes, les cétones, les acides, les esters, les lactones, les phénols, les composés soufrés et la caféine.

#### *Les hydrocarbures*

Parmi les hydrocarbures, de faibles proportions d'alcanes et d'alcènes ont été trouvées dans l'arôme du café. Ces derniers sont générés par transformation chimique sous haute température lors de la torréfaction (Ohloff, 1978). Les monoterpènes possèdent des notes aromatiques particulièrement appréciées (Guadani *et al.*, 1966). Parmi eux, un constituant typique du café : le 4,  $\alpha$ -diméthylstyrène qui possède une note d'« agrume » (Arctander, 1969).

#### *Les alcools*

Les alcools sont générés à partir de transformations enzymatiques d'acides aminés et d'acides-gras insaturés (De Lumen *et al.*, 1978). A l'inverse des alcools saturés, ceux dont la chaîne aliphatique est insaturée participent à l'arôme. C'est le cas typique du 1-octène 3-ol qui possède une odeur typique de champignon (désigné sous le nom de « mushroom alcohol ») (Cronin et Ward, 1971). La plupart des alcools sont d'importants précurseurs de la formation d'autres composés aromatiques. C'est le cas du linalol découvert seulement depuis 1967 et abondant dans le café (Stoll *et al.*, 1967).

#### *Les aldéhydes et les cétones*

Les aldéhydes sont issus de la dégradation enzymatique d'acides gras à chaîne hautement insaturée (Forss, 1969). Les aldéhydes à chaîne aliphatique saturée tels que le nonanal ou le décanal présentent des notes « fruité » et « fleurie » (Ohloff, 1978). Ces notes

<b>Composés aromatiques</b>	<b>Précurseurs majeurs</b> <i>des grains verts</i>	<b>Notes organoleptiques dominantes</b>
<b>Hydrocarbures</b>	Lipides	Fatty, Fruity, Oily
<b>Alcool</b>	Acides aminés, Acides gras	Green, Mushroom, Leafy
<b>Aldéhydes</b>	Lipides, Acide gras	Fruity, Honey, Green, Woody
<b>Cétones</b>	Lipides, Acide gras	Fruity, Banana, Flowery
<b>Acides</b>	Lipides	Fruity, Flowery
<b>Esters</b>	Lipides	Green, Fruity, Flowery
<b>Lactones</b>	Lipides	Creamy, Sweet, Flowery
<b>Phénols</b>	Acides chlorogéniques	Smoky, Spicy, Burnt
<b>Furannes</b>	Glucides	Earthy, Caramel, Nutty, Roasty
<b>Thiophènes</b>	Monosaccharides, Acides aminés	Sulfury, Nutty
<b>Pyrroles</b>	Monosaccharides, Acides aminés	Buttery, Caramel, Burnt
<b>Oxazoles</b>	Acides aminés	Sweet, Nutty
<b>Thiazoles</b>	Monosaccharides, Acides aminés	Green, Nutty, Roasty
<b>Pyridines</b>	Trigonelline, acides aminés	Green, Roasty, Earthy, Bitter, Burnt
<b>Pyrazines</b>	Acides aminés	Nutty, Green, Earthy
<b>Composés soufrés</b>	Acide aminés soufrés	Potato, Smoky, Sulfury

**Tableau 2.2.1.** : Constituants biochimiques des grains verts impliqués dans la formation et notes organoleptiques associées à l'arôme café (Tressl, 1989 ; Flament, 1991 ; Shibamoto, 1991).

sont aussi caractéristiques des cétones à chaîne aliphatique saturée, qui proviennent de l'oxydation des lipides (Forss, 1969). En revanche, les cétones insaturés présentent un grand intérêt organoleptique. Ils véhiculent des notes « caramélisé », indispensables aux cafés de grande qualité comme ceux de Colombie (Gianturco *et al.*, 1963).

#### *Les acides chlorogéniques*

Environ trente six pour cent de la fraction totale en acides chlorogéniques n'est pas détruite à l'issue de la torréfaction chez *C. canephora* var. Robusta (Lentner et Deatherage 1959 ; Feldman *et al.*, 1969), conduisant à un effet astringent dû à la présence des isomères appartenant à la classe des DQA (Ohiokpehai *et al.*, 1982 ; Clifford et Ohiokpehai, 1982). L'un des facteurs biochimiques de qualité du café Arabica en comparaison à celle du Robusta est lié à ses faibles teneurs en CGA (Clifford, 1985b ; Guyot *et al.*, 1988, 1996).

#### *Les lactones et les phénols*

Les lactones appartiennent à une famille très peu représentée en nombre dans l'arôme du café. Ils possèdent une note agréable « fruité » et « fleurie » (Keppler, 1977).

A l'opposé, les phénols sont des constituants essentiels de l'arôme café, avec plus de cinquante membres identifiés (Tressl *et al.*, 1978). Ils sont associés en général à une odeur de fumée (« smoky ») (Shimizu, 1970). Ils sont essentiellement formés à partir de la dégradation thermique des lignines et des acides chlorogéniques (Fiddler *et al.*, 1967) et participent ainsi à l'amertume du café (Leloup *et al.*, 1995).

#### *Les composés soufrés*

Les composés volatiles contenant un ou des atomes de soufre jouent un rôle essentiel dans le développement de l'arôme. Parmi les composés aliphatiques soufrés, le sulfide d'hydrogène (SH<sub>2</sub>) est l'élément le plus simple produit par dégradation de la cystéine. Il est à la fois un important constituant et un intermédiaire de la formation de nombreux autres composés aromatiques (Fujimaki *et al.*, 1969). Le méthane-thiol et le diméthyle sulfide formés à l'issue du clivage de la méthionine agissent comme « multiplicateur » de l'arôme du café (Wainwright *et al.*, 1972 ; Guadagni *et al.*, 1969).

#### *La caféine*

La caféine n'est pas dégradée lors de la torréfaction. D'un point de vue organoleptique, la caféine est une substance amère (Voilley *et al.*, 1977).

## Les composés hétérocycliques

Cette classe de composés hétérocycliques comprend par ordre quantitatif d'importance : les furannes, les pyrazines, les pyrroles, les oxazoles et thiazoles, les thiophènes, les pyridines, et les dérivés soufrés.

### *Les furannes*

Plus d'une centaine de furannes ont été identifiés. Cette famille de composé possède un cycle furannique dont les chaînes latérales peuvent inclure de nombreuses fonctions chimiques. La plupart des furannes possèdent des notes « caramélisée » et « grillée », comme le 5-méthyl-4-hydroxy-3 (2H)-furanone, le 2,5-diméthyl-4-hydroxy-3 (2H)-furanone (furanneol) et le 2-hydroxy-3-méthyl-2-cyclothène-1-one (cyclotène) (Shibamoto, 1991). En effet, les furannes sont directement produits par pyrolyse des sucres (Baltes et Bochmann, 1987a et 1987b ; Baltes et Mevissen, 1988). D'autres sont extensivement utilisés dans l'aromatization des boissons alcoolisées, des crèmes glacées et des sucreries. C'est le cas du furfural et du furanneol. Cependant, certains furannes n'ont pas un impact direct sur l'arôme. C'est le cas en particulier des furannes alkylés et alkenylés, dont le rôle majeur est leur capacité à produire indirectement des arômes dominants (furannes sulfurés), au travers de réactions secondaires avec des résidus pourvus de groupements sulfures (Shibamoto, 1991). Parmi les composés cycliques soufrés dérivés, le kahwéofuranne est caractéristique du café torréfié (Stoll *et al.*, 1967). Il possède une note « plaisante » et de « fumée » (Flament, 1991).

### *Les pyrazines*

Les pyrazines sont après les furannes, les composés hétérocycliques les plus abondants du café. Plus de 80 pyrazines ont été identifiés dans l'arôme. Les groupements fonctionnels présents sur les noyaux pyraziniques déterminent la nature de l'odeur et son seuil de détection (Shibamoto, 1991). Les méthylpyrazines possèdent une odeur de noisette grillée, alors que la majorité des pyrazines alkylées ont des notes « vert », voir « huileux » pour les dérivés hautement alkylés (Flament, 1991). Le seuil de détection est lié à la présence et au nombre d'atomes d'oxygène et/ ou de soufre présent(s) sur les chaînes latérales. Ainsi, les alkylpyrazines sont moins odorants (seuils plus élevés) que leurs homologues alkoxy-pyrazines ou alkythio-pyrazines.

Les pyrazines sont issues des réactions de Maillard (Maillard 1912, Ho *et al.*, 1993 ; Reineccius, 1995) qui ont lieu entre 100 et 150°C (lors de la torréfaction) entre acides aminés et monosaccharides. Ces réactions impliquent de nombreuses processus (condensation,

déshydratation, énoilisation, polymérisation...) entre un groupement carbonyle d'un glucide (polysaccharide ou monosaccharides) et un groupement amine (protéine ou acides aminés libres) (Baltes, 1990). Les monosaccharides impliqués dans ces réactions proviennent de l'hydrolyse du saccharose et de la dégradation des polysaccharides sous l'effet de la température. La formation des pyrazines est inhibée par les acides chlorogéniques (Guyot *et al.*, 1997).

#### *Les pyrroles*

Les pyrroles sont d'importants constituants de l'arôme café par leur nombre et leur diversité. De nombreux pyrroles ont été identifiés dans la fraction volatile de sérine, thréonine et de saccharose torréfiés en système modèle. Les teneurs en pyrroles sont d'autant plus élevées que la durée de torréfaction est longue. La pyrolyse de la trigonelline est une source en ces composés (De Maria *et al.*, 1994 et 1996). A l'état dilué, les alkyl et les acyl-pyrroles ont des arômes légèrement « caramélisé ».

#### *Les oxazoles et thiazoles*

Les oxazoles sont formés à partir de réaction entre acides aminés (Ho *et al.*, 1983). Près d'une trentaine d'oxazoles ont été identifiés dans l'arôme du café. Ils sont caractérisés par une note douce de « noisette », comme le 2,4,5-triméthylloxazole (Maga, 1981 ; Jin *et al.*, 1984).

Les thiazoles ont été pour la première fois découverts dans l'arôme du café (Stoll *et al.*, 1967). Certains composés de la famille possèdent des notes « vert » et « doux » et d'autres des notes « brûlé » (Pittet et Hruza, 1974).

#### *Les thiophènes*

Plus de 30 composés de la famille des thiophènes sont caractéristiques de l'arôme du café. La plupart des thiophènes possèdent une odeur sulfureuse (Shibamoto, 1980). Les thiophènes sont produits à la suite des réactions de Maillard entre le glucose et les acides aminés soufrés (Shibamoto et Russell, 1977 ; Sakaguchi and Shibamoto, 1978).

#### *Les pyridines*

Les pyridines résultent entre autre, de la dégradation thermique de la trigonelline (Viani et Horman, 1974 et 1975). Ils sont aussi formés à la suite de dégradations thermiques d'acides aminés et de réactions entre acides aminés et sucres (Suyama et Adachi, 1980). Des notes « vert », « amère », « brûlé », « noisetté » leurs sont associées.

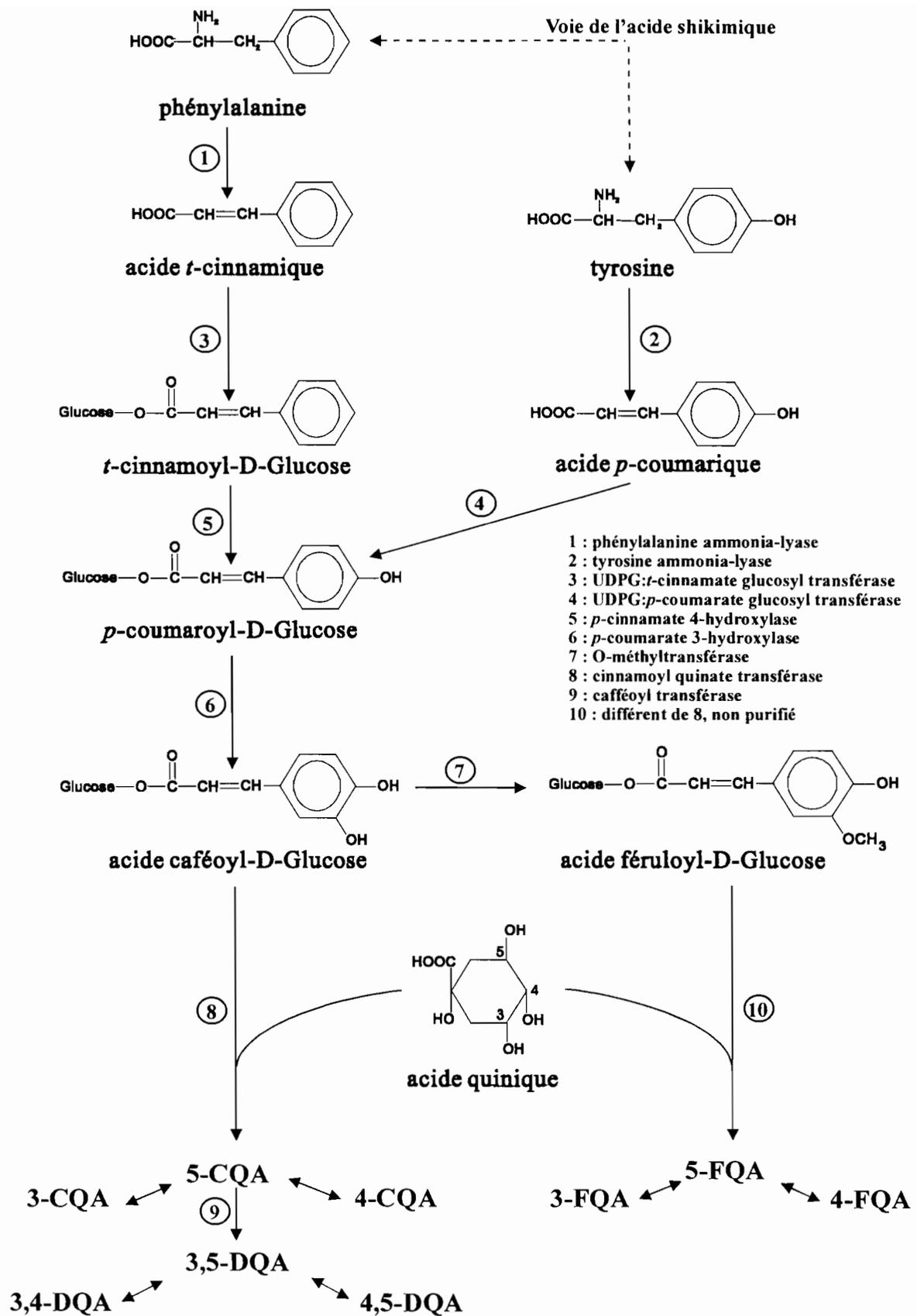


Figure 2.2.5. : Chaîne de biosynthèse conduisant à la formation des acides chlorogéniques.

### 2. 2. 3. Biosynthèse et génétique des composés biochimiques étudiés

Trois familles de composés biochimiques ont été retenues dans l'ensemble de notre étude en raison essentiellement de : i) leur implication plus ou moins directe dans le goût du café à la tasse, ii) leur relative facilité d'extraction et d'analyse (adaptable à un grand nombre d'échantillons et de répétitions, par rapport aux macromolécules comme les lipides, les polypeptides et les polysaccharides) et iii) de l'existence d'une grande variabilité pour ces composés dans le genre *Coffea*. Ces trois familles de composés sont : i) les acides chlorogéniques, ii) les deux alcaloïdes majoritaires, caféine et trigonelline et iii) le saccharose.

#### Les acides chlorogéniques

Les voies de biosynthèse conduisant à la formation des isomères de CGA sont illustrées sur la figure 2.2.5. Les enzymes présentées sur le schéma ont été purifiées sur différents matériels végétaux et par différents auteurs (Hahlbroch et Grisebach, 1979 ; Ulbrich et Zenk, 1979 et Friedman, 1997), mais existent certainement chez les caféiers (Clifford, 1985b).

Tous les acides chlorogéniques dérivent de deux acides aminés aromatiques (formés par la voie du shikimate) : la phénylalanine et la tyrosine (Villegaas et Kojima, 1986). Ces acides aminés sont transformés en acide caféique (caféate), dont la forme réactive est préférentiellement complexée au glucose (Fleuriet et Macheix, 1981 ; Moriguchi *et al.*, 1988 ; Tanaka et Kojima, 1991) et, dans une moindre mesure, au coenzyme A (Stöckigt et Zenk, 1974). Cette unité caféate réagit avec l'acide quinique (quinate : formé à partir du glucose), afin de former l'isomère prépondérant : le 5-CQA. L'enzyme responsable de cette estérification est le cinnamoyl quinate transférase (ou CQT) (enzyme 6 sur la figure 2.2.5.). Dans la classe des DQA, c'est le 3,5-DQA qui est initialement formé, lors de l'estérification d'une seconde unité caféate en position 3 à partir du 5-CQA.

A l'opposé, la formation des FQA est indépendante de celle du 5-CQA. Une quinate transférase, différente du CQT catalyse la réaction de condensation entre l'unité quinate et férulate (Ulbrich et Zenk, 1979).

Enfin, les autres isomères des classes CQA, DQA et FQA seraient formés par trans-estérification enzymatique du 5-CQA, du 3,5-DQA et du 5-FQA respectivement (Hanson, 1965 ; Friedman, 1997).

Colonna (1986) a montré que ces biosynthèses avaient lieu au niveau des feuilles. Les isomères d'acide chlorogénique ainsi formés sont exportés au niveau des fruits, où leur

accumulation est maximale lorsque ces derniers sont verts. Au cours de la maturation, le taux de CGA décroît (Petnga, 1986).

Une seule étude de génétique quantitative nous est connue et concerne d'une part la teneur globale en CGA et d'autre part des croisements intraspécifiques chez *C. canephora* (Montagnon *et al.*, 1998). L'héritabilité *s.s.*, qui reflète l'additivité, est relativement faible (36%).

### La caféine

La chaîne de biosynthèse conduisant à la formation de la caféine chez les caféiers est actuellement bien connue (Schulthess *et al.*, 1995 ; Schulthess et Baumann, 1995) (Annexe 2.2.). Cette dernière a lieu principalement au niveau de l'exocarpe et du mésocarpe du fruit (la lumière étant un stimulant de la méthylation du xanthosine). La caféine migre ensuite vers la graine durant la maturation. Bien que l'alcaloïde se forme aussi au niveau des feuilles, il ne semble pas y avoir de translocation vers les fruits (Baumann et Wanner, 1972). Dans les tissus de grains de café, la caféine est liée aux isomères d'acide chlorogénique et particulièrement au 5-CQA (Baumann *et al.*, 1993) et au 3,5-DQA (Horman et Viani, 1971 ; Kappeler *et al.*, 1987 ; Baumann et Röhrig, 1989).

Un croisement entre une espèce de caféiers avec caféine, *C. liberica dewevrei* (DEW), et une espèce sans, *C. pseudo-zanguebariae* (PSE), a montré que la présence *vs.* l'absence est déterminée par un gène majeur où l'absence est récessive (Barre *et al.*, 1998a).

Chez les espèces avec caféine, les études quantitatives ont principalement concerné les variétés de *C. canephora* et de *C. arabica*, les croisements intraspécifiques et les hybrides F1 interspécifiques (Charrier et Berthaud, 1975; Louarn, 1976; Le Pierres, 1987; Mazzafera et Carvalho, 1992). Ces études quantitatives ont conduit à deux résultats principaux, confirmés chez les hybrides de seconde génération avec caféine du croisement PSE x DEW (Barre *et al.*, 1998a) : i) la teneur en caféine est un caractère quantitatif à hérédité polygénique et ii) les effets génétiques sont prépondérants par rapport aux effets environnementaux et représentent 94% de la variabilité.

La teneur en caféine est souvent notée comme additive chez les hybrides intra- et interspécifique impliquant des espèces avec caféine, quel que soit le niveau de ploïdie (Capot, 1972; Charrier et Berthaud, 1975; Louarn, 1976; Berthaud, 1977; Le Pierres, 1987). Dans le croisement PSE x DEW, la teneur en caféine n'est pas additive, mais une transformation « racine carrée » suffit pour accepter cette hypothèse (Barre *et al.*, 1998a).

### La trigonelline

La trigonelline est abondante dans la chaîne de biosynthèse des nucléotides pyrimidiques conduisant à la formation du NAD (Nicotinamide Adénine Dinucléotide) (Tramontano *et al.*, 1982).

L'étude de génétique quantitative mentionnée pour les acides chlorogéniques chez *C. canephora* traite aussi de la trigonelline (Montagnon *et al.*, 1998). Comme précédemment, l'héritabilité *s.s.* est relativement faible (38%).

### Le saccharose

A ce jour, aucune étude physiologique concernant l'accumulation du saccharose chez le caféier n'a été publiée. En revanche, le métabolisme du saccharose a été bien étudié chez la tomate (*Lycopersicon* sp.). La biosynthèse du saccharose à partir du fructose et du glucose (métabolites primaires) est catalysée par la saccharose phosphate synthase (SPS) dans le cytoplasme des cellules (Hubart *et al.*, 1989 ; Bruneau *et al.*, 1991). La SPS catalyse la réaction : Fructose-6-phosphate + UDP-glucose → saccharose phosphate + UDP.

Deux autres enzymes sont impliquées dans la réaction inverse de clivage du saccharose : (1) une transférase et (2) une invertase (Ross et Davies, 1992). Les espèces sauvages de tomate (*L. peruvianum*, *L. hirsutum*, *L. chmielewski* du sous genre *Eriolycopersicon*) à fruits verts à maturité ont des teneurs élevées en saccharose et sont caractérisées par une faible activité acide invertase (Miron et Schaffer, 1991 ; Stommel, 1991, Sun *et al.*, 1992). En fait, l'accumulation de saccharose est déterminée par une balance entre synthèse et dégradation.

Chez le caféier, la génétique de la teneur en saccharose des grains verts n'a été étudiée que dans des croisements entre accessions de *C. canephora* impliquant de faibles différences entre géniteurs (Montagnon *et al.*, 1998). L'héritabilité *s.s.* est très faible (11%).

Des analyses génétiques plus avancées existent chez la tomate. Des descendance interspécifiques entre *L. chmielewskii* et *L. esculentum* (espèce cultivée) ont permis de montrer que le caractère "accumulation de saccharose" est : i) gouverné par un gène récessif et ii) associé à la fois à de faible activité acide invertase (Klann *et al.*, 1993 ; Stommel et Haynes 1993).

#### 2. 2. 4. En resume

L'arôme caractéristique et unique du café est constitué de plus de mille composés chimiques volatils identifiés. Les critères de qualité d'un café se basent sur le goût et l'arôme qu'il dégage. Les principales modalités du goût sont au nombre de quatre (amer, salé, aigre et sucré), auxquelles on peut rajouter la perception d'astringence. Par contre, les modalités de l'odorat restent en discussion et les 44 qualités définies par Harper (cité par Dart et Nurstein, 1985) sont un bon résumé de l'état actuel des connaissances. L'étude des mécanismes de formation des composés volatils a un intérêt considérable, mais reste d'une très grande complexité. Les voies proposées sont basées sur des considérations théoriques supportées par les études de systèmes modèles. La composition biochimique de la boisson est le résultat d'une suite d'opérations, qui se sont déroulées depuis la mise en culture des caféiers jusqu'à la torréfaction des grains verts. La composition de ces derniers apparaît donc comme un des facteurs intervenant dans la valeur du café marchand. La connaissance de cette composition n'est pas dépourvue d'intérêt dans la mesure où elle permet de mieux appréhender les facteurs biochimiques de la qualité du café. En effet, c'est sous l'action de la chaleur (les réactions de Maillard entre autres) que certains composés biochimiques caractéristiques du café vert vont se dégrader et conduire à la formation de nouveaux produits responsables de l'arôme. Ces composés biochimiques sont les précurseurs d'arôme.

Les acides chlorogéniques, les alcaloïdes (caféine et trigonelline) et le saccharose constituent les composés analysés tout au long de ce travail. Ce dernier repose sur la variabilité (Chap. 4), la génétique quantitative (Chap. 5) et la recherche de QTLs (Chap. 7) de ces composés. L'ensemble de ces études va concourir à donner une réponse au questionnement scientifique suivant :

- i) Quels sont les relations physiologiques et génétiques liant ces différents composés ?
- ii) Combien et où sont localisés les QTLs responsables des variations de leurs teneurs ?

Et à long terme, à se poser les suivantes :

- iii) Quels sont les gènes impliqués dans leur métabolisme ?
- iv) Quelles sont les conséquences sur la qualité du café de la modification d'une voie métabolique par surexpression ou suppression de l'accumulation d'autres composés biochimiques ?

## 2. 3. Cartographie génétique et recherche de QTLs

Cette partie est scindée en deux volets. Le premier traite de la cartographie génétique. Il constitue un rappel concernant le principe de construction de carte génétique, leurs applications et soulève les problèmes rencontrés dans le cas de croisement interspécifique de plantes allogames. Le second volet traite de la recherche des QTLs, en présentant brièvement le principe et les méthodes, ainsi que les principaux facteurs influençant leurs détections.

### 2. 3. 1. Cartographie génétique

La carte génétique d'une espèce est une représentation de son génome sur lequel des marqueurs sont régulièrement dispersés le long de groupes de liaison représentant les chromosomes. Le principe de construction d'une telle carte est fondé sur la recombinaison de segments de chromosomes homologues lors de la méiose. La méthodologie générale consiste à croiser deux individus différant pour un certain nombre de marqueurs. La ségrégation de ces marqueurs est analysée dans la descendance du croisement (De Vienne, 1998).

#### Marqueurs moléculaires et descendance en ségrégation

Les marqueurs les plus utilisés sont de type moléculaire et révèlent un grand nombre de locus polymorphes au niveau du génome. Les quatre types de marqueurs moléculaires les plus couramment utilisés sont : (i) les RFLP (Restriction Fragment Length Polymorphism) ou polymorphisme de longueur des fragments de restriction (Grodzicker *et al.*, 1974 ; Botstein *et al.*, 1980 ; Tanksley, 1983), (ii) les RAPD (Random Amplified Polymorphic DNA) ou polymorphisme d'amplification aléatoire de l'ADN (Williams *et al.*, 1990 ; Welsh et McClelland, 1990), (iii) les AFLP (Amplified Fragment Length Polymorphism) ou polymorphisme de longueur de fragment amplifié (Vos *et al.*, 1995) et (iv) les microsatellites, qui révèlent un polymorphisme du nombre de répétitions d'un motif de N nucléotides (Tautz, 1989 ; Beckmann et Soler, 1990).

Quatre types principaux de descendance en ségrégation dérivées de croisements entre lignées fixées permettent d'étudier les recombinaisons entre marqueurs. Des descendance F2 issues de l'autofécondation d'individus F1. Des descendance issues de croisement en retour (rétrocroisements ou backcross) d'un hybride avec un de ses parents récurrents. Des descendance issues d'haplodiploïdisation (HD) par régénération de plantes fonctionnelles à

partir de gamètes de la F1. Des populations de lignées recombinantes (RIL) obtenues par autofécondations successives à partir d'individus F2.

### Principe et méthode de construction

L'obtention d'une carte des marqueurs consiste successivement : i) à réunir les marqueurs en groupes de liaison, ii) à ordonner les marqueurs au sein des groupes et iii) à estimer au mieux les distances de recombinaison entre marqueurs.

L'ordre relatif des marqueurs et les distances de recombinaison les séparant les uns des autres le long des groupes de liaison sont déterminés par des tests statistiques. L'analyse « deux points » de toutes les paires de marqueurs permet de former des groupes de liaison. Puis, l'ordre le plus probable des marqueurs est déterminé par une analyse « multipoint ». Cette démarche est souvent conduite à l'aide du logiciel Mapmaker (Lander *et al.*, 1987) qui utilise la méthode du maximum de vraisemblance (Fisher, 1936) pour estimer les distances entre marqueurs et teste la liaison entre marqueurs par le calcul d'un *LOD score* (*LOD = Logarithm of the odds ratio* : Morton, 1955). Le *LOD score* correspond au rapport de la probabilité de liaison entre deux marqueurs sur la probabilité d'indépendance. Des fonctions de transformation rendent compte de la correspondance entre fréquence de recombinaison et distance de carte. La fonction de Haldane suppose que les événements de recombinaison se répartissent au hasard selon une loi de Poisson (Haldane, 1919). Les distances de Kosambi semblent mieux appropriées car la fonction correspondante tient compte des interférences chromosomiques (Kosambi, 1944). On aboutit alors, lorsque la carte est saturée, à un nombre de groupes de liaison égal au nombre chromosomique de l'espèce. Pour certains organismes, les groupes de liaison ainsi détectés ont été mis en correspondance avec les chromosomes.

### Applications des cartes génétiques

Chez les végétaux, deux applications majeures sont faites des cartes de liaison génétique.

Premièrement, elles permettent l'étude et la comparaison des génomes d'espèces plus ou moins éloignées en fournissant des informations sur l'évolution et la divergence de ces espèces. Ainsi, une bonne colinéarité des cartes du sorgho avec d'autres plantes ont été observée comme : le maïs (Binelli *et al.*, 1992 ; Dufour *et al.*, 1996), la canne à sucre (Dufour *et al.*, 1997 ; Glaszmann *et al.*, 1997) et le riz (Wilson *et al.*, 1999). De même, le cas tomate/pomme de terre présente une répartition et un ordre commun des marqueurs le long des

chromosomes, les seules différences majeures mises en évidence étant trois inversions paracentriques (Bonierbale *et al.*, 1988 ; Gebhardt *et al.*, 1991). En revanche, dans le cas tomate/ piment, bien que de nombreux gènes communs existent, leur ordre est fortement modifié (Tanksley *et al.*, 1988). Ces cartes servent aussi à comprendre la structuration des différentes espèces et leur différenciation à partir d'un ancêtre commun : dans le genre *Brassica* (Kianian et Quiros 1992 ; Mc Grath et Quiros 1992) et chez les graminées (Devos et Gale, 1997 ; Gale et Devos, 1998).

Deuxièmement, ces cartes servent à l'étude des caractères agronomiquement intéressants. Ces derniers concernent aussi bien des caractères à déterminisme monogénique que polygénique. Ce dernier cas fait l'objet de la suite de notre étude concernant la recherche des QTLs.

### Différences entre cartes génétiques intra- et interspécifiques

La moindre homologie des génomes lors des croisements interspécifiques conduit à :  
i) une réduction des taux de recombinaisons et ii) l'apparition de distorsions de ségrégation.

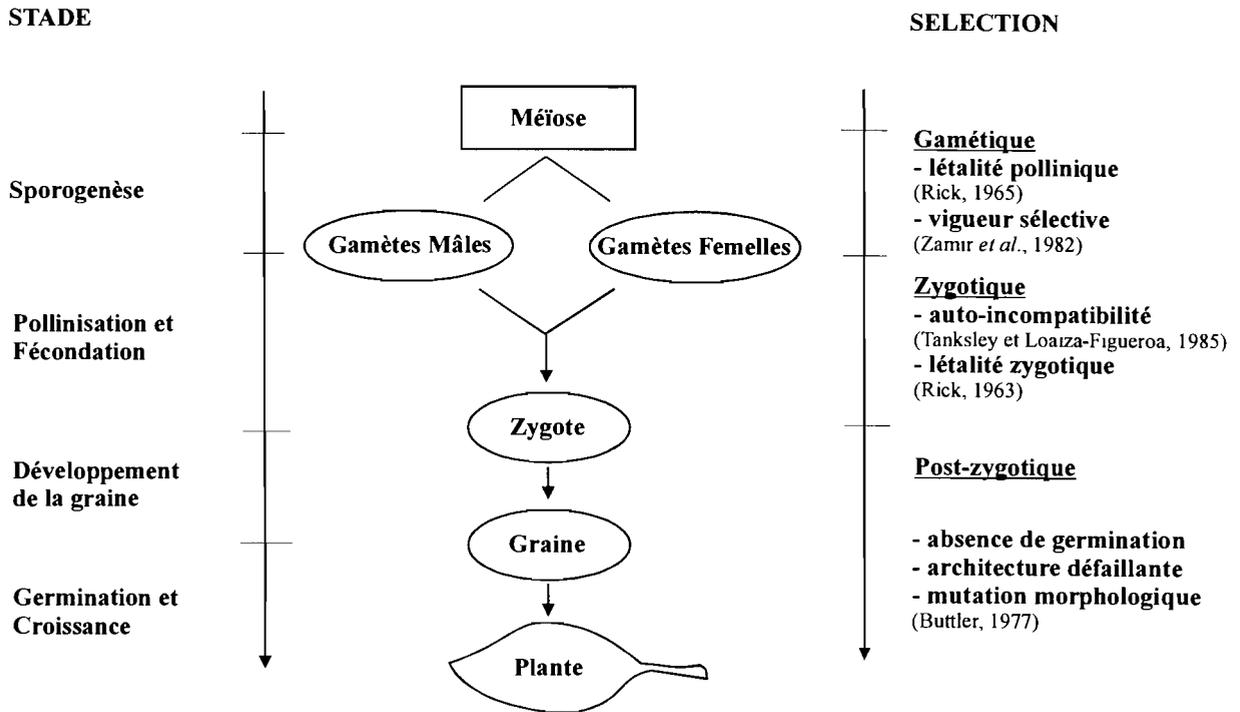
#### *Taux de recombinaison et taille de la carte*

La faible homologie entre chromosomes homologues chez les hybrides interspécifiques entraîne une réduction des taux de crossing-over durant la méiose (Radman et Wagner, 1993). Les descendances issues de croisements interspécifiques conduisent ainsi à l'obtention de cartes génétiques plus compactes et plus courtes que celles obtenues avec des croisements intraspécifiques. Cette réduction de taille de la carte a été confirmée chez la pomme de terre où la carte d'un croisement intraspécifique a une taille plus grande (65%) que celle d'un croisement interspécifique (Bonierbale *et al.*, 1988 ; Gebhardt *et al.*, 1991). La même tendance est observée chez la tomate (Paterson *et al.*, 1988) et chez le riz (Causse *et al.*, 1994).

#### *Distorsions de ségrégation*

Dans le cas particulier des croisements interspécifiques, où les génotypes parentaux sont souvent très différents, des ségrégations différentes de la ségrégation mendélienne attendue sont très souvent observées (Grant, 1975). L'écart constitue la distorsion de ségrégation et les marqueurs associés sont dits distordus (Zamir et Tadmor, 1986). Ces derniers peuvent former des segments chromosomiques, voire des chromosomes entiers, et concerner jusqu'à près de 38 % de la totalité des locus dans des croisements interspécifiques d'agrumes (Luro *et al.*,

**Figure 2.3.1.** : Origine biologique des distorsions de ségrégation observées chez les végétaux, en relation avec leur stade de développement. D'après Gadish et Zamir (1986).



1994). En règle générale, plus le degré d'homologie entre génome est grand, moins de régions chromosomiques présentent de distorsion (Thoquet, 1995). Chez les végétaux, de nombreuses distorsions en croisement interspécifique ont été observées (Xu *et al.*, 1997) et plus récemment chez *Pinus radiata* (Kuang *et al.*, 1999), *Arabidopsis thaliana* (Grini *et al.*, 1999) ou *Oryza sativa* (Virk *et al.*, 1998).

Les distorsions de ségrégation les plus courantes sont gamétiques, zygotiques et/ ou post zygotiques (Stebbins, 1958 ; Rick, 1963, 1965, 1969 ; Gadish et Zamir, 1986) (Fig. 2.3.1.) selon le stade où le processus intervient.

La sélection gamétique repose sur l'exclusion sélective de certains gamètes en relation avec leur contenu génétique. Elle est due à la combinaison d'allèles de gène à effet négatif ou létal. La fréquence de ces combinaisons découle du faible taux de recombinaison existant chez l'hybride interspécifique (Stam et Zeven, 1981) et/ ou de la présence de remaniements chromosomiques comme les translocations (Fauré, 1993). De nombreux exemples de combinaison de gènes létaux conduisant à une stérilité pollinique ou à une incompatibilité pollinique ont été étudiés. Chez la tomate, la présence de l'allèle *Ge* (« Gamete eliminator ») entraîne la mort des gamètes (Rick, 1966).

Les sélections zygotique et post- zygotique correspondent à l'ensemble des mécanismes qui concourent à réduire la viabilité, la vigueur et la fertilité des hybrides après la fécondation. Certains facteurs génétiques seraient contre-sélectionnés lors de la croissance du tube pollinique ou de la formation de l'embryon comme l'ont suggéré Gadish et Zamir (1986) en parlant de mortalité différentielle entre les zygotes. Certaines distorsions zygotiques dépendent de la présence d'un système d'incompatibilité. C'est le cas chez la pomme de terre et chez la tomate (*L. peruvianum*) où un tel système a été cartographié sur le chromosome I (Tanksley et Loaiza-Figueroa, 1985 ; Gebhardt *et al.*, 1991).

### Conséquence de l'allogamie sur l'établissement de cartes génétiques

Chez les végétaux, la cartographie repose couramment sur des descendance issues de parents de lignées fixées. Dans le cas d'une descendance de type backcross 1 par exemple, les gamètes formés par le parent récurrent génétiquement fixé sont identiques et seuls les gamètes produits par la F1 vont être responsables de la ségrégation observée dans la descendance.

Lorsque la descendance est issue d'espèces allogames où les géniteurs sont génétiquement non fixés, la construction des cartes est plus complexe et ceci est dû à deux facteurs : i) les géniteurs sont fortement hétérozygotes : jusqu'à quatre allèles peuvent alors

ségrèger par locus (il n'y en a que deux dans les descendance de lignées fixées) et ii) l'observation des phases des allèles (couplage ou répulsion) nécessite le calcul d'un maximum de vraisemblance (Ritter *et al.*, 1990) (cette connaissance est immédiate chez les descendance de parents fixés).

Différentes méthodes permettent cependant la réalisation de cartes génétiques chez des espèces non fixées : i) l'autofécondation d'un génotype hétérozygote assimilable à une descendance F2, ii) le croisement entre un géniteur fortement homozygote et une plante fortement hétérozygote assimilable à une descendance BC (c'est en fait un pseudo-backcross, où seuls les locus d'un des parents ségrègent comme dans un BC), iii) l'utilisation de populations haploïdes doublés comme chez *C. canephora* (Paillard *et al.*, 1996) et iv) considérer un croisement entre parents hétérozygotes comme un double pseudo-testcross (Grattapaglia et Sederoff, 1994). Cette dernière méthode consiste à décomposer ce croisement en deux testcross : le testcross d'un parent sur l'autre et réciproquement. On observe alors séparément deux ségrégations gamétiques : la ségrégation des allèles du premier parent dans la descendance, sans tenir compte des allèles apportés par le deuxième parent, et inversement pour ce dernier.

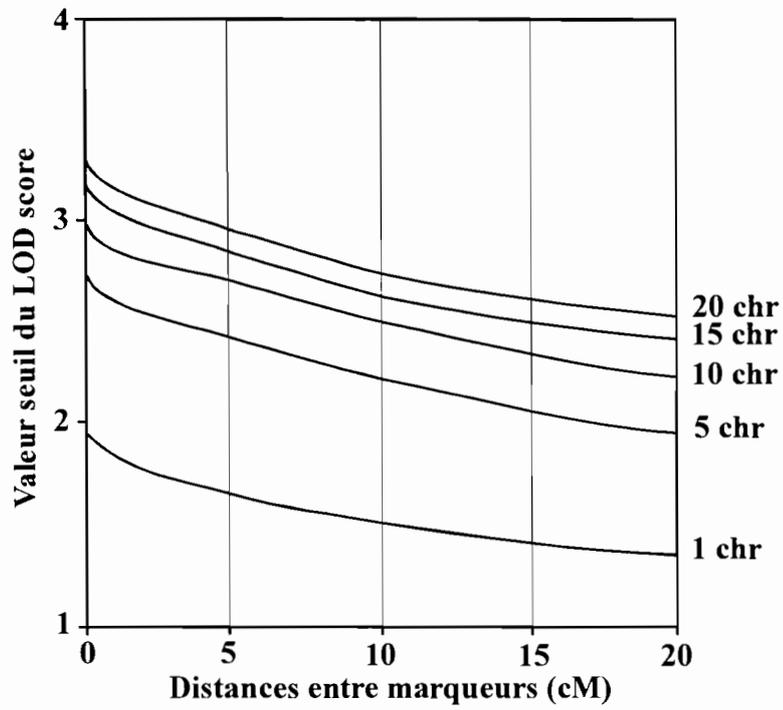
### 2.3.2. Recherche de QTLs

La majorité des caractères phénotypiques, comme par exemple le poids d'un fruit, varie de façon quantitative. De tels caractères sont gouvernés par plusieurs gènes (système polygénique et multi-locus) et influencés plus ou moins par les effets de l'environnement. Les locus agissant collectivement sur l'expression quantitative du caractère sont appelés des QTLs pour : « Quantitative Trait Loci » (Geldermann, 1975).

#### Principe et méthode de recherche de QTLs

Si les marqueurs sont suffisamment nombreux et correctement répartis sur le génome, chaque QTL sera encadré par des marqueurs relativement proches. Les recombinaisons entre les QTLs et les marqueurs bordants sont des événements d'autant plus rares que la liaison QTL-marqueur est forte. La clef essentielle de la détection des QTLs est donc de déterminer quels sont les marqueurs liés à une modification de la distribution du caractère ou de rechercher systématiquement les déséquilibres de liaison (Thoday, 1961). Pour cela, la cartographie de QTLs requiert une descendance cartographiée en ségrégation, où le caractère quantitatif a été mesuré pour chacun des individus. Les premiers QTLs cartographiés chez les végétaux l'ont été chez le maïs et la tomate sur des descendance F2 ou BC dérivant de

**Figure 2.3.2.** : Valeur seuil du *LOD score* (ordonnée) en fonction de l'espacement moyen entre les marqueurs (abscisse) pour différentes tailles du génome (exprimées en nombre de chromosomes). D'après Lander et Botstein (1989).



lignées pures (Tanksley *et al.*, 1982 ; Osborn *et al.*, 1981 ; Nienhuis *et al.*, 1981 ; Paterson *et al.*, 1988 ; Edwards *et al.*, 1987, 1992, Stuber *et al.*, 1992).

Ces méthodes utilisées peuvent être classées en deux catégories : i) celles qui s'appuient sur les marqueurs pris individuellement ; c'est la recherche de QTLs par analyse de variance et ii) celles qui prennent en compte conjointement deux ou plusieurs marqueurs ; c'est la recherche de QTLs par « interval mapping ».

L'utilisation de l'analyse de variance à un facteur conduit à un test F qui, lorsqu'il est significatif, suggère la présence d'une liaison génétique entre le marqueur et le caractère. Cette méthode négligeant l'effet des autres locus, il est conseillé d'utiliser l'analyse de covariance pour améliorer la puissance du test.

L'*interval mapping*, qui tient compte de l'information des marqueurs encadrant le QTL, est actuellement la méthode la plus utilisée. A chaque intervalle entre deux marqueurs, la probabilité de présence d'un QTL est calculée par la méthode du maximum de vraisemblance (Lander et Botstein, 1989). C'est une méthode puissante de détection, mais qui perd son pouvoir discriminant lorsque l'intervalle entre les marqueurs est supérieur à 20 cM (Darvasi *et al.*, 1993). Lander et Botstein (1989) ont déterminé par simulation une série de *LOD score* seuils en fonction du nombre de chromosomes et de l'espacement moyen entre marqueurs RFLP (Fig. 2.3.2.). Ils proposent, par exemple, un *LOD score* seuil de 2,4 pour un génome de 1000 cM avec un marqueur tous les 15 cM. Bien entendu, le choix de ces critères dépend de l'objectif de l'étude : une première évaluation probable de QTLs ou l'initiation d'un clonage de gènes.

### Principaux facteurs influençant les QTLs

Deux facteurs principaux concourent à influencer la détection des QTLs. Le premier, indépendant du fonds génétique correspond à la taille de la population et à la densité en marqueurs positionnés sur la carte génétique. L'efficacité de détection et la précision de la localisation des QTLs dépendent plus de l'effectif de la population que de la densité des marqueurs, lorsque celle-ci atteint un espacement moyen de 20 à 25 cM entre marqueurs (Darvasi *et al.*, 1993). Le second est directement lié au fonds génétique et correspond aux effets d'épistasie.

L'effet du milieu est défini en fonction du comportement d'un génotype dans un environnement donné. Ce dernier peut avoir une incidence importante sur la présence et la localisation des QTLs. L'exemple type est celui de l'étude de Paterson *et al.* (1991), où un

total de 29 QTLs a été cartographié dans trois environnements différents ; seuls quatre s'expriment dans les trois environnements (QTLs dits « généralistes »), 10 s'expriment dans deux et 15 s'expriment dans un seul (QTLs dits « spécialistes »). Cette expérimentation suggère que des dispositifs pluriannuels et multilocaux seront particulièrement adaptés à l'étude fine d'un caractère et à la recherche des QTLs le gouvernant.

Les effets d'épistasie correspondent aux interactions existant entre gènes et influençant l'expression du phénotype, tous les gènes n'agissant pas uniquement selon le mode additif. Les dispositifs expérimentaux constituent fréquemment une limite pour déterminer les effets d'épistasie, du fait des petits effectifs étudiés (Paterson *et al.*, 1991) et du nombre réduit de marqueurs (Edwards *et al.*, 1987). Généralement, il est possible d'étudier les effets d'épistasie d'ordre 1 (entre 2 QTLs) : par exemple pour la tolérance au sel chez la tomate (Breto *et al.*, 1994), l'architecture du maïs (Doebley *et al.*, 1995) ou encore les composantes de la résistance à *Phytophthora capsici* chez le piment (Lefèbvre et Palloix, 1996). Mais, au-delà de l'ordre 2 (3 QTLs), il est nécessaire de mettre en place des dispositifs expérimentaux lourds. A côté de ces interactions gène-gène, le fonds génétique peut aussi avoir un effet prononcé sur les caractères à effet quantitatif. Comme l'ont montré Doebley *et al.* (1995), les effets de deux QTLs de domestication du maïs sont beaucoup plus faibles lorsqu'ils ségrègent dans un fonds génétique « téosinte » que dans un fonds F2 maïs x téosinte.

#### **2. 3. 4. En résumé**

Le développement des marqueurs moléculaires en cartographie, et en particulier ceux générés en grand nombre par PCR, constitue un outil à la fois rapide et puissant pour des études très variées.

Chez les caféiers, la cartographie génétique et la recherche de QTLs de caractères d'intérêt agronomique, phénologique et/ ou biochimique au sein d'un croisement interspécifique sont en cours de développement. Cette étude concernant le croisement pionnier PSE x DEW est un préambule au programme d'amélioration des caféiers développé à l'IRD, en particulier pour la réalisation de sélection précoce.

Bien que de nombreuses combinaisons interspécifiques de caféiers soient réalisées (Louarn, 1992), aucune carte interspécifique n'a été obtenue avant cette étude (ceci est directement lié au temps long de génération des descendances de cartographie chez les arbres). Cependant, il ne faut pas perdre de vue le danger d'établir des programmes de sélection à partir des résultats d'un seul croisement. C'est pourquoi de nombreuses

combinaisons interspécifiques sont en cours d'analyse comme : *C. heterocalyx* x *C. canephora* et *C. pseudozanguebariae* x *C. canephora*.

## **Chapitre 3 : Matériel et méthodes**

---

## Chapitre 3 : Matériel et méthodes

Ce chapitre décrit l'ensemble du matériel végétal maintenu dans les stations de recherche IRD de Man et du CNRA à Divo (Côte d'Ivoire), ainsi que les méthodes d'acquisition des données (extraction et analyse par HPLC des composés biochimiques, extraction et dosage d'ADN génomique, marquage AFLP) et les méthodes d'analyse des données (étude de variabilité et de génétique quantitative des composé, cartographie et recherche de QTLs).

*N.B.* : Les solutions courantes sont référencées en annexe 3.1.

### 3. 1. Matériel végétal

#### 3. 1. 1. Les espèces cultivées

De nombreuses accessions naturelles des deux espèces *C. arabica* et *C. canephora* / existe<sup>nt</sup> au côté des cultivars. Ces accessions spontanées ont été considérées dans l'étude de diversité biochimique.

#### L'espèce *C. arabica*

*Coffea arabica* est originaire des forêts humides de montagne du sud-ouest de l'Ethiopie et de deux régions voisines, le Plateau de Boma au Soudan et le Mont Marsabit au Kenya. Sa propagation dans le monde s'est faite à partir d'un nombre limité d'origines génétiques, et les variétés agronomiques actuelles résultent de quelques mutants de ces origines et de croisements.

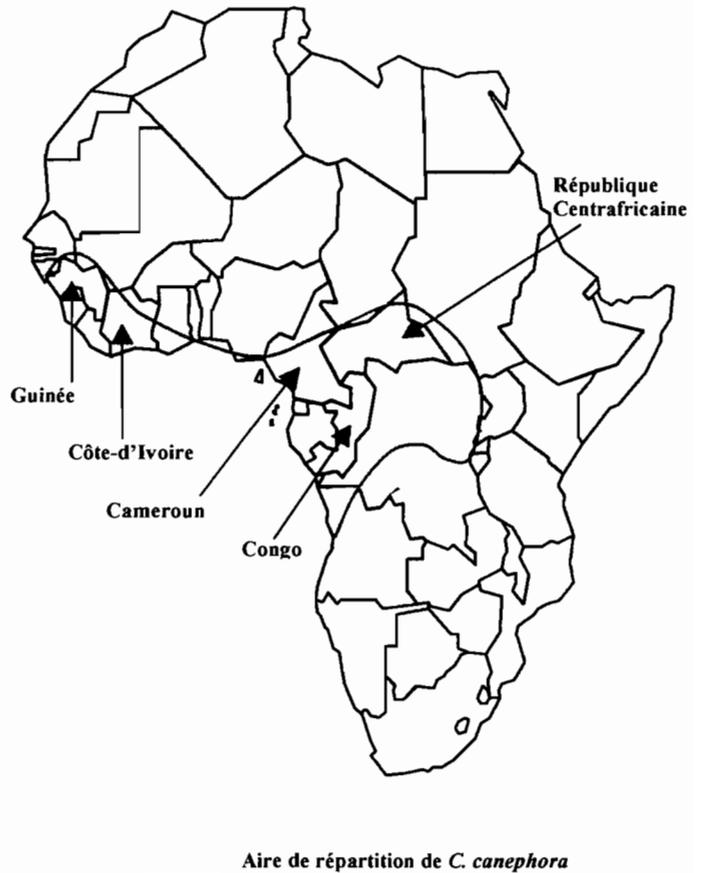
La collection correspond essentiellement à des descendances d'arbres spontanés et sub-spontanés issus d'une prospection ORSTOM réalisée en Ethiopie (Guillaumet et Hallé, 1967).

**Tableau 3.1.** : Origines spontanées et sub-spontanées éthiopiennes (36) et kenyanes (2) de l'espèce *Coffea arabica*. Ces génotypes correspondent à une partie de la collection maintenue au Mont Tonkoui, à Man (Côte d'Ivoire).

Numéro d'origine	Provenance	Altitude (m.)
1 ; 2	Vallée du Gojeb	1420
3	Bonga	1730
6 ; 7 ; 8	Wush-Wush	1860
9 ; 10	Bonga	1600
13	Limu	1660
14	Limu-Gimma	1530
15 ; 17	Gimma-Goré	1350
20 ; 21 ; 22 ; 23	Tippi	1300
35b	Mizan-Teferi	1500
35t ; 35q	Mizan-Teferi	1465
36b	Kollo	1460
37 ; 38	Shasha	1500/ 2000
39	Decchia	2000
40	Wush-Wush	2050
42 ; 43 ; 44	Decchia	2040
46	Decchia-Bonga	2000
48 ; 49 ; 52 ; 53 ; 56 ; 58 ; 59 ; 61	Bonga	1700
4 ; 7	Mont Marsabit (Kenya)	1500

**Tableau 3.2.** : Origines spontanées et sub-spontanées (38) de l'espèce *Coffea canephora* collectées dans cinq pays africains. Ces génotypes correspondent à une partie de la collection maintenue à la station de Divo (Côte d'Ivoire). Les lettres A à E correspondent aux groupes génétiques révélés par les marqueurs RFLP (Dussert *et al.*, 1999).

Génotype	Pays d'origine
2650 E 2651 E 2663 A 2664 C 2667 C 2678 C 2683 C 2685 C	Cameroun
2009 D 2105 D 2121 D 2128 D 2321 E 2350 D 2358 D 2395 D	Côte d'Ivoire
2722 A 2723 A 2725 E 2727 E 2729 E 2730 E 2739 E 2740 B	Congo
2815 2824 2844 2862 2867 2886 2891 2918	Guinée
2501 B 2504 E 2511 B 2516 E 2518 B 2555 B	République Centrafricaine



Introduites sous forme de graines, les accessions sont actuellement en collection au Mont Tonkoui en Côte d'Ivoire.

Trente huit origines ont été retenues pour les différentes évaluations biochimiques (Tab. 3.1.).

#### L'espèce *C. canephora*

*Coffea canephora* possède une aire de répartition très large : située dans le continent Africain, elle s'étend, d'ouest en est, de la Guinée au Soudan, et, du nord au sud, du Cameroun en Angola. Des missions de prospection ont été réalisées par l'ORSTOM, en collaboration avec le CIRAD, l'IPGRI et la FAO, de 1975 à 1987 dans cinq pays africains : la Côte d'Ivoire, la Guinée en Afrique de l'Ouest, le Congo, le Cameroun, et la République Centrafricaine. L'ensemble du matériel collecté a été rassemblé en Côte d'Ivoire (station de Divo). La diversité génétique a été évaluée par marqueurs RFLP (Dussert *et al.*, 1999), excepté pour les plants originaires de Guinée. Les lettres symbolisant les groupes génétiques ont été conservées et sont reportées dans le tableau 3.2.

Huit génotypes par pays ont été retenus pour les évaluations biochimiques, excepté 6 génotypes pour la République Centrafricaine (Tab. 3.2.).

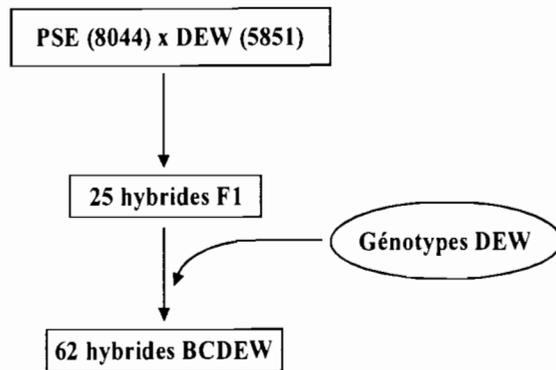
### **3. 1. 2. Le croisement interspécifique : PSE x DEW**

Les études d'héritabilité des teneurs en composés biochimiques, la cartographie génétique et la recherche des QTLs associés aux différents caractères d'intérêt agronomique ont été réalisées sur une descendance hybride issue d'un croisement entre deux espèces de caféiers hautement différenciés pour l'ensemble des caractères considérés : *C. liberica* var. *dewevrei* et *C. pseudozanguebariae*.

#### Les espèces parentales

DEW est une espèce anciennement cultivée et caractérisé par des arbres unicaules de 15 à 20 m de haut aux larges feuilles (15-30 cm x 5-15 cm). L'inflorescence est composée de plus de quatre fleurs ayant chacune cinq ou six pétales. La floraison a lieu six jours après la pluie déclencheuse et la maturation des fruits dure de 10 à 11 mois. Les fruits, de taille moyenne (15-20 mm x 13-16 mm), sont rouges à maturité. Son aire de répartition, très étendue, s'étend de la Guinée en Ouganda avec une extension en Angola (Berthaud et Guillaumet, 1978). En Centrafrique, il se situe plutôt dans des galeries forestières. Le

**Tableau 3.3.** : Schéma d'obtention et géotypes de la descendance de cartographie génétique. Les espèces parentales (12 géotypes par espèce) proviennent de Côte d'Ivoire et de la collection maintenue en serre à Montpellier (codes EB et H). Les 25 géotypes hybrides F1 (parcelle D16) et les 62 hybrides BCDEW (parcelle L13) sont codés suivant les numéros de ligne et de colonne de plantation.



Géotype						
DEW	PSE	F1	BCDEW			
5462	8044 a	23/26	1/03	9/21	19/06	
5495	8044 b	23/27	1/09	10/05	19/09	
5572	8044 c	23/28	2/02	10/08	19/13	
5585	8056	23/29	2/12	10/10	19/21	
5588	H 51	23/30	4/03	10/14	19/22	
5769	H 54	23/31	4/04	12/03	20/03	
5851	H 55	23/32	4/05	12/04	20/04	
5990	H 57	23/33	4/08	12/07	20/06	
EB 52	H 60	23/34	4/11	12/12	20/17	
EB 55	H 64	23/35	4/13	12/15	20/18	
EB 56	H 65	24/26	4/14	14/17	20/19	
EB 66	H 70	24/27	4/15	16/04	20/21	
		24/28	5/05	16/05		
		24/30	7/09	16/11		
		24/31	7/11	16/12		
		24/32	7/12	16/13		
		24/33	7/13	16/15		
		25/26	7/16	16/18		
		25/28	9/02	16/19		
		25/29	9/03	17/09		
		25/31	9/06	17/13		
		26/26	9/07	17/18		
		26/28	9/13	17/19		
		26/29	9/16	19/02		
		27/26	9/19	19/05		

**Tableau 3.4.** : Génotype correspondant à l'étude de génétique quantitative (Chapitre 5) pour les acides chlorogéniques (CGA), le saccharose (SAC) et la trigonelline (TRI).

		Génotype					
		PSE	BCPSE-1	BCPSE-2	F1	BCDEW	DEW
CGA	8013		L13 1/12 L13 5/18 L13 9/08 L13 12/09 L13 14/02		D16 23/26 D16 23/29 D16 23/30 D16 23/32 D16 24/27	L13 1/04 L13 1/09 L13 1/10 L13 2/02 L13 2/12	5303
	8038		L13 14/03 L13 17/21			D16 24/30 D16 25/29	L13 4/04 L13 4/13
SAC	8044		L13 1/12 L13 5/18 L13 7/21	L22 15/10 L22 14/08 L22 20/02			5372
	8053		L13 9/08 L13 12/09 L13 14/03 L13 17/21	L22 3/21 L22 12/10 L22 18/06 L22 19/06	D16 23/29 D16 23/32 D16 23/34 D16 24/26 D16 25/26	L13 1/04 L13 1/09 L13 2/02 L13 2/12	5390
	8056				D16 25/26 D16 25/31 D16 25/35	L13 4/04 L13 4/13 L13 4/15	5751
TRI	8066		L13 1/12 L13 9/08 L13 12/09 L13 14/03 L13 17/21 L13 20/14				5801
	8107						



Photo 3 : Floraison de l'espèce sauvage *Coffea pseudozanguebariae* Bridson.

génotype 5851 ayant servi à la génération des hybrides F1 a été prospecté en Centrafrique près de Siolo Nvuna (Berthaud et Guillaumet, 1978).

PSE est une espèce sauvage arbustive (1,5-3,5 m) à petites feuilles (7,8-10,2 cm x 2,9-4,1 cm) et à fleurs éparses (Photo 3), dont les pétales à tube peu évasé sont au nombre de 5 à 8. Les fruits sont petits (8,3-11,3 mm x 5,7-8,4 mm), allongés ou ovoïdes et violets foncés à maturité (Hamon *et al.*, 1984). La floraison a lieu 7 à 8 jours après la pluie déclencheuse et la maturation dure de 2,5 à 3,5 mois. Cette espèce a été trouvée dans les zones côtières tanzaniennes et kenyanes (Anthony *et al.*, 1987 ; Berthaud *et al.*, 1980 ; Berthaud *et al.*, 1983). Le génotype 8044 utilisé pour l'obtention des hybrides F1 a été prospecté au Kenya à Diani (Berthaud *et al.*, 1980).

#### Les hybrides de première et seconde génération

Le croisement contrôlé entre PSE (parent femelle) et DEW (parent mâle) a été réalisé par Louarn en 1980 à Man en Côte-d'Ivoire (Schéma, tableau 3.3.). Du pollen du génotype 5851 (DEW) a été apporté manuellement sur le génotype 8044 (PSE).

Vingt cinq hybrides F1 ont été créés par pollinisation contrôlée et plantés en essai en 1981 sur une parcelle isolée (parcelle D16). Les hybrides de seconde génération ont été obtenus par pollinisation libre. Néanmoins, les travaux de Barre (1997) ont montré, grâce aux marqueurs moléculaires, que ces hybrides sont issus de la pollinisation des F1 par, en majorité, des arbres DEW (hybrides BCDEW) et, en minorité, des arbres PSE (hybrides BCPSE-1). Ces hybrides BCDEW et BCPSE-1 sont implantés sur la parcelle L13. Enfin, des hybrides BCPSE-2 issus de croisements contrôlés (croisement sous manchon) ont été réalisés et sont plantés sur la parcelle L 22.

#### Le matériel utilisé dans notre étude

L'étude de l'additivité (Chap. 5) utilise un modèle équilibré avec sept arbres dans chacun des groupes suivant : DEW, BCDEW, F1, BCPSE et PSE. Les génotypes considérés diffèrent d'un composé à un autre (Tab. 3.4.). L'origine géographique des espèces parentales figure en annexe 3.2.

L'importance relative des effets génétiques, de l'environnement (variation inter-annuelle) et de l'interaction est estimée pour les années 1995 et 1997 avec : 7 hybrides pour le saccharose et les alcaloïdes et 10 génotypes pour les acides chlorogéniques.

Pour la cartographie génétique et la recherche des QTLs, soixante-deux hybrides BCDEW ont été retenus. Les 25 hybrides F1 ont aussi été considérés, ainsi qu'un échantillon de 12 DEW et de 12 PSE, afin de sélectionner les bandes spécifiques de PSE pour la cartographie (Tab. 3.3.).

### **3. 2. Les méthodes d'acquisition des données**

#### **3. 2. 1. Extraction, séparation et analyse des composés biochimiques**

##### Préparation des échantillons

L'extraction des différents composés biochimiques a lieu sur les grains verts de café. Ces derniers sont issus du dépulpage par voie humide des drupes de café à maturité. Ils subissent deux séchages successifs : i) en Côte d'Ivoire sous le soleil (jusqu'à une teneur en eau d'environ 10 %) et ii) au laboratoire dans un dessiccateur contenant du silicagel (teneur en eau finale d'environ 5%).

Pour chaque génotype analysé, un lot de 50 graines est réduit en poudre à l'aide d'un broyeur à bille (Dangoumill) durant 2 minutes. La poudre obtenue sert : i) à déterminer la teneur des différents composés biochimiques (5 fractions) et ii) à estimer la teneur en eau (trois fractions pesées avant et après une déshydratation à 100°C dans une étuve durant 15 heures). Les différentes solutions obtenues après extraction sont filtrées à 0,2 µm (Millipore) et conservées à -20°C jusqu'à l'analyse en HPLC.

##### Extractions des composés biochimiques

###### *Les acides chlorogéniques*

Cinq méthodes différentes d'extraction ont été préalablement comparées (Ky *et al.*, 1997) (voir article annexe A). La méthode retenue est la suivante : l'extraction est effectuée à partir de 50 mg de poudre. Chaque fraction est placée dans un tube contenant 50 ml de méthanol à 70 % et 0,5 % de bisulfite de sodium (agent antioxydant), durant 15 heures, sous agitation (table d'agitation, 125 rpm) à froid (4°C) et à l'obscurité. L'extrait aqueux est purifié par 50 µl de chacun des réactifs de Carrez A (21,9 g d'acétate de zinc dihydraté mélangé à 3 g d'acide acétique glacial dans 100 ml) et B (10,6 g de ferrocyanide de potassium trihydraté dans 100 ml).

*La trigonelline et la caféine*

La caféine et la trigonelline étant deux alcaloïdes hydrophiles, leur extraction se fait dans une solution d'extraction alcaline : 50 mg de poudre de café vert sont mélangés à 500 mg d'oxyde de magnésium (MgO) dans 25 ml d'eau dans un tube à essai de 50 ml clos par un capuchon. Le mélange d'extraction est maintenu à 110°C (catalyseur thermique de la réaction d'extraction) pendant 30 minutes, sous pression en autoclave (Barre *et al.*, 1998a). Après refroidissement, le mélange est dilué deux fois par adjonction d'eau filtrée.

*Le saccharose*

L'extraction du saccharose s'effectue dans un tube de 50 ml capuchonné contenant 50 mg de poudre de café vert et 25 ml d'eau distillée, à 60°C dans un bain-marie durant 15 minutes. Les tubes sont agités manuellement toutes les cinq minutes. Les tubes refroidis, les matières colloïdales des extraits sont précipitées par adjonction de 50 µl de chacun des réactifs de Carrez.

Séparation et analyse par chromatographie liquide haute pression (HPLC) des acides chlorogéniques et des alcaloïdes.*Principe de l'HPLC en phase inverse*

La séparation et l'analyse des isomères d'acides chlorogéniques, de la caféine et de la trigonelline s'effectue par HPLC en phase inverse. Cette technique est basée sur la différence de solubilité des composés biochimiques entre deux phases : la phase mobile polaire (mélange d'éluants) et la phase stationnaire (silice poreuse sur laquelle sont greffées de longues chaînes carbonées en C<sub>18</sub>). L'extrait contenant les composés biochimiques est injecté dans la phase mobile polaire, passe à travers la phase solide et la rétention de chaque composé à l'intérieur de la colonne dépend de son adsorption sur la phase stationnaire apolaire. L'HPLC en phase inverse permet d'améliorer la vitesse et la sensibilité de la méthode. Le passage de l'échantillon sous de fortes pressions raccourcit le temps de séparation des composés et améliore la résolution de cette technique d'analyse.

*Montages chromatographiques*

Le montage de l'appareillage HPLC est constitué d'un ordinateur : NEC Power Mate 486/ 33i, avec comme logiciel : *Millenium 2010 Chromatography manager* (option PDA), de 2 pompes Waters 510, d'un injecteur automatique d'échantillon (Waters<sup>TM</sup> 717 plus

*autosampler*), d'une lampe UV et d'un détecteur à barrette de diode (Waters™ 996 *Photodiode array detector*), d'une précolonne C<sub>18</sub> et d'une colonne C<sub>18</sub> (250 x 4 mm ; Merck *Superspher 100 RP 18*, taille des particules : 5 µm). Tous les solvants utilisés sont de qualité analytique et dégazés par ultrason (Ney, 300 Ultrasonik) pendant 20 minutes.

#### *Analyses des acides chlorogéniques*

Le programme d'éluion utilise deux solvants. Le solvant A est une solution d'acide phosphorique à 2 mM (275 µl d'acide phosphorique 85,7 % pour 2,5 l d'eau distillée). Le solvant B est du méthanol. Les échantillons (volume d'injection de 10µl) sont analysés à température ambiante selon le programme d'éluion suivant : un mélange des solvants A-B dans la proportion initiale 75-25, vers une solution pure de B, suivant un gradient linéaire en 45 minutes. Le palier à 100% de solvant B permet de nettoyer la colonne après chaque injection. Le débit des solvants est fixé à 0,8 ml/ min. La détection UV des CGA s'effectue à une longueur d'onde de 325 nm correspondant à leur maximum d'absorption. Tous les 40 échantillons, la précolonne C18 est renouvelée, et la colonne 100 RP 18 est nettoyée durant une nuit avec du méthanol pur à un débit de 0,2 ml/ min. Tous les dix passages, l'injection d'un échantillon témoin de 5-CQA (Ref. D11.080-9, Aldrich-Chimie) est effectuée, afin de se rendre compte d'une éventuelle évolution des mesures au cours du temps. Les mesures ont été faites en randomisation totale.

Les neuf isomères principaux constituant les acides chlorogéniques ont été identifiés par comparaison des chromatogrammes lus à 280 nm et à 325 nm (maximum d'absorption), des temps de rétention et des spectres UV avec les résultats obtenus par Rakotomalala (1993). L'identification des trois isomères de la classe des CQA a été confirmée par isomérisation du 5-CQA, selon le protocole de Trugo et Macrae (1984). L'expérience consiste à dissoudre 100 mg de 5-CQA dans 10 ml d'eau distillée (ajustement du pH de la solution à 8,0 avec une solution d'ammoniac diluée), à porter la solution à 100°C dans un bain bouillant (30 min), puis à ajuster le pH à 2,5 - 3,0 avec une solution d'acide chlorhydrique diluée. Cette solution finale est ensuite utilisée en HPLC.

La détermination des concentrations (en mg/ l) des différents isomères est rapportée à celle d'une gamme étalon du standard 5-CQA. La teneur des différents isomères (comme celle des autres composés) est exprimée en pourcentage de matière sèche des grains verts (% MS).

### *Analyses de la caféine et de la trigonelline*

Le programme d'élution utilise deux solvants. Le solvant A est une solution aqueuse (pH = 5,3) contenant du triéthylamine (2 ml pour 2 l de solution) et de l'acide acétique glacial (1ml pour 2 l de solution). Le solvant B est du méthanol. Les échantillons (volume d'injection de 10 µl) sont analysés à température ambiante selon le programme d'élution suivant : un mélange des solvants A-B dans la proportion 70-30 maintenu constant par gradient isochratique durant 15 min. Le palier à 100% de solvant B permet de nettoyer la colonne après chaque injection. Le débit des solvants est fixé à 1,0 ml/ min. Les détections UV de la caféine et de la trigonelline se font à une longueur d'onde respective de 272,8 nm et 263,3 nm.

L'identification et la détermination des concentrations en caféine et trigonelline ont été réalisées par comparaison à la gamme étalon correspondant aux standards purs du commerce (Sigma).

### Séparation et analyse du saccharose par couplage : HPLC/ ampérométrie pulsée.

Le saccharose est analysé par une technique de couplage entre la chromatographie liquide d'échange d'ions (à pH compris entre 12 et 14) et de la détection ampérométrique pulsée. Cette technique est simple d'emploi, ne nécessite aucun pré-traitement de l'échantillon et est sensible.

La séparation par chromatographie échange d'ions met à profit l'exaltation à pH basique de la fonction légèrement acide des sucres. Ceux-ci vont donc se comporter comme des acides faibles qui vont se séparer par échange d'anions en fonction de leur pKa. Les résines utilisées (Dionex) sont constituées de sphère de polystyrène-divinyl-benzène de 10 µm de diamètre, sur lesquelles sont fixées des microbilles de latex synthétiques. L'échange d'ions se réalise à la périphérie de ces microbilles, porteuses de groupements ammoniums quaternaires. La colonne utilisée est une colonne de 4 x 250 mm (Carbopac PA1, Dionex). Afin de pouvoir résister à des pH élevés, l'intégralité du système analytique (pompe, injecteur automatique, tubulures) est construit en matériaux synthétiques. La pompe GP 40 sert à propulser les deux éluants : éluant A composé d'eau distillée filtrée à 0.2 µm et éluant B de soude à 150 mM, selon le gradient isochratique suivant : 25 % de A et 75 % de B. Le débit des éluants est fixé à 1,0 ml/min.

La détection par ampérométrie pulsée met à profit la possibilité des glucides d'être facilement oxydés. A un potentiel précis, les fonctions alcools secondaires –CHOH des glucides sont oxydées sur une électrode d'or. Il en résulte un courant proportionnel à la quantité de –CHOH présents. Suite à cette oxydation, des résidus viennent se déposer sur l'électrode d'or qui se salit rapidement et perd toute sensibilité. Afin de remédier à cet inconvénient, l'électrode est automatiquement et successivement soumise à trois potentiels : i)  $E1 = +50\text{mV}$  (300 ms): oxydation des –CHOH ; ii)  $E2 = + 60 \text{ mV}$  (300 ms) : oxydation totale des sous-produits organiques ; et iii)  $E3 = - 0,8 \text{ V}$  (480 ms) : l'électrode retrouve son état d'origine. L'ampérométrie pulsée présente les avantages suivants : i) spécificité (seuls les groupements alcools secondaires des sucres répondent à ce potentiel), ii) sensibilité (concentration de quelques picomoles détectable) et iii) détection non destructive.

### **3. 2. 2. Extraction et dosage de l'ADN**

#### Principe

La méthode d'extraction d'ADN utilisée repose sur une extraction des noyaux, suivie de leur lyse. Les feuilles sont finement broyées et la poudre mise en suspension dans un tampon d'extraction. Celui-ci a pour objet de libérer et de nettoyer les noyaux (Triton-X 100), de maintenir leur structure grâce à la pression osmotique (KCl, NaCl), d'inhiber les nucléases (EDTA), d'éviter une oxydation ( $\beta$ -mercaptoéthanol) et de maintenir un pH constant (Tris). Les noyaux sont isolés par centrifugation, puis lysés dans un tampon contenant du MATAB. Les protéines et les composés hydrophobes sont éliminés dans du chloroforme isoamylique (24/ 1). L'ADN peut alors être précipité et repris dans un petit volume de Tris-EDTA (TE).

#### Protocole d'extraction d'ADN

Les feuilles proviennent de Man et sont plongées dans l'azote liquide à leur arrivée à Montpellier. Elles sont ensuite conservées à  $- 80^{\circ}\text{C}$  dans l'attente de leur lyophilisation pendant au moins 72 h. Elles sont ensuite stockées, avant broyage, dans une chambre froide ( $4^{\circ}\text{C}$ , humidité relative de 18 %). Aucune altération n'a été observée après plus d'un an de conservation.

Les feuilles sont réduites en poudre à l'aide d'un broyeur à bille "Dangoumill" (Prolabo) durant 3 minutes. Un gramme et demi de poudre est pesé dans des Erlenmeyers de 250 ml. Cent millilitres de tampon de Dolezel sont utilisés pour l'extraction (15 mM Tris,

2 mM EDTA, 80 mM KCl, 20 mM NaCl, 0,1 %  $\beta$ -mercaptoéthanol, 0,5 % Triton-X 100, pH 9) (Dolezel, 1991 ; Dolezel *et al.*, 1989, 1992). La solution est homogénéisée à l'aide d'un agitateur magnétique, puis agitée pendant 2 h à température ambiante sur table agitante (25 rpm). La suspension de noyaux est transvasée dans deux tubes de 50 ml, puis centrifugée à 4000 g pendant 20 min. Les culots sont rassemblés dans 20 ml de tampon de lyse (4 % MATAB, 0,1 M Tris, 1,25 M NaCl, 0,02 M EDTA, pH 8). Il est possible à ce stade de congeler les noyaux à -20°C.

La lyse des noyaux est réalisée dans un bain-marie à 65°C pendant 4 h, avec une agitation manuelle des tubes toutes les heures. La purification consiste à ajouter un volume de 20 ml de chloroforme isoamylique (24/1) à la solution. Après agitation par retournement, les tubes sont centrifugés à 4000 g pendant 10 min et la phase aqueuse est récupérée. Cette étape est réalisée deux fois. Après quoi, l'extrait aqueux est soumis à l'action de 100  $\mu$ l de RNase A (10 mg/ml, Boehringer) durant 30 minutes à 37°C (étuve). L'ADN est ensuite précipité avec 20 ml d'isopropanol, puis un culot est obtenu après 10 min de centrifugation à 3000 g. Le culot est repris dans 0,7 ml de TE, puis transféré dans un tube Eppendorf de 2 ml. L'ADN est alors précipité avec 0,6 ml d'isopropanol et 70  $\mu$ l d'acétate de sodium (3 M, pH 5,2). Un culot est à nouveau obtenu après centrifugation à 12000 g pendant 10 min. Il est nettoyé avec de l'éthanol 70 % et séché pendant 30 min au " Speed-Vac ", puis dissout dans 200  $\mu$ l de TE.

### Dosage de l'ADN

Le dosage de l'ADN est réalisé sur mini-gel d'agarose (1%) dans du tampon TAE 1X sous une migration à 50V. Une quantité de 0,4 g d'agarose est pesée dans un Erlenmeyer contenant 40 g (environ 40 ml) de tampon TAE 1X. Le mélange est porté à ébullition durant 1 min. au four à micro-onde. Le gel légèrement refroidi est coulé aussitôt dans la cuve de migration munie d'un peigne de dépôt. La solution d'ADN à doser est diluée à 1/50 et 1/100 et le volume du dépôt d'environ 10  $\mu$ l. La détermination des concentrations s'effectue pour chaque génotype par comparaison avec les deux premières bandes du marqueur lambda HindIII/ EcoRI pour les dilutions  $\lambda$ 1 : 200 ng et 100 ng et  $\lambda$ 2 : 100 ng et 50 ng. Les bandes sont visibles sous UV après coloration du gel au BET durant environ 15 min. L'étape du dosage permet aussi de se rendre compte de la qualité de l'ADN : bonne migration, pas dégradé. L'ADN est ensuite conservé à -20°C. Les rendements de l'extraction varient de 250 à 1500 ng par g de poudre sèche selon les types d'hybride.

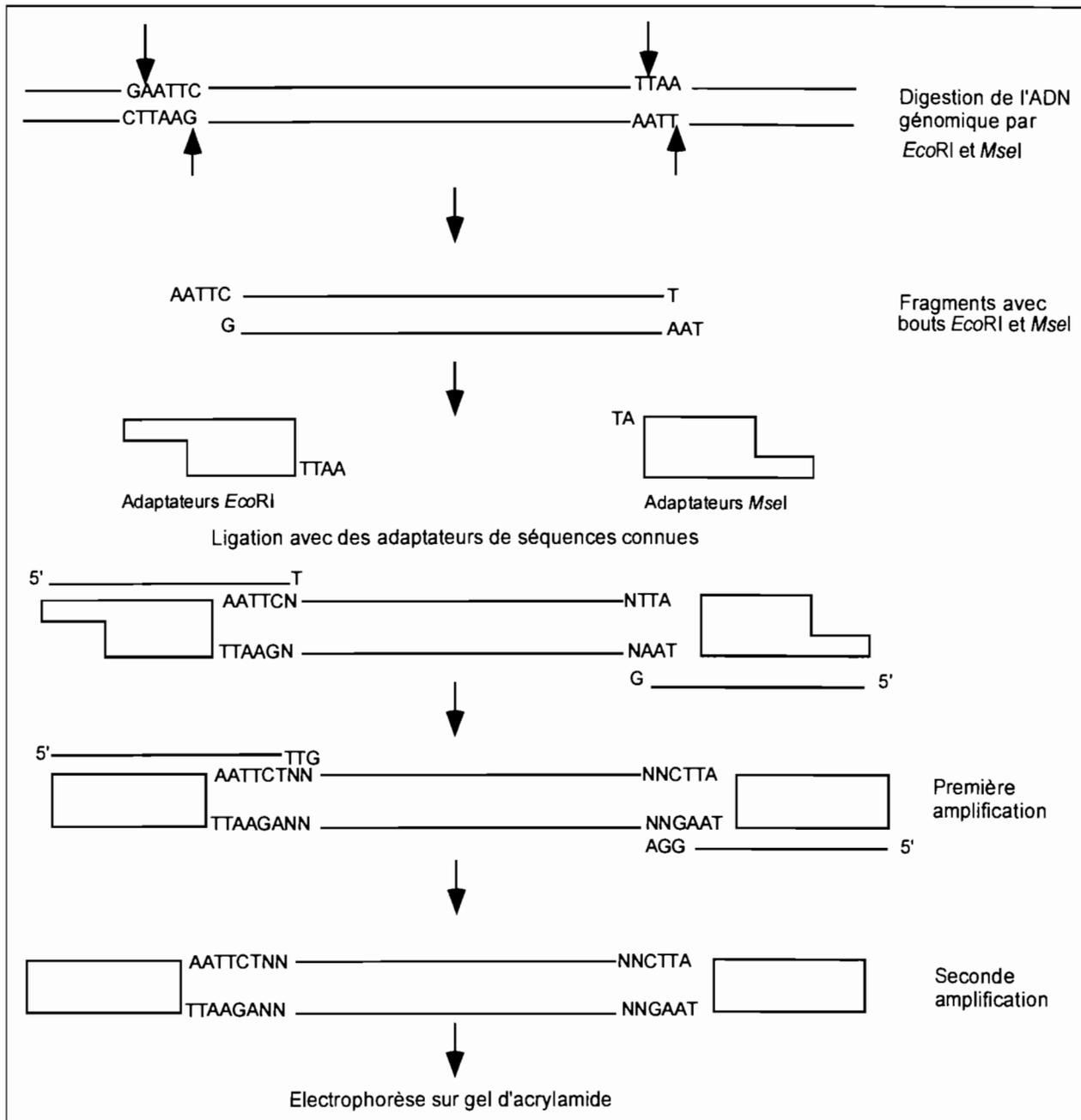
### 3. 2. 3. Technique de marquage par AFLP

#### Principe

Les marqueurs AFLP (Amplified Fragment Length Polymorphism) sont générés par l'amplification sélective par PCR de fragments de restriction, issu d'une digestion génomique totale (Zabeau et Vos, 1993). Cette technique révèle à la fois un polymorphisme : i) de site de restriction et ii) des sites d'hybridation des bases arbitraires. Le choix de cette technique a été principalement guidé par le fait qu'elle permet d'obtenir un grand nombre de marqueurs en un temps relativement court. Le principe de la méthode est simple : i) l'ADN est digéré par deux enzymes de restriction (endonucléases), l'une à 6 bases *Eco* RI et l'autre à 4 bases *Mse* I (étape de digestion) ; ii) des adaptateurs *Eco* RI et *Mse* I de séquence connue d'une vingtaine de nucléotides et complémentaires des fragments digérés sont ensuite ajoutés aux extrémités (étape de ligation) ; iii) les fragments sont pré-amplifiés par PCR en utilisant comme amorce des oligonucléotides correspondant à la séquence des adaptateurs prolongés en 3' (vers l'intérieur du fragment à amplifier) par une base arbitraire ; iv) les fragments sont amplifiés, de nouveau par PCR, en utilisant les même amorces, mais prolongées en 3' par deux bases arbitraires supplémentaires. Seuls les fragments ayant des bases complémentaires de ces 6 (2 x 3) bases arbitraires vont donc être amplifiés. Les amorces utilisées et les conditions de PCR privilégient l'amplification des fragments *Eco*RI-*Mse*I au détriment des fragments *Eco*RI-*Eco*RI et *Mse*I-*Mse*I. Après électrophorèse en gel d'acrylamide, les fragments (jusqu'à une centaine) vont être révélés. Ils sont marqués soit radioactivement, soit chimiquement, afin de repérer leur signal sur film autoradiographique. Le schéma de la figure 3.1. illustre les différentes étapes de la technique AFLP.

#### Digestion de l'ADN génomique

La digestion est réalisée sur des solutions d'ADN génomique diluées à une concentration de 50 ng/  $\mu$ l. Deux enzymes de restriction sont utilisées : *Eco*RI et *Mse*I. La solution de digestion comprend pour un individu : 5  $\mu$ l d'ADN à 50 ng/  $\mu$ l, 2 unités de chacun des enzymes, 5  $\mu$ l de tampon 5 X spécifique de l'enzyme T4 ligase et de l'eau de qualité AFLP (volume final de 25  $\mu$ l). La réaction a lieu dans un tube Eppendorf de 0,5 ml et dure 2 heures à 37°C (dans une machine PCR). Les enzymes sont inactivées en incluant en fin de programme une période de 15 minutes à 70°C.



**Figure 3.1.** : Schéma illustrant le principe de la technique AFLP.

### Ligation des adaptateurs

L'objectif de cette étape est de lier les extrémités des fragments d'ADN digérés à deux adaptateurs spécifiques des sites de coupures *EcoRI* (adaptateur I) et *Mse I* (adaptateur II). La solution de ligation comprend pour un individu : les 25 µl d'ADN digérés, 1 µl de chacun des adaptateurs, 1 unité de T4 ligase, 5 µl de tampon 5 X correspondant à l'enzyme et 17 µl d'eau de qualité AFLP (volume final de 50 µl). La ligation dure trois heures à 37°C dans une machine PCR. En fin de programme, la ligase est thermiquement dénaturée à 60°C durant 10 minutes. La solution digérée/ liguée est ensuite diluée 10 fois avec du tampon TE.

### Amplifications

L'amplification sélective des fragments d'ADN muni de leurs adaptateurs se fait en deux étapes. La première étape de pré-amplification s'effectue, comme précédemment, dans des tubes de 0,5 ml. Par génotype, 5 µl d'ADN digéré et ligué, 0,5 µl de chacun des amorces +1 à 150 ng/ µl, 2 µl de dNTPs (5 mM), 5 µl de tampon 10X Proméga, 3 µl de MgCl<sub>2</sub> (25 mM), 0,2 µl de Taq Polymérase (5U/ µl), le tout ajusté avec de l'eau de qualité AFLP à 50 µl. Vingt cycles d'amplification par PCR sont effectués suivant la programmation : 30 sec. à 94°C, 1 min., à 56°C, 1 min. à 72°C. Les échantillons pré-amplifiés sont dilués au 30<sup>ème</sup>.

La seconde étape d'amplification nécessite : 5 µl d'ADN pré-amplifié, 2 µl de tampon 10X Proméga, 0,8 µl de dNTPs (5 mM), 1,2 µl de MgCl<sub>2</sub> (25 mM), 0,1 µl de Taq Polymérase (Proméga, Réf. : M 1865) (5 U/ µl), 0,5 µl de A (A est une solution contenant l'amorce *EcoRI* +3 marqué au gamma <sup>33</sup>P au niveau de l'ATP (voir le paragraphe suivant pour le protocole de marquage), 0,4 µl d'amorce *MseI* +3 (100 ng/ µl ) et de l'eau de qualité AFLP pour compléter à 20 µl. L'amplification suit le programme : 30 sec à 94°C, 30 sec à 65°C, 1 min à 72°C (13 cycles avec - 0,7°C par cycle), 30 sec à 94°C, 30 sec à 56°C, 1 min à 72°C (33 cycles). A la suite de cette amplification, 20 µl de bleu d'arrêt sont ajoutés aux échantillons.

### Marquage de l'amorce

Le marquage de l'amorce *EcoRI* +3 s'effectue à l'aide de l'ATP γ <sup>33</sup>P à 250 µCi (Amersham, Réf. : AH 9968). Dans un tube de 0,5 ml et pour 50 distributions (de cette solution appelée A), le mélange réactionnel est le suivant : 9 µl d'ammorce, 2,5 µl de tampon kinase 10 X, 1 µl de T4 polynucléotide kinase (Appligène, Réf. : 120104), 5 µl de <sup>33</sup>P et 7,5 µl d'eau AFLP. L'ensemble du mélange est placé dans une étuve à 37 °C pendant une heure. L'enzyme est ensuite dénaturé 10 min à 70 °C et le mélange conservé à 4 °C.

### Préparation des gels de migrations.

Les produits d'amplification sont analysés sur un gel de séquence dans des conditions dénaturantes. La composition du gel est la suivante : dans un Erlenmeyer à bec verseur, 40 g d'urée, 12 ml d'acrylamide-bis 40%, 8 ml de tampon TBE 10 X et 28 ml d'eau milliQ sont successivement ajoutés. La solution (volume total de 80 ml) est ensuite mélangée par agitation magnétique en chauffant très légèrement. Après dissolution totale de l'urée et au moment de couler le gel, 300 µl d'une solution d'ammonium persulfate à 10% (100 mg/ 1ml) et 30 µl de TEMED sont ajoutés. Le coulage du gel se fait sur un côté en maintenant le montage des plaques en biais de manière à éviter la création de bulle d'air. Un peigne sans dent est ensuite inséré.

### Préparation et montage des plaques de migration.

Le gel est coulé entre deux plaques de verre dont les faces internes ont été préalablement rincées à l'éthanol. La petite plaque subit un traitement complémentaire au "répulsif" (diméthylchlorosilane). Deux millilitres de ce "répulsif" sont soigneusement étalés sur la plaque à l'aide d'un "Kimwipes". Cette opération permettra en fin de migration des échantillons l'adhérence du gel à la grande plaque. Le montage des plaques se fait de la façon suivante : i) la grande plaque est placée à niveau, face interne vers le haut, ii) deux espaceurs de 0.4 mm d'épaisseur sont positionnés le long de la plaque, iii) la petite plaque est placée ensuite par-dessus la grande, face interne vers le bas et iv) quatre pinces à dessin maintiennent le montage, le temps d'appliquer le boudin en caoutchouc de maintien.

### Dépôt et migration des échantillons.

Après polymérisation complète du gel (durant au moins 45 minutes), le boudin de maintien est retiré délicatement. Le montage entier est rincé à l'eau chaude sous le robinet, afin de retirer les excédents de gel et les dépôts d'urée. Le tout est ensuite placé sur la cuve de migration, dont les compartiments du haut et du bas contiennent chacun 500 ml du tampon de migration (TBE 1X). Le peigne sans dent est retiré, et l'urée chassée à l'aide d'une seringue contenant du tampon. Un peigne à 60 dents est ensuite placé permettant d'aménager 62 espaces de dépôts. Le gel subit un préchauffage d'environ 15 minutes à 50 W. Parallèlement, les échantillons sont dénaturés. Le volume de dépôt des échantillons est de 5 µl. Afin d'éviter des défauts de migration liés au bordure du gel, du bleu est déposé dans le premier et dernier puit. Les échantillons sont déposés, les uns à la suite des autres et de gauche à droite, avec un

témoin interne afin d'orienter la lecture du gel après révélation. La migration s'effectue à 50W durant environ 2 h 30, correspondant à un front de migration du bleu ciel à environ 27 cm du haut des plaques.

#### Révélation du gel.

Après la migration, la petite plaque est retirée et le gel reste collé sur la grande plaque. Cette dernière est placée dans une cuve contenant la solution de fixation (un mélange d'acide acétique glacial, d'éthanol et d'eau distillée (10/ 20/ 70)). Après 30 minutes de fixation, le gel est transféré sur du papier Wattman n°3 (40 x 33 cm) par simple application du papier sur le gel. L'ensemble gel/ papier Wattman est entouré d'un film plastique alimentaire et séché sous vide à 80°C durant 30 minutes. En chambre noire sous lumière rouge, le film alimentaire est retiré. Le gel sec, toujours sur son support Wattman, est mis en contact d'un film autoradiographique Kodak (Biomax) dans une cassette d'exposition pendant 48 à 72 heures. Après l'exposition, le film est révélé selon les techniques habituelles : révélation, rinçage à l'eau, fixation, rinçage à l'eau et séchage à l'air.

### **3. 3. Les méthodes d'analyse des données**

#### **3. 3. 1. Etude de la diversité**

L'ensemble des analyses statistiques a été conduit en utilisant le logiciel statistique Statistica (version 5.1, 1997).

#### Structuration de la diversité

La structuration de la diversité a été étudiée par analyse en composantes principales (ACP) normées. Il s'agit de rechercher les axes principaux d'un nuage multidimensionnel afin de condenser l'information contenue par les N variables en un lot restreint de nouvelles variables synthétiques, les facteurs. Le but est de mettre en évidence des facteurs biologiques sous-jacents dont les effets sont à l'origine de la diversité observée (Noirot, 1990). L'option Varimax Normalisé du logiciel Statistica a été utilisée dans cette ACP normée. Cette option permet de rechercher les groupes indépendants de variables les plus corrélées entre elles.

### Autres méthodes

L'étude de la diversité au sein de chacune des collections *C. arabica* et *C. canephora* a été réalisée par analyse de variance à un critère de classification (modèle fixe). Des régressions linéaires ont aussi été utilisées.

La comparaison des variances entre les deux collections ont été effectuée en utilisant le test  $F = (\sigma_{\text{Max}}^2 / \sigma_{\text{Min}}^2)$ . Les variances étant inégales, le test non paramétrique de Kruskal et Wallis a été utilisé afin de comparer les moyennes entre les deux collections, avec le test de Newman et Keuls.

### **3. 3. 2. Approche quantitative du déterminisme génétique**

L'ensemble des analyses statistiques a été conduit en utilisant le logiciel statistique Statistica (version 5.1 1997).

#### Importance des effets génétiques et environnementaux

L'héritabilité des teneurs en composés biochimiques a été étudiée en utilisant un modèle croisé, mixte, équilibré et orthogonal d'analyse de variance à deux critères de classification : l'année de récolte (effet fixe) et le génotype (effet aléatoire). Les espérances mathématiques des carrés moyens sont :

$$E(CM_H) = \sigma^2 + pn \sigma_H^2, \text{ pour l'effet "génotype"}$$

$$E(CM_A) = \sigma^2 + n \sigma_{HA}^2 + qn \Sigma a_i^2 / (p-1), \text{ pour l'effet "année"}$$

$$E(CM_{HA}) = \sigma^2 + n \sigma_{HA}^2, \text{ pour l'interaction}$$

$$E(CM_R) = \sigma^2, \text{ pour la résiduelle}$$

avec  $n$  le nombre d'extraction par génotype,  $p$  le nombre de génotype par groupe et  $q$  le nombre d'année où a été effectué les mesures et où  $\sigma_H^2$ ,  $\sigma_{HA}^2$ ,  $\sigma^2$  et  $\Sigma a_i^2 / (p-1)$  représentent respectivement : la variance génotypique, la variance de l'interaction, la variance de la résiduelle, et la variance inter-annuelle (estimation fixe).

Pour les effets significatifs, les composantes des variances peuvent être estimées :

$$\sigma_H^2 = (CM_H - CM_R) / pn$$

$$\Sigma a_i^2 / (p-1) = (CM_A - CM_R) / qn$$

$$\sigma_{HA}^2 = (CM_{HA} - CM_R)/n$$

$$\sigma^2 = CM_R$$

L'importance relative (exprimée en pourcentage) de chacun des effets peut aussi être calculée :

$$\sigma_H^2 / (\sigma_H^2 + \sigma_{HA}^2 + \Sigma a_i^2 / (p-1) + \sigma^2), \text{ pour l'effet " génotype "}$$

$$\sigma_{HA}^2 / (\sigma_H^2 + \sigma_{HA}^2 + \Sigma a_i^2 / (p-1) + \sigma^2), \text{ pour l'effet interaction}$$

$$(\Sigma a_i^2 / (p-1)) / (\sigma_H^2 + \sigma_{HA}^2 + \Sigma a_i^2 / (p-1) + \sigma^2), \text{ pour l'effet " année "}$$

Ces équations correspondent aux coefficients de corrélation intra-classe de Snedecor.

### Différences inter-groupe et inter-arbres au sein des groupes

Les différences inter-groupes (PSE, BCPSE, F1, BCDEW et DEW) et inter-arbres (intra-groupe) ont été testées à l'aide du modèle mixte, équilibré et orthogonal d'analyse de variance hiérarchisée à deux critères de classification : les groupes (effet fixe) et les arbres (effet aléatoire) hiérarchisées aux groupes. Les groupes étant cultivés sur des parcelles différentes (excepté pour les groupes BCDEW et BCPSE-1), la mise en évidence de différences inter-groupes peuvent refléter en partie une différence de parcelle de culture.

### Additivité des teneurs en composés biochimiques

La méthode utilisée a pour objectif de supprimer en partie la confusion possible entre effet " parcelle " et effet " groupe ". L'additivité a été en effet testée en utilisant la régression linéaire : les parts de génome DEW attendues pour chacun des groupes sont de 0, 25, 50 et 100 %, respectivement pour PSE, BCPSE, F1, BCDEW et DEW. Avec l'hypothèse d'additivité, la moyenne attendue pour le caractère est proportionnelle à la part de génome DEW présente chez la plante.

Dans ces conditions, observer cinq points alignés ne peut pas s'expliquer par d'éventuelles différences de parcelles disposées en différents points de la station.

### **3. 3. 3. Cartographie de la descendance BCDEW**

Préalablement à la cartographie du croisement interspécifique, un fichier de simulation de cartographie contenant les principales caractéristiques de la descendance BCDEW (62 individus et présence de marqueurs distordus et non distordus) a été construit. Ce fichier de

simulation permettra de connaître la démarche et les valeurs de *LOD score* adaptées à la cartographie de la descendance issue du croisement PSE x DEW.

### Construction du fichier de simulation

Le fichier de simulation comprend 80 marqueurs répartis équitablement en huit groupes de marqueurs simulés et notés de GF-1 à GF-8. Leurs principales caractéristiques sont les suivantes :

- i) Les marqueurs de GF-1 à GF-4 correspondent à des marqueurs dominants qui suivent une ségrégation de type Mendélienne en backcross : 50 % de présence (codage en « 1 ») et 50 % d'absence (codage en « 0 »). Le premier marqueur de chacun des groupes a été simulé de façon aléatoire (loi de Bernoulli). Par la suite, chaque marqueur au sein des groupes est issu du précédent par inversion de deux codages (le « 1 » transformé en « 0 » et vice et versa) pour ~~10~~<sup>2/50</sup> % des individus. Chacun des groupes est par conséquent indépendant et les marqueurs d'un même groupe sont liés.
- ii) Les marqueurs des groupes GF-5, -6 et -7 sont distordus vers PSE (75 % de 1 et 25 % de 0). GF-5 est indépendant des 4 premiers groupes et ses marqueurs sont liés. La construction du groupe GF-6 diffère de GF-5 par le fait que le premier marqueur n'est pas obtenu par simulation. Il correspond en réalité au premier marqueur de GF-1 pour lequel nous avons transformé la moitié des « 0 » en « 1 ». Ainsi, le groupe GF-6 est un groupe de dix marqueurs liés distordus, devant théoriquement se rattacher au groupe GF-1. Les marqueurs de GF-7 sont indépendants entre eux, ainsi que de tous les autres marqueurs.
- iii) Les marqueurs du GF-8 sont indépendants entre eux et simulent la superposition de 2 bandes conduisant à un artéfact de distorsion de type 3 : 1. Ce codage a été effectué en tirant au hasard 10 couples de marqueurs au sein des groupes 1 à 4.

### Choix des bandes AFLP codées :

Seuls les marqueurs AFLP spécifiques de PSE seront retenus. Trente combinaisons d'amorces AFLP ont été utilisées pour la cartographie. Compte tenu de l'indisponibilité des génotypes parentaux chez PSE et DEW (parent PSE 8044 mort et intervention de géniteurs DEW dans la pollinisation non contrôlée des F1), les combinaisons ont été testées sur 12 génotypes représentatifs de chacune des espèces parentales, les 25 hybrides F1 et les 62 hybrides BCDEW. Les 12 génotypes PSE choisis constituent une bonne représentation de la

collection prospectée au Kenya en 1977 (Berthaud *et al.*, 1980). Les 12 génotypes DEW ont été retenus parmi ceux situés le plus proche de la parcelle des hybrides F1 (environ 1 km).

Les critères de choix des bandes polymorphes générées sont les suivants : les bandes PSE spécifiques retenues pour la cartographie doivent :

- i) être présentes chez les 12 génotypes PSE,
- ii) être absentes chez les 12 génotypes DEW,
- iii) être présentes chez les 25 hybrides F1,
- iv) ségréger dans la descendance composée de 62 hybrides BCDEW.

La double lecture des autoradiogrammes a été effectuée en adoptant le codage conventionnel ; présence de la bande PSE spécifique : « 1 » et absence de cette même bande : « 0 ».

#### Test des distorsions de ségrégation

La ségrégation des marqueurs a été comparée à la ségrégation mendélienne attendue pour des marqueurs dominants (ségrégation 1 : 1) par des tests de  $\chi^2$ .

#### Construction de la carte génétique

La cartographie du fichier de simulation et de la descendance BCDEW a été réalisée au moyen du logiciel MAPMAKER/ EXP version 3.0 b (Lander *et al.*, 1987, Lincoln *et al.*, 1992). La fonction de cartographie retenue est celle de Kosambi (Kosambi, 1944), qui prend en compte le phénomène d'interférence génétique lié au double recombinaison (Lorieux, 1993).

A la suite des résultats obtenus en cartographiant le fichier de simulation, la construction de la carte a été effectuée en trois étapes successives :

- i) Cartographie des marqueurs non-distordus à une valeur de *LOD score* de 5,5 pour une distance maximale entre locus de 30 cM. L'ordre des marqueurs a été confirmé par des tests multipoints (commande "ripple" du logiciel MAPMAKER). On obtient des groupes de liaison de marqueurs non distordus de forte vraisemblance.
- ii) Cartographie des marqueurs distordus suivant la même démarche que précédemment. On obtient des groupes de liaison de marqueurs distordus de forte vraisemblance.

iii) Agglomération des groupes de liaison de forte vraisemblance entre eux. Les jonctions ont été réalisées à des valeurs de *LOD score* > 3.

### 3. 3. 4. Recherche de QTLs

L'ensemble des variables servant à la recherche de QTLs est décrit dans l'annexe 3.3.

#### Identification des QTLs

Deux approches de mise en évidence des QTLs de ces caractères ont été adoptées :

- i) La recherche de QTLs en considérant les marqueurs individuellement par analyses de variance. Ces dernières sont réalisées afin de rechercher les associations significatives entre les variables quantitatives et les marqueurs intégrés dans les groupes de liaison de la carte. Ces analyses de variance sont réalisées à l'aide du logiciel Statistica.
- ii) La recherche de QTLs à partir de plusieurs marqueurs à l'aide du logiciel : Mapmaker/QTLs (Lincoln *et al.*, 1992). L'avantage de ce logiciel est la recherche de QTLs par cartographie d'intervalle composite (en anglais "composite interval mapping"). Cette méthode de détection part de l'hypothèse qu'il y a un QTL au plus dans l'intervalle entre deux marqueurs liés à un taux de recombinaison  $r$ . Le pourcentage de variance expliquée par un QTL ou par le modèle multi-QTLs a été calculé avec le même logiciel.

#### Variable quantitative simulée

Une variable quantitative simulée a été créée (à l'aide du logiciel Statistica). Cette variable non mesurée dans la descendance BCDEW, nous permettra d'évaluer les risques d'erreur dans l'identification des QTLs (Chap. 7).

**Chapitre 4 : Variabilité biochimique  
des espèces cultivées**

---

## Chapitre 4 : Variabilité biochimique des espèces cultivées

La variabilité biochimique (acides chlorogéniques, caféine, trigonelline et saccharose) a été évaluée à partir de 38 accessions spontanées de chacune des espèces *C. arabica* et *C. canephora*.

Cette étude a été entreprise pour :

- i) mettre en exergue les composés biochimiques impliqués dans le goût et qui différencient les deux espèces cultivées ;
- ii) évaluer la diversité présente à l'intérieur de ces deux espèces ;
- iii) analyser la structuration de cette variabilité en relation avec les voies connues de biosynthèse.

### 4. 1. Diversité biochimique chez l'espèce *C. arabica*

L'ensemble des composés analysés représente en moyenne 16 % MS. Cette quantité constitue la part extraite. Dans la suite de ce paragraphe, ces composés seront présentés les uns à la suite des autres par ordre d'importance.

#### 4. 1. 1. Le saccharose

Le saccharose est majoritaire avec 56 % de la part extraite. Des différences significatives existent entre accessions avec une teneur variant de 7,4 à 11,1 % MS (Tab. 4.1.).

#### 4. 1. 2. Les acides chlorogéniques

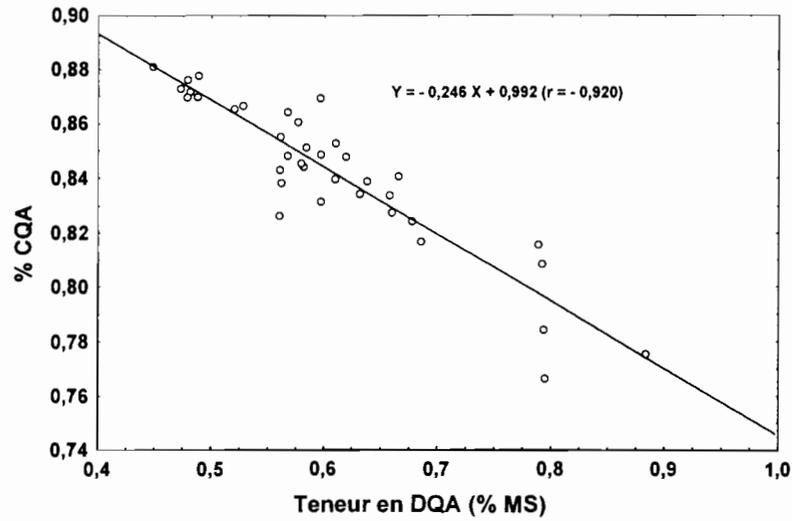
Les acides chlorogéniques constituent le second groupe de composés avec 25 % de la part extraite. Les trois classes CQA, DQA et FQA représentent respectivement 80, 15 et 5 % de la teneur globale en acides chlorogéniques. Deux isomères (3- et 4-FQA) totalisant moins

**Tableau 4.1. :** Moyennes et amplitudes (Min-Max) des variables biochimiques des accessions spontanées de *C. arabica* (exprimée en % MS).

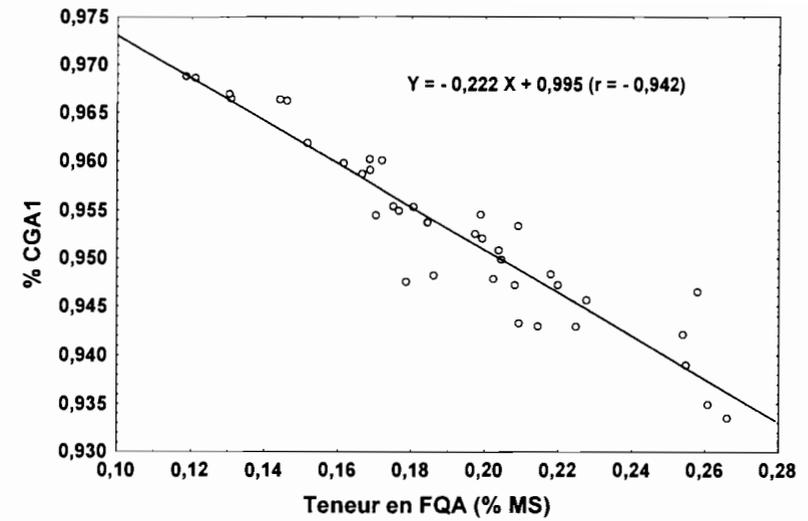
Variabiles	Moyenne	Min	Max	F	ddl <sub>1</sub> , ddl <sub>2</sub>	p
Saccharose	9,23	7,40	11,1	40,1	38, 117	< 0,001
3-CQA	0,20	0,16	0,25	17,3	38, 117	< 0,001
5- & 4-CQA	3,06	2,43	3,72	7,98	38, 117	< 0,001
CQA	3,26	2,61	3,97	7,46	38, 117	< 0,001
5-FQA	0,15	0,08	0,22	47,5	38, 117	< 0,001
FQA	0,19	0,12	0,27	34,9	38, 117	< 0,001
3,4-DQA	0,12	0,08	0,17	30,3	38, 117	< 0,001
3,5-DQA	0,23	0,17	0,36	31,9	38, 117	< 0,001
4,5-DQA	0,25	0,18	0,38	29,5	38, 117	< 0,001
DQA	0,60	0,43	0,88	27,9	38, 117	< 0,001
CGA1	3,9	3,2	4,6	4,98	38, 117	< 0,001
CGA	4,1	3,4	4,8	4,48	38, 117	< 0,001
%CQA	84,3	76,7	88,1	247	38, 117	< 0,001
%CGA1	95,1	89,1	96,9	2,15	38, 117	< 0,001
%3-CQA	5,0	4,0	6,4	111	38, 117	< 0,001
%5-CQA	75,3	67,7	80,3	16,7	38, 117	< 0,001
%5-FQA	3,7	2,1	5,4	234	38, 117	< 0,001
%3,4-DQA	2,9	1,9	4,6	171	38, 117	< 0,001
%3,5-DQA	5,7	4,0	8,6	164	38, 117	< 0,001
%4,5-DQA	6,2	4,6	9,3	155	38, 117	< 0,001
Caféine	1,22	0,96	1,62	10,5	38, 117	< 0,001
Trigonelline	1,19	0,88	1,77	20,9	38, 117	< 0,001

**Tableau 4.2.** : Composantes principales de la diversité intraspécifique de *C. arabica* pour les acides chlorogéniques. Les coefficients de corrélation sont représentés par les chiffres sans parenthèse (en gras : les corrélations significatives). Les contributions des composantes à la variable de diversité sont représentées par les chiffres entre parenthèse (exprimée en pourcentage  $100 \times R^2$ ).

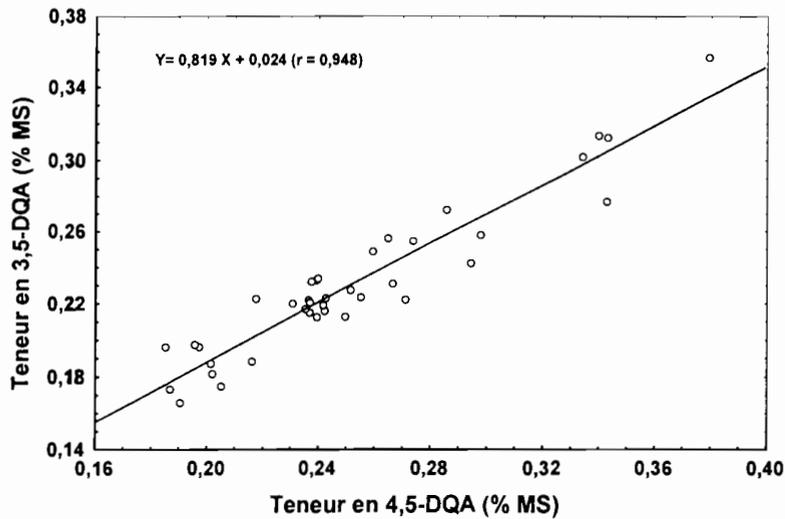
	Facteur 1	Facteur 2	Facteur 3	Facteur 4	Somme des R <sup>2</sup>
3-CQA	-0,036	<b>0,360</b> (13,0)	0,030	<b>0,925</b> (85,6)	97,9
5-CQA	<b>-0,362</b> (13,1)	<b>0,930</b> (86,5)	0,016	-0,046	99,6
5-FQA	-0,085	0,194	<b>0,976</b> (95,3)	-0,024	92,7
3,4-DQA	<b>0,849</b> (72,1)	0,034	-0,053	<b>0,485</b> (23,5)	95,5
3,5-DQA	<b>0,945</b> (89,3)	0,121	-0,084	-0,237	88,2
4,5-DQA	<b>0,972</b> (94,5)	0,065	-0,141	-0,075	94,9
CQA	<b>-0,354</b> (12,5)	<b>0,934</b> (87,2)	0,019	0,042	99,5
FQA	-0,087	0,200	<b>0,974</b> (94,9)	0,036	92,9
DQA	<b>0,989</b> (97,8)	0,084	-0,108	-0,023	98,1
CGA	0,036	<b>0,991</b> (98,2)	0,121	0,038	95,3
CGA1	0,051	<b>0,998</b> (99,6)	-0,026	0,033	99,4
%CQA	<b>-0,951</b> (90,4)	0,294	0,091	0,028	90,6
%CGA1	-0,103	0,135	<b>-0,984</b> (96,8)	-0,011	62,4
%3-CQA	-0,065	-0,198	-0,045	<b>0,965</b> (93,1)	92,7
%5-CQA	<b>-0,883</b> (78,0)	0,287	-0,173	-0,193	77,3
%5-FQA	-0,106	-0,112	<b>0,983</b> (97,7)	-0,055	90,6
%3,4-DQA	<b>0,807</b> (65,1)	-0,318	-0,095	<b>0,436</b> (10,0)	84,2
%3,5-DQA	<b>0,915</b> (83,7)	-0,259	-0,117	-0,240	83,0
%4,5-DQA	<b>0,926</b> (85,7)	-0,284	-0,171	-0,079	86,1



**Figure 4.1.** : Relation entre les variables % CQA et teneur DQA chez *C. arabica*.



**Figure 4.3.** : Relation entre les variables %CGA1 et teneur en FQA chez *C. arabica*.



**Figure 4.2.** : Relation entre teneur en 4,5-DQA et teneur en 3,5-DQA chez *C. arabica*.

de 2 % de la teneur en CGA n'ont pas été pris en considération dans cette étude. Les autres isomères présentent des différences significatives entre accessions à la fois pour les teneurs absolues et relatives (% en composés) (Tab. 4.1.).

Quatre composantes principales expliquent respectivement 41 %, 23 %, 19 % et 13 % de la diversité des teneurs en acides chlorogéniques. La première composante exprime l'accumulation des isomères de la classe des DQA, au détriment de celle des CQA (Tab. 4.2.). Une forte relation linéaire et négative existe entre le %CQA et la teneur en DQA (Fig. 4.1.). La valeur à l'origine (0,992) ne diffère pas de 1, signifiant que l'équation peut être reformulée comme suit :  $CQA + DQA = \text{constante}$ . Ceci correspond à un *trade-off* entre les deux composés. Une forte relation positive existe aussi entre les teneurs en 3,5-DQA et 4,5-DQA (Figure 4.2.). Ici, la valeur à l'origine (0,024) ne diffère pas de zéro, signifiant que le rapport 3,5-DQA/ 4,5-DQA est constant.

La deuxième composante principale représente principalement l'accumulation simultanée en CQA et DQA (CGA1). La troisième composante principale correspond à l'accumulation de FQA, au dépend de CGA1 (Fig. 4.3.). Comme pour la figure 4.1., la droite coupe l'axe des ordonnées à  $y = 1$  et l'équation peut ainsi se réécrire  $CQA + DQA = \text{constante}$  (*trade-off*). La dernière composante principale explique les variations de teneur en 3-CQA, qui apparaît ainsi indépendante de la teneur en 5-CQA.

#### 4. 1. 3. Les alcaloïdes

La caféine représente 7,5 % de la fraction extraite. Comme pour les composés précédents, la caféine montre des différences significatives entre accessions, avec une teneur variant de 0,96 à 1,62 % MS (Tab. 4.1.). La teneur en caféine (CAF) est corrélée avec la troisième composante principale de l'analyse sur les CGA ( $r = 0,50$ ), c'est-à-dire avec le « *trade-off* » : teneur en FQA vs teneur en CGA1. Plus il y a de FQA, plus il y a de caféine. Cependant, la valeur du  $R^2$  indique que seulement 25 % de la variance de CAF est expliquée par cette composante principale. Le rapport CAF/ CQA devrait être constant si la caféine et le CQA n'existaient que sous forme de chlorogénates. En fait, ce rapport varie de 4,39 à 8,31 avec une moyenne de 6,18. De plus, il est corrélé avec la quatrième composante ( $r = - 0,73$ ), c'est-à-dire avec la teneur en 3-CQA et non avec le deuxième axe comme on aurait pu s'y attendre.

**Tableau 4.3.** : Moyennes et amplitudes (Min-Max) des variables biochimiques des accessions spontanées de *C. canephora* (en % MS). F et p correspondent aux résultats des ANOVA.

Variable	Moyenne	Min	Max	F	ddl <sub>1</sub> , ddl <sub>2</sub>	p
3-CQA	0,81	0,43	1,33	66,1	36, 111	< 0,001
5- & 4-CQA	6,85	4,68	8,96	9,49	36, 111	< 0,001
CQA	7,66	5,12	9,50	7,88	36, 111	< 0,001
5-FQA	1,17	0,62	2,04	37,2	36, 111	< 0,001
FQA	1,43	0,77	2,23	33,8	36, 111	< 0,001
3,4-DQA	0,77	0,54	0,95	11,9	36, 111	< 0,001
3,5-DQA	0,71	0,38	1,13	39,8	36, 111	< 0,001
4,5-DQA	0,83	0,41	1,24	22,6	36, 111	< 0,001
DQA	2,31	1,57	3,03	15,2	36, 111	< 0,001
CGA1	9,96	6,81	12,21	6,90	36, 111	< 0,001
CGA	11,3	7,88	14,4	7,61	36, 111	< 0,001
%CQA	76,8	68,5	84,2	192	36, 111	< 0,001
%CGA1	87,5	82,6	93,1	187	36, 111	< 0,001
%3-CQA	7,11	3,71	11,4	2720	36, 111	< 0,001
%5- & 4-CQA	60,1	54,0	66,8	304	36, 111	< 0,001
%5-FQA	10,2	6,06	14,7	166	36, 111	< 0,001
%3,4-DQA	6,75	4,66	9,19	330	36, 111	< 0,001
%3,5-DQA	6,26	3,35	8,86	1273	36, 111	< 0,001
%4,5-DQA	7,30	4,67	11,9	58,3	36, 111	< 0,001
Saccharose	5,45	4,05	7,05	25,6	36, 111	< 0,001
Caféine	2,54	1,51	3,33	16,7	36, 111	< 0,001
Trigonelline	1,01	0,53	1,34	9,90	36, 111	< 0,001

La trigonelline (TRI) représente aussi 7,5 % de la fraction extraites et des différences significatives existent entre les différentes accessions : la teneur varie de 0,88 à 1,77 % MS à l'intérieur de l'espèce. La trigonelline n'est corrélée à aucun autre composé étudié.

## 4. 2. Diversité de l'espèce *C. canephora*

La part extraite de la poudre représente 20 % MS des grains verts. Les composés analysés sont présentés par ordre d'importance.

### 4. 2. 1. Les acides chlorogéniques

Les acides chlorogéniques forment le premier groupe des composés analysés avec 53 % de la fraction extraite. Les classes CQA, DQA et FQA représentent respectivement 67%, 20% et 13 % de la teneur globale en CGA. Les isomères 4-FQA et 3-FQA totalisant moins de 2 % de la teneur en CGA n'ont pas été pris en considération dans la suite de cette étude. Les autres isomères présentent des différences significatives entre accessions, aussi bien pour les teneurs absolues que relatives (Tab. 4.3.). Les tests F montrent clairement que ces teneurs relatives sont plus discriminantes que les teneurs absolues, et particulièrement pour le %3-CQA et le %3,5-DQA.

La diversité en acides chlorogéniques des accessions peut être structurée selon : i) l'origine géographique, ii) les groupes génétiques observés d'après les marqueurs moléculaires et iii) la teneur en 3-CQA.

L'origine géographique des différentes accessions (Cameroun, Congo, Côte-d'Ivoire, Guinée et RCA, République Centricaine) est une source de diversité pour les six variables suivantes : 3-CQA, 5-FQA, %CGA1, %3-CQA, %5-FQA et %4,5-DQA (Tab. 4.4.). L'importance de cette source de diversité varie de 29 % à 39 %, suggérant que les variations à l'intérieur de chaque pays demeurent prépondérantes. A l'exception du %4,5-DQA, qui oppose la RCA des autres, les autres caractères opposent l'Ouest (Guinée et Côte d'Ivoire) à l'Est (RCA et Cameroun). Les pays de l'Ouest montrent de faibles valeurs pour le 3-CQA, le %3-CQA et le %CGA1 et de fortes pour le 5-FQA et le %5-FQA.

Les groupes génétiques définis par les analyses RFLP (Dussert *et al*, 1999) montrent des différences uniquement pour le 4,5-DQA (teneur absolue et relative) (Tab. 4.5.), opposant le groupe B aux groupes C et D. L'importance de la diversité intra-groupe est proche de 45 %. Pour les autres caractères, la variance intra-groupe représente 100 % de la diversité.

**Tableau 4.4.** : Teneurs moyennes (exprimées en % MS) et résultats des ANOVA (valeurs de F et de p), des caractères biochimiques présentant des différences significatives entre les origines géographiques de *C. canephora*. Les codes des pays sont les suivants : Guinée (GUI), Côte-d'Ivoire (COD), Cameroun (CAM), République Centrafricaine (RCA) et Congo (CON). Les résultats des test de Newman et Keuls sont représentés par les lettres a et b en exposant.

Caractère	GUI	COD	CAM	RCA	CON	F	ddl <sub>1</sub> , ddl <sub>2</sub>	p
3-CQA	0,69 <sup>a</sup>	0,68 <sup>a</sup>	1,05 <sup>b</sup>	0,90 <sup>ab</sup>	0,70 <sup>a</sup>	3,74	4, 32	0,013
5-FQA	1,40 <sup>a</sup>	1,33 <sup>ab</sup>	1,08 <sup>ab</sup>	0,88 <sup>b</sup>	1,10 <sup>ab</sup>	3,16	4, 32	0,027
% CGA1	86,3 <sup>a</sup>	86,1 <sup>a</sup>	87,2 <sup>ab</sup>	89,8 <sup>b</sup>	88,6 <sup>ab</sup>	3,10	4, 32	0,029
% 3-CQA	5,81 <sup>a</sup>	6,25 <sup>ab</sup>	9,18 <sup>b</sup>	8,20 <sup>ab</sup>	6,29 <sup>a</sup>	3,65	4, 32	0,015
% 5-FQA	11,6 <sup>a</sup>	11,9 <sup>a</sup>	9,52 <sup>ab</sup>	7,97 <sup>b</sup>	9,70 <sup>ab</sup>	4,37	4, 32	0,006
% 4,5-DQA	6,93 <sup>a</sup>	6,45 <sup>a</sup>	6,83 <sup>a</sup>	9,09 <sup>b</sup>	7,57 <sup>ab</sup>	3,14	4, 32	0,027

**Tableau 4.5.** : Teneurs moyennes (exprimées en % MS) et résultats (valeurs de F et de p) des ANOVA, des caractères biochimiques présentant des différences significatives entre les origines géographiques de *C. canephora*. Les groupes sont nommés de A à E. Les résultats des tests de Newman et Keuls sont représentés par les lettres a et b en exposant.

caractères	A	B	C	D	E	F test	df <sub>1</sub> , df <sub>2</sub>	p
4,5-DQA	0,90 <sup>ab</sup>	1,09 <sup>b</sup>	0,67 <sup>a</sup>	0,69 <sup>a</sup>	0,84 <sup>ab</sup>	4,84	4, 24	0,005
%4,5-DQA	7,58 <sup>ab</sup>	9,63 <sup>b</sup>	6,00 <sup>a</sup>	6,43 <sup>a</sup>	7,54 <sup>ab</sup>	5,20	4, 24	0,004

**Tableau 4.6.** : Teneurs moyennes (exprimées en % MS) et résultats (valeurs de F et de p) des ANOVA, des caractères biochimiques présentant des différences significatives entre les deux groupes CGA-1 et CGA-2 discriminés par leur teneur en 3-CQA de *C. canephora*.

Caractères	Groupe CGA-1	Groupe CGA-2	F	ddl <sub>1</sub> , ddl <sub>2</sub>	p
3-CQA	0,61	1,11	169	1, 35	< 0,001
3,5-DQA	0,80	0,56	20,9	1, 35	< 0,001
%CQA	76,0	78,0	4,20	1, 35	0,048
%3-CQA	5,54	9,69	89,1	1, 35	< 0,001
%5-CQA	61,2	58,2	7,70	1, 35	0,009
%3,5-DQA	7,12	4,86	46,3	1, 35	< 0,001

**Tableau 4.7.** . Composantes principales de la diversité intraspécifique de *C. canephora* pour les acides chlorogéniques. Les coefficients de corrélation sont indiqués sans parenthèses (en gras : les corrélations significatives). Les contributions des composantes à la variable de diversité sont indiquées entre parenthèses (exprimée en pourcentage 100 x R<sup>2</sup>).

	Facteur 1	Facteur 2	Facteur 3	Facteur 4	Facteur 5
3-CQA	0,134	<b>0,951</b> (90,4)	-0,085	0,109	0,164
5-CQA	<b>0,957</b> (91,5)	-0,205	0,088	-0,135	-0,116
5-FQA	<b>0,385</b> (14,8)	-0,288	<b>0,871</b> (75,9)	0,004	-0,055
3,4-DQA	<b>0,363</b> (13,2)	0,067	0,053	0,289	<b>0,877</b> (76,9)
3,5-DQA	<b>0,417</b> (17,4)	<b>-0,837</b> (70,1)	0,132	0,245	0,085
4,5-DQA	0,292	-0,006	0,008	<b>0,944</b> (82,3)	0,088
CQA	<b>0,987</b> (97,4)	0,085	0,061	-0,100	-0,065
FQA	<b>0,380</b> (14,4)	0,031	<b>0,922</b> (85,0)	0,048	-0,009
DQA	<b>0,468</b> (21,9)	<b>-0,405</b> (16,4)	0,086	<b>0,701</b> (49,1)	<b>0,342</b> (11,7)
CGA	<b>0,926</b> (85,7)	-0,053	<b>0,334</b> (11,2)	0,154	0,060
CGA1	<b>0,977</b> (95,4)	-0,074	0,080	0,168	0,069
%CQA	0,221	<b>0,461</b> (21,3)	-0,045	<b>-0,767</b> (58,8)	<b>-0,382</b> (14,6)
%CGA1	0,034	-0,095	<b>-0,992</b> (98,4)	0,014	0,011
%3-CQA	-0,144	<b>0,948</b> (89,9)	-0,165	0,066	0,140
%5-CQA	0,286	<b>-0,357</b> (12,7)	<b>-0,485</b> (23,5)	<b>-0,622</b> (38,7)	<b>-0,383</b> (14,7)
%5-FQA	0,005	-0,317	<b>0,937</b> (87,7)	-0,066	-0,092
%3,4-DQA	<b>-0,362</b> (13,1)	0,094	-0,199	0,170	<b>0,885</b> (78,3)
%3,5-DQA	-0,006	<b>-0,948</b> (89,9)	-0,031	0,205	0,073
%4,5-DQA	-0,132	0,024	-0,148	<b>0,970</b> (94,1)	0,057

La teneur en 3-CQA structure les accessions en deux groupes : CGA-1 et CGA-2 (Fig. 4.4.). La teneur moyenne en 3-CQA est respectivement de 0,61 et 1,11 % MS pour CGA-1 et CGA-2. Le groupe CGA-1 est aussi caractérisé par de faibles valeurs pour le %CQA et le %3-CQA et des valeurs élevées pour le 3,5-DQA, le %5-CQA et le %3,5-DQA (Tab. 4.6.).

Cinq composantes principales expliquent 26 %, 22 %, 21 %, 19 % et 11 % de la diversité des teneurs en acides chlorogéniques. La première composante s'interprète comme l'accumulation des CGA, avec peu ou pas d'effet sur l'importance relative des teneurs de chacun des isomères (Tab. 4.7.). La deuxième composante représente le « *trade-off* » entre le 3-CQA et le 3,5-DQA (teneurs absolues et relatives). La figure 4.5. illustre bien ce phénomène : plus il y a de 3-CQA, moins il y a de 3,5-DQA. Cette composante présente deux intérêts : i) elle discrimine les groupes CGA-1 et CGA-2 et ii) la pente de la droite indique un rapport de type 2 : 1 (deux isomères 3-CQA pour un 3,5-DQA). La troisième composante correspond à l'accumulation du 5-FQA (teneur absolue et relative) au détriment du CGA1. La relation entre les variables %CGA1 et %5-FQA (Fig. 4.6.) peut s'écrire  $y = 81,5 (1 - x/10)$ , révélant un *trade-off* de type 10 : 1. La quatrième composante représente l'accumulation du 4,5-DQA (valeurs absolue et relative) au détriment du %CQA et du %5-CQA. La relation entre le %5-CQA et %4,5-DQA représente aussi un *trade-off* (Fig. 4.7.). Enfin, la dernière composante correspond à l'accumulation du 3,4-DQA (valeurs absolue et relative) avec ses effets négatifs sur le %CQA et le %5-CQA et positifs sur le %3,4-DQA. Un *trade-off* entre %4-CQA et %3,4-DQA devrait être observé. Malheureusement, un seul pic correspondant aux isomères 4-CQA et 5-CQA est mis en évidence d'après notre méthode d'analyse (Ky *et al.*, 1997 – annexe A) et nous pouvons seulement déceler une faible relation entre le %5-CQA, qui inclut une faible part de 4-CQA, et le %3,4-DQA.

#### 4. 2. 2. Le saccharose

Le saccharose représente 26 % de la fraction extraite et se classe ainsi au deuxième rang après les acides chlorogéniques. Sa teneur présente des différences significatives entre les accessions (Tab. 4.3.). Aucune différence n'a été détectée ni entre les différentes origines géographiques ( $F_{4,32} = 1,16$  ;  $p = 0,35$ ), ni entre les groupes génétiques ( $F_{4,24} = 0,86$  ;  $p = 0,50$ ) et ni même entre les groupes CGA-1 et CGA-2 ( $F_{1,35} = 2,72$  ;  $p = 0,11$ ).

#### 4. 2. 3. Les alcaloïdes

La caféine et la trigonelline représentent respectivement 10 et 5 % de la partie extraite de la poudre. Les deux alcaloïdes présentent des différences significatives entre accessions

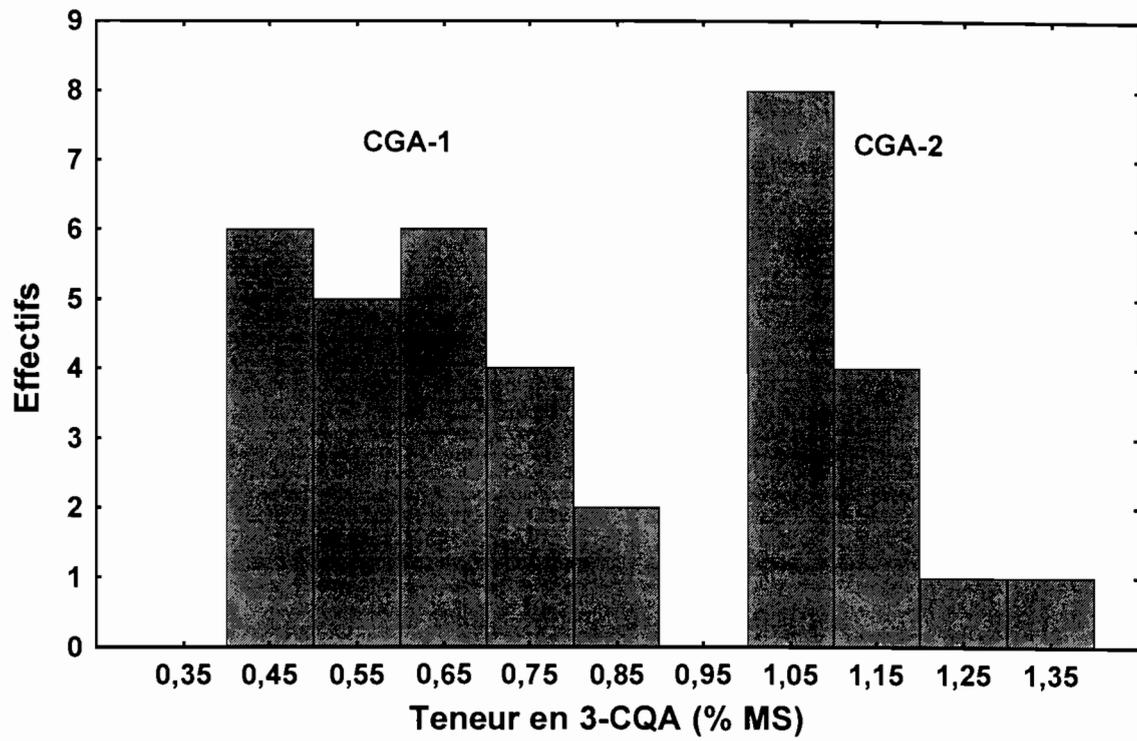
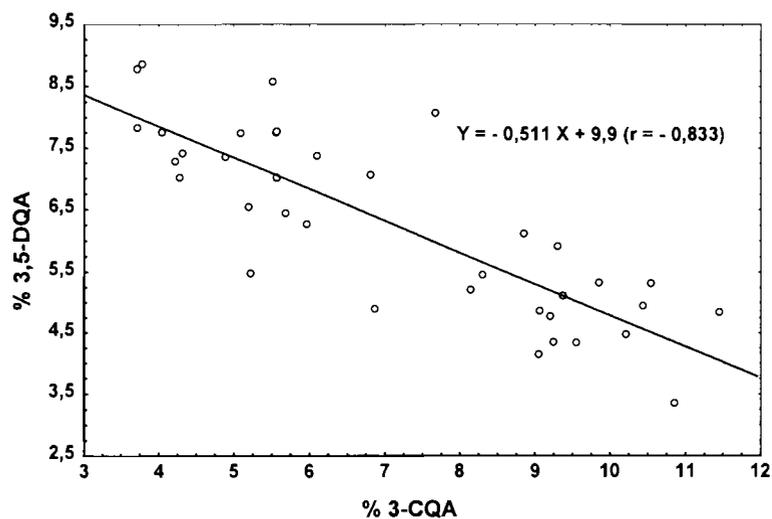
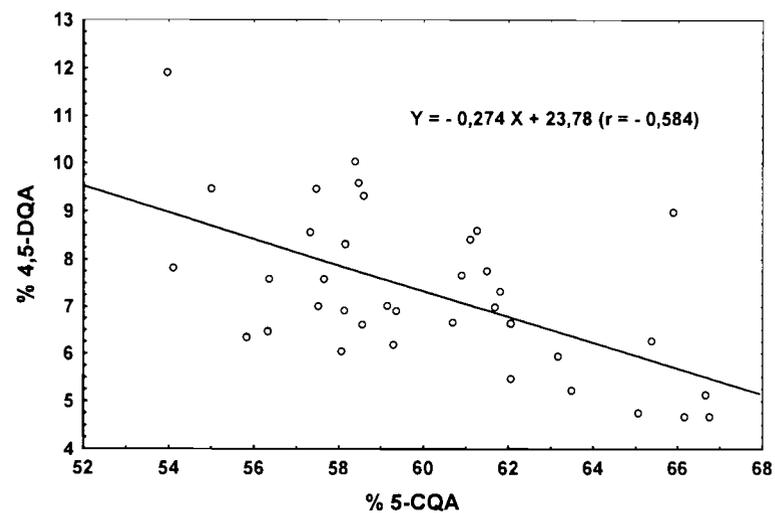


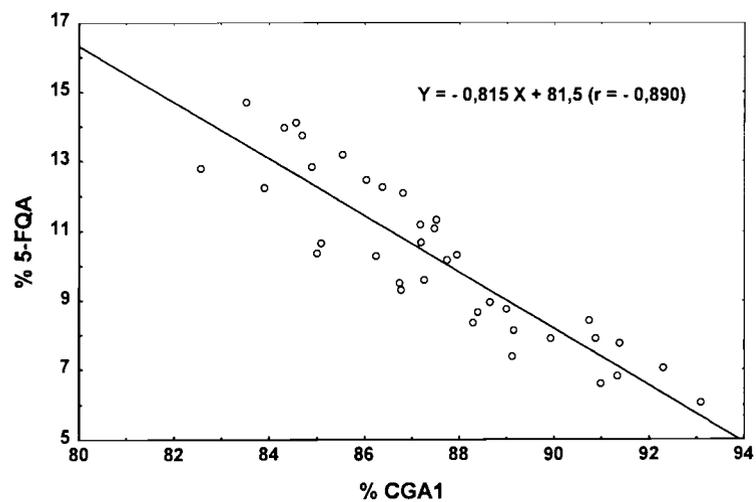
Figure 4.4. : Distribution bi-modale de la teneur en 3-CQA chez *C. canephora*.



**Figure 4.5.** : Relation entre les variables % 3-CQA et % 3,5-DQA chez *C. canephora*.



**Figure 4.7.** : Relation entre les variables % 5-CQA et % 4,5-DQA chez *C. canephora*.



**Figure 4.6.** : Relation entre les variables % 5-FQA et % CGA1 chez *C. canephora*.

**Tableau 4.8. :** Différences interspécifiques entre les accessions spontanées *C. arabica* et *C. canephora*.

Caractère	Variance	Variance	F obs	ddl <sub>1</sub> , ddl <sub>2</sub>	p
	<i>C. arabica</i>	<i>C. canephora</i>			
3-CQA	0,000603	0,073506	122	36, 39	< 0,001
5-&4-CQA	0,059795	0,784545	13,1	36, 39	< 0,001
5-FQA	0,001182	0,118766	100	36, 39	< 0,001
3,4-DQA	0,000475	0,012741	26,8	36, 39	< 0,001
3,5-DQA	0,001734	0,038369	22,1	36, 39	< 0,001
4,5-DQA	0,002262	0,041491	18,3	36, 39	< 0,001
CQAt	0,063703	0,802612	12,6	36, 39	< 0,001
FQAt	0,001479	0,142241	96,1	36, 39	< 0,001
DQAt	0,011025	0,151556	13,7	36, 39	< 0,001
CGA	0,072575	1,704695	23,5	36, 39	< 0,001
CGA1	0,060681	1,190674	19,6	36, 39	< 0,001
%CQAt	7,776089	9,281518	1,19	36, 39	0,29
%CGA1	1,795817	6,777209	3,77	36, 39	< 0,001
%CQA3	0,319205	5,804806	18,2	36, 39	< 0,001
%CQA5	7,868435	12,4055019	1,58	36, 39	0,08
%FQA5	0,671280	5,559940	8,28	36, 39	< 0,001
%3,4-DQA	0,324901	0,893529	2,75	36, 39	0,001
%3,5-DQA	1,083151	2,186997	2,02	36, 39	0,017
%4,5-DQA	1,484140	2,734469	1,84	36, 39	0,032
Saccharose	0,946919	0,665679	1,42	37, 36	0,015
Caféine	0,019332	0,198145	10,2	36, 39	< 0,001
Trigonelline	0,025995	0,022062	1,18	39, 36	0,31

(Tab. 4.3.). Comme pour la teneur en saccharose, aucune différence n'a été détectée entre les niveaux de structuration des accessions et CAF et TRI : entre les origines géographiques ( $F_{4,32} = 0,42$  ;  $p = 0,79$ ) et ( $F_{4,32} = 2,57$  ;  $p = 0,06$ ), entre les groupes génétiques ( $F_{4,24} = 0,60$  ;  $p = 0,67$ ) et ( $F_{4,24} = 1,29$  ;  $p = 0,30$ ), et entre les groupes CGA-1 et CGA-2 ( $F_{1,35} = 2,09$  ;  $p = 0,15$ ) et ( $F_{1,35} = 0,64$  ;  $p = 0,43$ ).

### 4. 3. Comparaison entre *C. arabica* et *C. canephora*

La comparaison de la diversité (à l'aide du test F) montre que la diversité est plus large chez *C. canephora* (CAN) pour l'ensemble des variables, à l'exception de la teneur en trigonelline, du %CQA, du %CGA1 et du %5-CQA (pour ces quatre dernières variables, c'est la diversité chez *C. arabica* qui est plus importante). Le rapport des variances est particulièrement élevé ( $> 100$ ) pour les teneurs en 3-CQA et 5-FQA (Tab. 4.8.). Il est compris entre 10 et 23 pour les autres composés. Dans tous les cas, ce rapport est plus élevé pour les variables absolues que pour les variables relatives.

La comparaison des teneurs moyennes des variables montre des différences interspécifiques significatives pour l'ensemble des caractères, à l'exception du %3,5-DQA (Tab. 4.9.). Les teneurs en saccharose et en trigonelline sont respectivement 1,7 et 1,2 fois plus importante chez *C. arabica* que chez *C. canephora*. Les importances relatives du 5-CQA dans CGA, du CQA dans CGA1 et de CGA1 dans CGA sont aussi plus élevées chez la première espèce. A l'opposé, les teneurs en caféine et des isomères d'acides chlorogéniques sont plus importantes chez *C. canephora*. Le rapport « *C. canephora/C. arabica* » varie de 2,1 à 7,8 respectivement pour les teneurs en caféine et en 5-FQA.

## 4. 4. Discussion

### 4. 4. 1. Comparaison entre collections d'accessions spontanées et de variétés cultivées

La comparaison des teneurs en composés biochimiques entre échantillons de caféiers spontanées (que nous appellerons collections spontanées) et cultivars permet d'avoir une idée sur l'existence ou non, de la réduction de variabilité depuis la mise en culture.

Pour *C. canephora*, les résultats de teneur en CGA, comparés à des variétés de Robustas (Balyaya et Clifford, 1995), révèle respectivement que : i) la collection étudiée présente une teneur moyenne plus élevée (11,3 % MS vs. 9,6 % MS) et ii) l'amplitude est plus importante : 7,9 - 14,4 % MS contre 7,0 - 10,0 % MS. En revanche, les teneurs moyennes en caféine y sont

**Tableau 4.9.** : Comparaison des accessions spontanées *C. arabica* et *C. canephora* par le test de Kruskal & Wallis.

Caractère	Moyenne	Moyenne	p
	<i>C. arabica</i>	<i>C. canephora</i>	
Saccharose	9,23	5,45	< 0,001
3-CQA	0,20	0,81	< 0,001
5- & 4-CQA	3,06	6,85	< 0,001
CQA	3,26	7,66	< 0,001
5-FQA	0,15	1,17	< 0,001
FQA	0,19	1,43	< 0,001
3,4-DQA	0,12	0,77	< 0,001
3,5-DQA	0,23	0,71	< 0,001
4,5-DQA	0,25	0,83	< 0,001
DQA	0,6	2,31	< 0,001
CGA1	3,9	9,96	< 0,001
CGA	4,1	11,3	< 0,001
%CQA	84,3	76,8	< 0,001
%CGA1	95,1	87,5	< 0,001
%3-CQA	5	7,11	< 0,001
%5-CQA	75,3	60,1	< 0,001
%5-FQA	3,7	10,2	< 0,001
%3,4-DQA	2,9	6,75	< 0,001
%3,5-DQA	5,7	6,26	0,25
%4,5-DQA	6,2	7,3	< 0,001
Caféine	1,22	2,54	< 0,001
Trigonelline	1,19	1,01	0,006

plus faibles : 2,54 % MS contre 3,15 % MS. Pour la trigonelline, la teneur moyenne de la collection (1,01 % MS) est supérieure à celle observée par l'équipe de Stennert et Maier (1993 et 1994) (0,67 % MS) pour les cultivars. La teneur en saccharose suit la même tendance, avec un gain d'environ 1 % MS en faveur de la collection (Tressl *et al.*, 1982).

Pour la collection de *C. arabica* spontanées, les teneurs en CGA (moyenne : 4,1 % MS; amplitude : 3,4 - 4,8 % MS) sont plus faibles et plus groupées que celles trouvées au sein de 21 variétés cultivées (moyenne : 5,26 % MS; amplitude : 4,43 - 6,63 % MS) (Clifford *et al.*, 1989b). Les teneurs moyennes en caféine sont semblables, mais l'amplitude est plus importante chez les variétés cultivées (0,76 à 1,82 % MS), incluant nos valeurs. La teneur en trigonelline des variétés est plus importante en terme de moyenne et d'amplitude chez les variétés cultivées (Stennert et Maier, 1993 et 1994). A l'opposé, la teneur en saccharose est plus importante dans notre collection (d'environ 1 % MS) (Tressl *et al.*, 1982).

Un point intéressant est la teneur plus élevée en caféine observée chez les cultivars. Ce composé est impliqué dans les systèmes de défense des caféiers contre les agressions de petits insectes (Anthony, 1992). Il est probable que la sélection de clones résistants et adaptés à des cultures à grande densité ait indirectement contribué à sélectionner des génotypes capables de synthétiser davantage de caféine.

La même tendance est observée chez *C. arabica* pour les CGA, qui assurent aussi une résistance aux stress de nature biologique et mécanique (Zuluaga *et al.*, 1971 ; Echandi et Fernandez, 1962 ; Macheix *et al.*, 1990). En revanche, les variétés cultivées de *C. canephora* ont une teneur moyenne en CGA plus faible que celle des accessions spontanées. Il faut souligner que *C. canephora* est une espèce très riche en CGA et que la teneur présente chez les variétés est suffisante pour assurer la résistance aux pathogènes. Une sélection sur l'amertume des cafés pourrait alors expliquer leur moindre teneur. Une autre explication apparaît en comparant la diversité cultivée à la diversité prospectée : les arbres introduits en caféiculture proviendraient de zones géographiques plus réduites (*bottleneck*).

#### 4. 4. 2. Voies de biosynthèse des acides chlorogéniques

Les composantes principales mises en évidence chez les acides chlorogéniques permettent de confirmer le schéma de biosynthèse (Fig. 2.2.5.) actuellement admis. Les résultats montrent qu'à une composante principale correspond une portion de la chaîne de biosynthèse et il est alors possible d'attribuer le fonctionnement d'une enzyme ou sa régulation pour expliquer la diversité liée à chaque composante. Pour résumer, on peut dire :

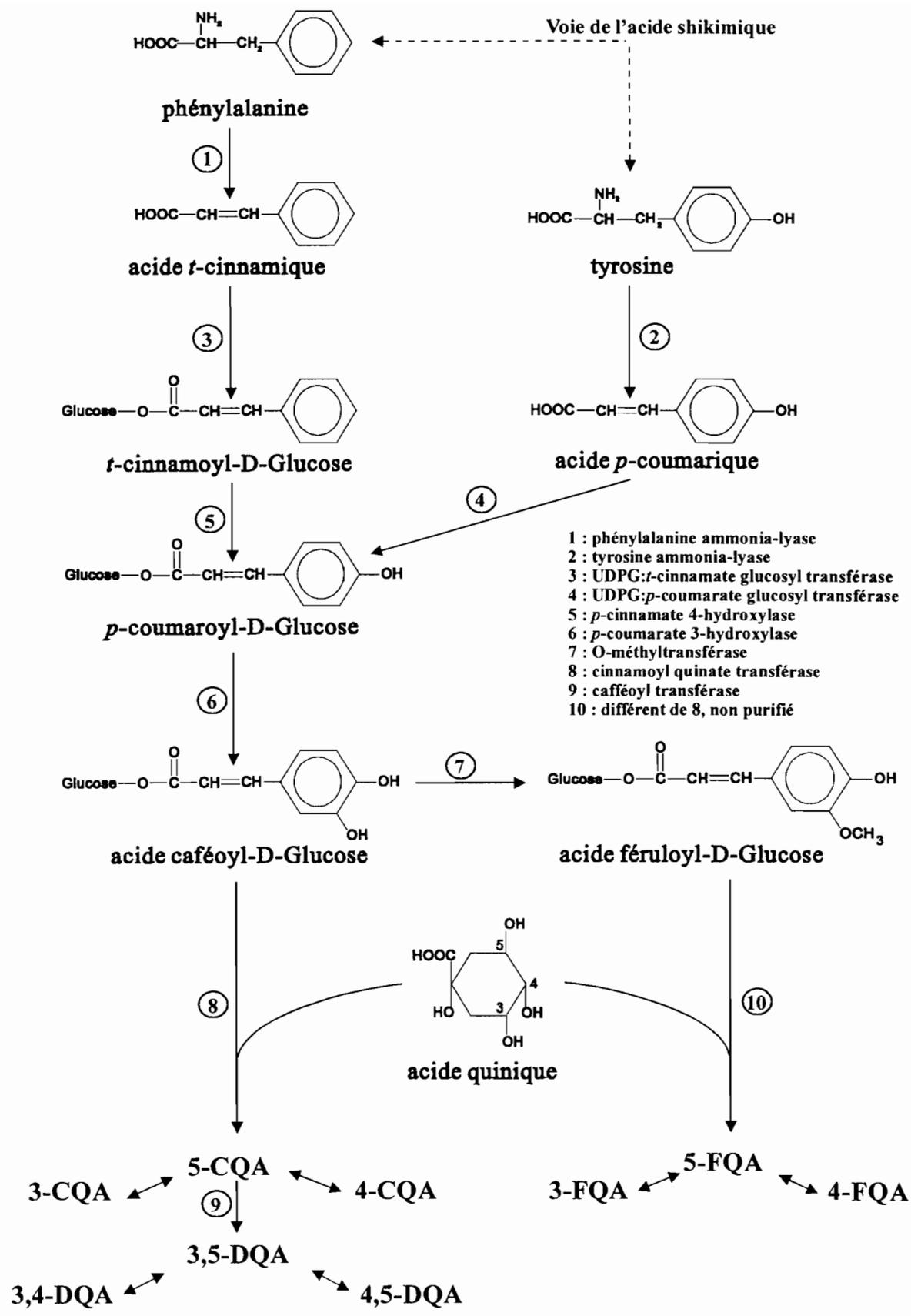


Figure 2.2.5. : Chaîne de biosynthèse conduisant à la formation des acides chlorogéniques.

une composante, une enzyme. Cette approche, utilisée aussi chez l'hévéa (Chrestin et Noirot, 1982 ; Chrestin, 1985), permet de vérifier la chaîne de biosynthèse proposée sur la figure 2.2.5. chez les caféiers.

L'analyse réalisée sur les accessions spontanées de *C. arabica* suggère que : i) la première composante, caractérisant l'accumulation des DQA au détriment de celle des CQA, pourrait correspondre à la cafféoyl transférase (enzyme 9 de la Fig. 2.2.5.), ii) la deuxième composante définissant la variable CGA1 (accumulation simultanée des CQA et DQA) correspondrait à la cinnamoyl quinate transférase (enzyme 8 de la Fig. 2.2.5.) et iii) la troisième composante caractérisant l'accumulation du 5-FQA (et à fortiori celle des FQA) au détriment de CGA1 pourrait correspondre aux enzymes 7 ou 10 de la même figure. La dernière composante correspondant à l'accumulation du 3-CQA (et dans une moindre mesure à celle du 3,4-DQA : Tab. 4.2.) est en « contradiction » avec la séquence de formation :  $\rightarrow 5\text{-CQA} \rightarrow 3,5\text{-DQA} \rightarrow 3,4\text{-DQA}$ . Ceci laisse penser qu'une voie parallèle pourrait exister suivant la séquence :  $3\text{-CQA} \rightarrow 3,4\text{-DQA}$ .

Certaines composantes existent chez les deux espèces, même si leur importance relative est différente. Ainsi, la première composante observée chez *C. canephora* ressemble très fortement à la deuxième, mise en évidence chez *C. arabica*. De même, la troisième composante observée chez les deux espèces ne diffèrent pas entre elles. En revanche, les composantes 2, 4 et 5 sont propres à *C. canephora* : la première sous-entend l'existence d'une enzyme entre le 3-CQA et le 3,5-DQA du type cafféoyl transférase, la seconde résulterait d'une régulation entre %5-CQA et %4,5-DQA, tandis que la troisième ferait intervenir une enzyme aboutissant à l'accumulation de 3,4-DQA.

Le point intéressant qui se dégage de ces analyses est l'existence possible d'une séquence  $3\text{-CQA} \rightarrow 3,4\text{-DQA}$  chez *C. arabica* et  $3\text{-CQA} \rightarrow 3,5\text{-DQA}$  chez *C. canephora* qui se grefferait sur la voie de biosynthèse actuellement considérée. Cette hypothèse reste plausible lorsque l'on sait que la formation des isomères à l'intérieur d'une même classe demeure encore incertaine et qu'elle semble dépendre du matériel végétal. En effet, chez la chicorée par exemple, seule la séquence  $\rightarrow 5\text{-CQA} \rightarrow 3,5\text{-DQA}$  existe (Taylor, 1968 et Mialoundama et Paulet, 1975), alors que chez la pomme, il semblerait que l'isomère 4-CQA soit le premier synthétisé. L'existence de deux groupes discriminés par la teneur en 3-CQA chez l'espèce *C. canephora* révèle l'importance de cette voie parallèle.

#### 4. 5. En résumé

Des points ont été confirmés :

- i) les grains verts de *C. canephora* sont plus riches en acides chlorogéniques et en caféine. A l'intérieur des acides chlorogéniques, l'importance relative des différentes classes change: les CQA sont ainsi plus importants chez *C. arabica* ;
- ii) les grains verts de *C. arabica* sont plus riches en sucre et en trigonelline ;
- iii) l'espèce *C. arabica* est nettement moins variable.

D'autres sont des nouveautés :

- i) la mise en culture a contribué à modifier la variabilité des teneurs en composés biochimiques étudié, en comparaison aux génotypes spontanés ;
- ii) l'analyse en composantes principales est une approche originale pour décortiquer les voies de biosynthèse aboutissant à la formation en cascade de composés d'une même famille, comme les acides chlorogéniques ;
- iii) en particulier, des voies additionnelles à la formation des isomères de la classe des CQA sont proposées, suivant les séquences :  
 $\rightarrow 3\text{-CQA} \rightarrow 3,4\text{-DQA}$  et  $\rightarrow 3\text{-CQA} \rightarrow 3,5\text{-DQA}$ .

**Chapitre 5 : Génétique quantitative  
des composés biochimiques**

---

## Chapitre 5 : Génétique quantitative des composés biochimiques

Les fortes teneurs en saccharose et en trigonelline, l'absence de caféine et la faible teneur en acides chlorogéniques sont les caractéristiques biochimiques de l'espèce *C. pseudozanguebariae* qu'il serait intéressant d'introduire chez *C. canephora*. Une meilleure connaissance du déterminisme génétique en terme quantitatif s'avère nécessaire. Il en est de même des interrelations existantes. L'analyse quantitative du déterminisme génétique de la teneur en caféine a déjà été étudiée dans le croisement PSE x DEW (Barre *et al.* 1998).

Les différents profils chromatographiques enregistrés chez les deux espèces sont présentés en annexe (5.1. ; 5.2. ; 5.3.). Nous allons aborder dans un premier temps l'importance relative du génotype dans la diversité observée. Chaque génotype n'étant représenté que par un arbre, l'effet de l'environnement n'est estimé qu'au travers des variations enregistrées entre années. Nous aborderons ensuite l'hérédité quantitative de chaque composé en comparant les espèces parentales, les hybrides F1 et les descendances de backcross (BCDEW et BCPSE). L'hypothèse d'additivité y sera aussi testée. Enfin, dans le cas des acides chlorogéniques, les relations phénotypiques entre composés seront étudiées.

*N.B.* : Les résultats obtenus dans ce chapitre concernant le saccharose et les acides chlorogéniques ont été publiés (Ky *et al.*, 1999 et 2000) et les articles correspondants sont joints en annexe B et C.

### 5. 1. Importance relative des variations génotypiques et inter-annuelles

#### 5. 1. 1. Pour la teneur en saccharose

L'effet « année » a été testée sur sept hybrides : cinq issus d'un backcross sur DEW et deux d'un backcross sur PSE, ceci pour 1995 et 1997 (Tab. 5.1.).

**Tableau 5.1.** : Variations des teneurs en saccharose (exprimée en % MS) chez 5 hybrides BCDEW et 2 hybrides BCPSE pour les années de récolte : 1995 et 1997.

Génotype	Groupe	1995	1997	Moyenne
L13 9/08	BCPSE-1	5,34	5,22	5,28
L13 14/03	BCPSE-1	5,33	4,36	4,84
L13 1/09	BCDEW	3,96	3,67	3,82
L13 9/03	BCDEW	3,32	4,04	3,68
L13 12/12	BCDEW	8,25	8,57	8,41
L13 19/13	BCDEW	5,55	6,54	6,04
L13 20/17	BCDEW	5,90	6,54	6,22
Moyenne		5,38	5,56	

**Tableau 5.2.** : Variations des teneurs en trigonelline (exprimée en % MS) chez 4 hybrides BCDEW et 3 hybrides BCPSE pour les deux années de récoltes : 1995 et 1997.

Génotype	Groupe	1995	1997	Moyenne
L13 9/08	BCPSE	0,98	1,03	1,00
L13 14/03	BCPSE	1,03	1,09	1,06
L13 1/12	BCPSE	1,70	1,52	1,61
L13 1/09	BCDEW	1,15	1,18	1,16
L13 9/03	BCDEW	0,76	0,85	0,80
L13 12/12	BCDEW	1,02	1,08	1,05
L13 20/17	BCDEW	0,73	0,77	0,75
Moyenne		1,05	1,07	

La teneur en saccharose diffère entre génotypes ( $F_{6, 56} = 121$  ;  $p < 0,001$ ), mais ne semble pas varier sur les deux années de récoltes ( $F_{1, 6} = 0,50$  ;  $p = 0,51$ ). Une interaction « année de récolte x génotype » existe ( $F_{6, 56} = 5,34$  ;  $p < 0,001$ ), qui montre que les différences entre années dépendent des génotypes. Néanmoins, en terme de contribution à la diversité observée, les différences entre arbres apparaissent nettement majoritaires (86,5 %), l'interaction et la résiduelle n'expliquant respectivement que 6,3 % et 7,2 % de la variance totale.

### 5. 1. 2. Pour la teneur en trigonelline

Afin de tester l'effet année, les grains de quatre hybrides BCDEW et trois hybrides BCPSE ont été récoltés en 1995 et 1997 (Tab. 5.2.). La variation des teneurs en trigonelline sont uniquement expliquée par les différences entre hybrides ( $F_{6, 42} = 45,7$  ;  $p < 0,001$ ). Les différences de teneur entre les deux années de récolte ( $F_{1, 42} = 1,06$  ;  $p = 0,58$ ) et l'interaction « année de récolte x génotype » ( $F_{6, 42} = 1,17$  ;  $p = 0,34$ ) ne sont pas significatives.

### 5. 1. 3. Pour les acides chlorogéniques

L'effet génotypique a été testé sur douze hybrides BCDEW pour les années 1995 et 1997. Des différences entre arbres sont observées pour l'ensemble des variables (Tab. 5.3.). La majorité des caractères présente à la fois un effet « année » et une interaction « année x génotypes ». Certains caractères ne montrent qu'une interaction (CGA, CQA, 5-CQA, 5-CQA/CQA, 3,5-DQA/DQA, 5-FQA/FQA, 4-FQA/FQA et 3-FQA/FQA) et un seul, le ratio 4,5-DQA/DQA, qu'un effet « année ». L'importance de la variance génotypique varie de 43% à 92 %. En général, les isomères présentant de faibles teneurs montrent un effet environnemental plus important.

## 5. 2. Diversité des espèces parentales et des hybrides : test d'additivité

### 5. 2. 1. Pour la teneur en saccharose

La teneur moyenne en saccharose est nettement plus faible chez DEW (- 30%) et cette différence est très hautement significative ( $F_{1, 12} = 16,80$  ;  $p < 0,001$ ). Les variations entre arbres à l'intérieur de chaque espèce sont aussi très hautement significatives ( $F_{12, 56} = 13,78$  ;  $p < 0,001$ ). Les contributions inter-, intra- spécifiques et résiduelles sont respectivement de 55 , 33 et 12 %.

**Tableau 5.3.** : Analyses de variance, ~~partiellement hiérarchisées~~, réalisées pour évaluer les différences entre deux années de récolte (effet fixe), pour 12 géotypes (effet aléatoire) ~~et pour 4 arbres par géotype~~ (effet aléatoire hiérarchisé dans les géotypes), pour les teneurs en acides chlorogéniques. Les contributions factorielles sont représentées entre parenthèse. NS : Non significatif ( $p > 0.05$ ); \* : Significatif ( $0.01 < p \leq 0.05$ ); \*\* : Hautement significatif ( $0.001 < p \leq 0.01$ ); \*\*\* : Très hautement significatif ( $p < 0.001$ ).

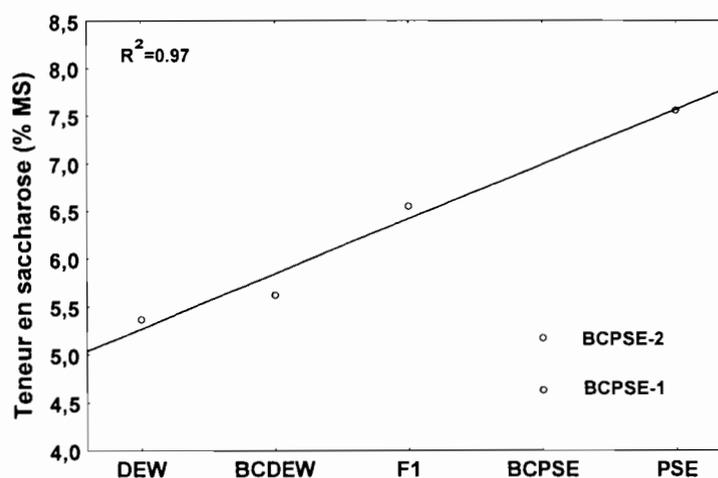
	Géotypes	Année	Interaction
	ddl =11, 48	ddl =1, 11	ddl=11, 48
CGA	48,4 *** (77,7 %)	0,25 NS -	7,80 *** (22,3 %)
CQA	71,4 *** (77,8 %)	1,49 NS -	11,0 *** (22,2 %)
DQA	23,5 *** (65,3 %)	24,7 *** (26,9 %)	2,35 * (7,8 %)
FQA	121 *** (71,3 %)	20,8 *** (18,6 %)	9,48 *** (22,0 %)
CQA/ CGA	82,2 *** (42,7 %)	66,4 *** (49,3 %)	8,63 *** (8,0 %)
DQA/ CGA	56,3 *** (73,2 %)	54,0 *** (24,0 %)	2,06 * (2,8 %)
FQA/ CGA	166 *** (68,2 %)	15,3 *** (17,8 %)	18,0 *** (14,1 %)
5-& 4-CQA	77,8 *** (77,7 %)	1,14 NS -	12,0 *** (22,3 %)
3-CQA	21,5 *** (61,3 %)	8,46 * (17,1 %)	4,60 *** (21,6 %)
5-& 4-CQA/ CQA	55,6 *** (84,3 %)	4,63 NS -	6,07 *** (15,7 %)
4,5-DQA	21,6 *** (74,4 %)	13,9 ** (17,0 %)	2,18 * (8,6 %)
3,5-DQA	22,0 *** (62,0 %)	21,0 *** (27,4 %)	2,78 ** (10,5 %)
3,4-DQA	13,9 *** (44,8 %)	29,0 *** (43,5 %)	2,69 ** (11,7 %)
4,5-DQA/ DQA	4,26 *** (76,9 %)	4,90 * (23,1 %)	1,50 NS -
3,5-DQA/ DQA	5,54 *** (59,5 %)	0,026 NS -	2,55 * (40,5 %)
3,4-DQA/ DQA	8,50 *** (52,1 %)	6,27 * (18,9 %)	3,09 ** (29,1 %)
5-FQA	121 *** (76,3 %)	20,6 *** (15,5 %)	7,45 *** (8,2 %)
4-FQA	65,1 *** (62,4 %)	4,49 * (8,9 %)	15,8 *** (28,7 %)
3-FQA	261 *** (73,4 %)	12,2 ** (13,1 %)	24,9 *** (13,5 %)
5-FQA/ FQA	173 *** (67,3 %)	0,074 NS -	42,7 *** (32,7 %)
4-FQA/ FQA	232 *** (69,7 %)	0,001 NS -	51,2 *** (30,3 %)
3-FQA/ FQA	205 *** (91,6 %)	4,58 NS -	10,4 *** (8,4 %)

**Tableau 5.4.** : Teneur en saccharose (en % MS) chez les espèces parentales, les hybrides F1 et les hybrides backcross (BCDEW et BCPSE). Les différences inter-génotypes sont données par les résultats du test F (\*\*\*: Très hautement significatif,  $p < 0,001$ ).

	DEW	BCDEW	F1	BCPSE	PSE
Moyenne	5,37	5,62	6,55	4,63	7,57
Min-Max	3,76-6,24	3,67-7,47	5,09-8,18	3,83-5,22	5,80-9,25
Amplitude	2,48	3,80	3,09	1,39	3,45
F <sub>6,28</sub>	43,5 ***	64,0 ***	49,6 ***	8,1 ***	10,9 ***

**Tableau 5.5.** : Teneur en saccharose (en % de MS) chez des hybrides BCPSE-2 issus de deux croisements contrôlés F1 x PSE.

Génotype	Croisement F1 x PSE	Saccharose
L22 15/10	D16 24/29 x Mo 9/010	5,09
L22 14/08	D16 24/29 x Mo 9/010	4,94
L22 20/02	D16 24/29 x Mo 9/010	4,54
L22 3/21	D16 24/29 x Mo 9/010	4,85
L22 12/10	D16 24/29 x 8056	5,51
L22 18/6	D16 24/29 x 8056	5,65
L22 19/6	D16 24/29 x 8056	5,15
L22 4/22	D16 24/29 x 8056	5,62



**Figure 5.1.** : Régression linéaire illustrant l'additivité de la teneur en saccharose, excepté pour les hybrides BCPSE.

La teneur moyenne en saccharose des sept hybrides F1 est proche de la moyenne des valeurs parentales (6,47 % MS) suggérant l'additivité (Tab. 5.4.). La variance « inter-hybrides F1 » ne diffère pas de la variance « inter-arbres » au sein des espèces parentales ( $F_{12, 6} = 1,17$ ); ceci confirme l'hétérozygotie parentale pour ce caractère vraisemblablement polygénique.

Les teneurs moyennes sont de 4,63 % MS pour BCPSE-1 contre 5,62 % MS pour BCDEW. Les amplitudes des deux groupes d'hybrides se chevauchent (Tab. 5.4.). Comme pour les espèces parentales et les hybrides F1, les différences « inter-hybrides » au sein de chacun des deux backcross sont hautement significatives ( $F_{13, 56} = 40,1$  ;  $r < 0,001$ ). Cependant, la variance « intra-BCPSE » est quatre fois moins importante (0,252) que chez les hybrides F1, tandis que la variance « intra-BCDEW » est double (2,130) de celle de la F1.

Concernant l'additivité, la teneur moyenne en saccharose des hybrides BCDEW (5,62 % MS) est proche de la valeur attendue [ $(F1 + DEW) / 2 = 5,15$  % MS]. A l'opposé, la teneur moyenne en saccharose des hybrides BCPSE-1 (4,63 % MS) est loin de cette valeur [ $(F1 + PSE) / 2 = 7,06$  % MS]. Cet écart au modèle des hybrides BCPSE-1 a été vérifié sur un second groupe de sept hybrides différents, notés BCPSE-2, récoltés la même année (1997). Leurs teneurs en saccharose (Tab. 5.5.) varient de 4,54 à 5,65 % MS avec une moyenne de 5,10 % MS. Notons que la différence entre les groupes BCPSE-1 et BCPSE-2 est significative ( $F_{1, 12} = 4,90$  ;  $p = 0,045$ ).

L'hypothèse d'additivité suggérée par les valeurs moyennes est confirmée par la régression linéaire (Fig. 5.1.), excepté pour les hybrides BCPSE.

### 5. 2. 2. Pour la teneur en trigonelline

La teneur moyenne en trigonelline des grains verts est de 0,57 % MS chez DEW et de 1,02 % chez PSE. La différence est très hautement significative ( $F_{1, 12} = 59,4$  ;  $p < 0,001$ ) et les amplitudes observées ne se recouvrent pas (Tab. 5.6.) Les différences notées entre les arbres à l'intérieur des espèces sont aussi très hautement significatives ( $F_{12, 42} = 17,4$  ;  $p < 0,001$ ), mais ne contribuent que pour 8 % à la variance totale (les contributions interspécifiques et résiduelles sont respectivement de 71 et 21 %).

La teneur en trigonelline chez les hybrides F1 est de 1,13 % en moyenne, et cette valeur est nettement plus élevée que la moyenne parentale (0,80 % MS). L'analyse de variance ( $F_{2, 18} = 28,5$  ;  $p < 0,001$ ) et le test de Newmann et Keuls (Tab. 5.6.) montrent que la moyenne des hybrides F1 est similaire à celle de l'espèce PSE, mais diffère de celle de DEW : la trigonelline ne serait pas un caractère additif. Des différences hautement significatives ( $F_{6, 21}$

**Tableau 5.6.** : Teneurs moyennes en trigonelline (en % MS) des espèces parentales, des hybrides F1 et des hybrides backcross (BCDEW et BCPSE). Les différences inter-arbres sont données par les résultats du test F (\*\*\*: très hautement significatif,  $P < 0.001$ ). Les résultats du test de Newmann & Keuls sont indexés avec des lettres.

	DEW	BCDEW	F1	BCPSE	PSE
Moyenne	0,57 <sup>a</sup>	1,08 <sup>b</sup>	1,13 <sup>b</sup>	1,21 <sup>b</sup>	1,02 <sup>b</sup>
Min-Max	0,51-0,66	0,74-1,38	0,87-1,44	1,03-1,52	0,86-1,19
Amplitude	0,15	0,64	0,57	0,49	0,33
F <sub>6, 21</sub>	17,8 ***	6,43 ***	22,3 ***	18,0 ***	17,3 ***

**Tableau 5.7.** : Moyennes des teneurs en acides chlorogéniques (en % de MS) des groupes DEW, BCDEW, F1, BCPSE et PSE. Les résultats des tests de Newmann & Keuls sont indiqués par des lettres en exposant.

	DEW	BCDEW	F1	BCPSE	PSE
CGA	7,97 <sup>a</sup>	5,81 <sup>b</sup>	4,36 <sup>c</sup>	2,50 <sup>d</sup>	1,51 <sup>e</sup>
CQA	6,33 <sup>a</sup>	4,41 <sup>b</sup>	3,34 <sup>c</sup>	1,81 <sup>d</sup>	1,40 <sup>d</sup>
DQA	1,18 <sup>a</sup>	0,97 <sup>b</sup>	0,67 <sup>c</sup>	0,45 <sup>d</sup>	0,04 <sup>e</sup>
FQA	0,465 <sup>a</sup>	0,431 <sup>a</sup>	0,344 <sup>a</sup>	0,236 <sup>ab</sup>	0,075 <sup>b</sup>
CQA/ CGA	0,794 <sup>b</sup>	0,760 <sup>ab</sup>	0,767 <sup>ab</sup>	0,725 <sup>a</sup>	0,925 <sup>c</sup>
DQA/ CGA	0,149 <sup>a</sup>	0,167 <sup>a</sup>	0,155 <sup>a</sup>	0,176 <sup>a</sup>	0,025 <sup>b</sup>
FQA/ CGA	0,058 <sup>a</sup>	0,072 <sup>a</sup>	0,078 <sup>a</sup>	0,099 <sup>a</sup>	0,049 <sup>a</sup>
4-&5-CQA	5,80 <sup>a</sup>	4,15 <sup>b</sup>	3,21 <sup>c</sup>	1,74 <sup>d</sup>	1,37 <sup>d</sup>
3-CQA	0,532 <sup>a</sup>	0,257 <sup>b</sup>	0,134 <sup>c</sup>	0,073 <sup>d</sup>	0,026 <sup>e</sup>
4-&5-CQA/ CQA	0,914 <sup>a</sup>	0,941 <sup>b</sup>	0,959 <sup>c</sup>	0,960 <sup>c</sup>	0,981 <sup>d</sup>
4,5-DQA	0,596 <sup>a</sup>	0,512 <sup>a</sup>	0,325 <sup>b</sup>	0,214 <sup>c</sup>	0,025 <sup>d</sup>
3,5-DQA	0,321 <sup>a</sup>	0,254 <sup>b</sup>	0,227 <sup>b</sup>	0,166 <sup>b</sup>	0,007 <sup>c</sup>
3,4-DQA	0,262 <sup>a</sup>	0,204 <sup>b</sup>	0,122 <sup>c</sup>	0,071 <sup>d</sup>	0,005 <sup>e</sup>
4,5-DQA/ DQA	0,501 <sup>b</sup>	0,527 <sup>b</sup>	0,488 <sup>b</sup>	0,481 <sup>b</sup>	0,657 <sup>a</sup>
3,5-DQA/ DQA	0,273 <sup>b</sup>	0,264 <sup>b</sup>	0,332 <sup>a</sup>	0,371 <sup>a</sup>	0,184 <sup>c</sup>
3,4-DQA/ DQA	0,225 <sup>a</sup>	0,209 <sup>a</sup>	0,180 <sup>ab</sup>	0,148 <sup>b</sup>	0,159 <sup>b</sup>
5-FQA	0,382 <sup>a</sup>	0,296 <sup>a</sup>	0,253 <sup>a</sup>	0,200 <sup>ab</sup>	0,064 <sup>b</sup>
4-FQA	0,049 <sup>b</sup>	0,115 <sup>c</sup>	0,091 <sup>c</sup>	0,036 <sup>b</sup>	0,011 <sup>a</sup>
3-FQA	0,036 <sup>a</sup>	0,020 <sup>b</sup>	0,000 <sup>c</sup>	0,000 <sup>c</sup>	0,000 <sup>c</sup>
5-FQA/ FQA	0,818 <sup>a</sup>	0,662 <sup>b</sup>	0,734 <sup>ab</sup>	0,810 <sup>a</sup>	0,849 <sup>a</sup>
4-FQA/ FQA	0,106 <sup>a</sup>	0,298 <sup>c</sup>	0,266 <sup>c</sup>	0,190 <sup>b</sup>	0,151 <sup>b</sup>
3-FQA/ FQA	0,076 <sup>a</sup>	0,040 <sup>b</sup>	0,000 <sup>c</sup>	0,000 <sup>c</sup>	0,000 <sup>c</sup>

= 22,3 ;  $p < 0,001$ ) existent entre les hybrides F1 avec des teneurs variant de 0,87 à 1,44 % MS (Tab. 5.6.).

Les teneurs moyennes en trigonelline sont de 1,03 % et 1,08 % MS chez BCPSE et BCDEW. Ces valeurs, similaires de celles notées chez PSE et les hybrides F1, diffèrent de celles caractérisant DEW. Ainsi, l'hypothèse d'additivité doit être rejetée.

Des différences très hautement significatives existent entre les arbres à l'intérieur des groupes précédents (Tab. 5.6.).

### 5. 2. 3. Pour les teneurs en acides chlorogéniques

#### Teneur globale en acides chlorogéniques

Les teneurs en CGA des grains verts diffèrent fortement entre les espèces parentales : elle est d'environ 8 % MS chez DEW, soit cinq fois celle de PSE (Tab. 5.7.). Les variations qui existent entre les arbres à l'intérieur des espèces ne se recouvrent pas (6-9 % chez DEW vs. 1,3 -1,8 % chez PSE).

Chez les hybrides F1, la teneur moyenne est CGA (4,36 % MS) est proche de la moyenne parentale (4,74 % MS) et diffère significativement de celles des espèces parentales (Tab. 5.7.).

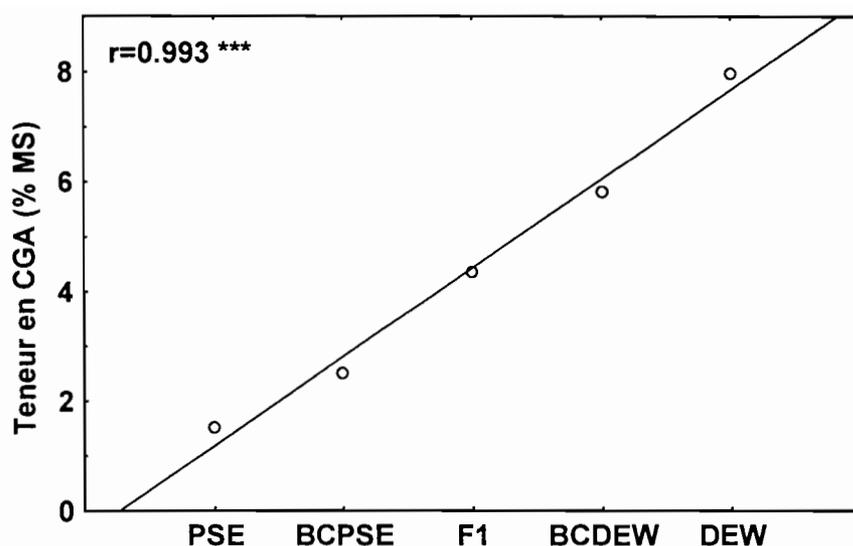
La teneur en CGA chez BCPSE est intermédiaire entre celles de la F1 et de PSE. Il en est de même de la teneur moyenne chez BCDEW par rapport à la F1 et à DEW.

Ceci suggère une hypothèse d'additivité pour ce caractère, que nous pouvons vérifier sur la figure 5.2.

#### Teneurs en acides caféylquiniques

Les CQA représentent respectivement 79 % et 93 % des CGA chez DEW et PSE et ces valeurs sont significativement différentes (Tab. 5.7.). Néanmoins, l'hérédité quantitative de ce taux relatif n'est pas simple. En effet, les hybrides F1 et BCDEW montrent un taux proche de celui de DEW, tandis que le taux observé chez BCPSE (72%) est significativement plus faible que celui de DEW.

A l'inverse, l'hérédité quantitative de la teneur en CQA est plus simple. La teneur en CQA, 4,5 fois élevée chez DEW que chez PSE, montre chez les hybrides F1 une valeur intermédiaire entre les espèces parentales. L'hypothèse d'additivité est confirmée par les valeurs enregistrées chez les deux rétrocroisements (Tab. 5.8.).



**Figure 5.2.** : Régression linéaire illustrant l'additivité de la teneur totale en acides chlorogéniques (CGA). \*\*\* : Très hautement significatif.

**Tableau 5.8.** : Caractères biochimiques vérifiant l'hypothèse d'additivité avec ou sans transformation mathématique des données. Les caractères non additifs n'ont pas été présentés. \*\*, Hautement significatif ( $0.001 < p \leq 0.01$ ); \*\*\*, Très hautement significatif ( $p < 0.001$ ).

	Transformation	r	Signification
CQA	Sans transformation	0,981	**
4-&5-CQA/CQA	Sans transformation	0,965	**
4-&5-CQA	Sans transformation	0,984	**
3-CQA	$y=\sqrt{x}$	0,984	**
DQA	Sans transformation	0,993	***
4,5-DQA	Sans transformation	0,993	***
3,5-DQA	$y=x^2$	0,998	***
3,4-DQA	Sans transformation	0,998	***
FQA	$y=x^2$	0,994	***
5-FQA	$y=x^2$	0,984	**

La teneur groupée des isomères 5- et 4-CQA varie de 1,4 % MS chez PSE à 5,8 % MS chez DEW et l'hypothèse d'additivité s'ajuste bien aux données (Tab. 5.8.). Ces deux isomères représentent 91 % des CQA chez DEW et 98 % chez PSE, les deux valeurs étant significativement différentes (Tab. 5.7.). L'hérédité quantitative de ce taux relatif semble de type additif (Tab. 5.8.).

La teneur en 3-CQA content varie de 0,026 % MS chez PSE à 0,532 % MS chez DEW. Cette teneur n'est pas un caractère additif, mais sa racine carrée l'est (Table 3).

#### Teneurs en acides dicaféylquiniques

Le taux de DQA (DQA/CGA) représente respectivement 15 % et 2,5 % de la fraction totale en acides chlorogéniques dans les grains verts de DEW et PSE et ces deux valeurs sont significativement différentes (Annexe 5.4.). L'hérédité de ce taux en diCQA n'est pas additive. En effet, les moyennes des BCDEW, F1 et BCPSE ne diffèrent pas de la moyenne de DEW. Notons la présence d'un hybride BCPSE ayant un taux relatif de 32 %. Aucune explication n'a pu être donnée à cette exception, cet hybride ayant toutes les caractéristiques de son groupe pour les autres.

Chez les espèces parentales, la teneur moyenne en DQA est près de 30 fois plus élevée chez DEW que chez PSE (Tab. 5.7.). Son hérédité quantitative est plus simple que celle de son taux relatif. Ainsi, la teneur moyenne chez les hybrides F1 est intermédiaire entre celle des parents et l'additivité est confirmée par les valeurs moyennes des groupes BCDEW et BCPSE (Tab. 5.8).

La teneur moyenne de l'isomère 4,5-DQA varie de 0,025 à 0,060 % MS chez les espèces PSE et DEW respectivement (Tab. 5.7.). Les hybrides F1 présentent une moyenne intermédiaire à celles des parents et l'additivité est confirmée par les valeurs des hybrides BCPSE et BCDEW. Le taux relatif de cet isomère (4,5-DQA/DQA) représente respectivement 50 % et 66 % de la fraction totale en DQA chez les espèces DEW et PSE (ces taux sont significativement différents) (Annexe 5.4.). L'additivité n'est pas observée : les valeurs correspondantes aux groupes d'hybrides BCDEW, BCPSE et F1 ne sont pas différentes de celles de DEW.

La teneur moyenne de l'isomère 3,5-DQA est 46 fois plus importante chez DEW que chez PSE (Tab. 5.7.). L'additivité peut être mise en évidence, mais les écarts entre valeurs moyennes attendues et observées ne sont pas négligeables. En revanche, la transformation des valeurs en utilisant la fonction «  $y = x^2$  » conduit à améliorer l'additivité (Tab. 5.8.). Le taux

relatif de cet isomère (3,5-DQA/DQA) représente respectivement 27 % et 18 % de la fraction totale en DQA chez DEW et PSE (ces taux sont significativement différents) (Annexe 5.4.). L'hérédité quantitative de ce taux est complexe : le groupe BCDEW n'est pas différent de celui de DEW et la F1, tandis que le groupe BCPSE montre un taux plus élevé (35 %).

La teneur en 3,4-DQA est 52 fois plus importante chez l'espèce DEW que chez PSE (Tab. 5.7.) et l'additivité est vérifiée pour la teneur moyenne de cet isomère (Tab. 5.8.). Son taux relatif (3,4-DQA/DQA) présente des valeurs similaires, mais légèrement inférieures à celui du 3,5-DQA. L'hypothèse d'additivité peut être acceptée pour les hybrides F1 et BCDEW, mais doit être rejetée pour les hybrides BCPSE.

#### Teneurs en acides férulylquiniques

Les FQA représentent en moyenne 7 % de la teneur totale en acides chlorogéniques. Aucune différence significative de ce taux relatif n'a été décelée parmi les cinq groupes : PSE, BCPSE, F1, BCDEW et DEW (Annexe 5.4.). En revanche, la teneur moyenne en FQA diffère énormément entre DEW et PSE. Comme pour l'isomère 3,5-DQA, l'additivité de la teneur en FQA s'observe après une transformation des données du type «  $y = x^2$  » (Tab. 5.8.).

Le taux relatif en 5-FQA (5-FQA/FQA) est identique chez les deux espèces parentales (Tab. 5.7.). Les hybrides F1 et BCPSE présentent des taux relatifs comparables à celui des parents, alors que celui du groupe BCDEW est significativement inférieur. A l'opposé du taux relatif, la teneur moyenne varie beaucoup entre les espèces, PSE présentant six fois moins de 5-FQA que DEW. Comme pour la teneur moyenne en FQA, l'additivité est acceptée pour le 5-FQA en utilisant la fonction «  $y = x^2$  » (Tab. 5.8.).

Les taux relatifs en 4-FQA (4-FQA/FQA) des espèces parentales ne diffèrent pas significativement (Annexe 5.4.). Cependant, le taux augmente linéairement de PSE (15 %) à BCDEW (30 %). Le même comportement génétique est observé pour la teneur moyenne en 4-FQA qui croît de PSE à BCDEW. L'absence d'additivité est due à une teneur deux fois plus faible chez DEW que celle de BCDEW.

Le 3-FQA est un isomère mineur, absent des groupes PSE, BCPSE et F1. Chez l'espèce DEW, il ne représente que 7,6 % et 0,4 % de la fraction totale en FQA et en CGA (Tab. 5.7.). Cet isomère est cependant intéressant du point de vue génétique : son absence serait contrôlée par un gène majeur dominant. Les arbres sans 3-FQA seraient homozygotes (FF) chez l'espèce PSE, hétérozygotes chez les hybrides F1, homozygotes ou hétérozygotes (Ff ou FF) chez les hybrides BCPSE. Les arbres avec 3-FQA seraient homozygotes récessifs (ff) pour

l'espèce DEW et pour 50% des hybrides BCDEW. Chez ces derniers, on s'attend donc à avoir une valeur moyenne en 3-FQA intermédiaire entre celle des hybrides F1 et de l'espèce DEW, ce qui est effectivement observé (Tab. 5.7).

*N.B.* : Les teneurs en acides chlorogéniques des différents génotypes parentaux figurent en annexe 5.5.

### 5. 3. Relations phénotypiques entre acides chlorogéniques

Aucune relation entre le saccharose, la trigonelline et les acides chlorogéniques n'a été détectée. En revanche, des relations ont été mises en évidence entre les isomères d'acides chlorogéniques. Ces relations sont de deux types : i) entre isomères d'une même classe et ii) entre isomères de classes différentes.

#### 5. 3. 1. Relation à l'intérieur des classes d'acides chlorogéniques

##### Chez les acides caféylquiniques

Une relation curvilinéaire a été mise en évidence entre les teneurs en 3-CQA et en 4-& 5-CQA (Fig. 5.3.a.), représentant statistiquement 89 % de la variance. Un modèle d'allométrie ( $y = ax^b$ ) s'ajuste aux données et son paramètre  $b$ , proche de 2 ( $b = 1,803$ ) montre que la relation est de type quadratique. Une telle relation curvilinéaire explique pourquoi le taux relatif de 4-& 5-CQA varie entre les groupes.

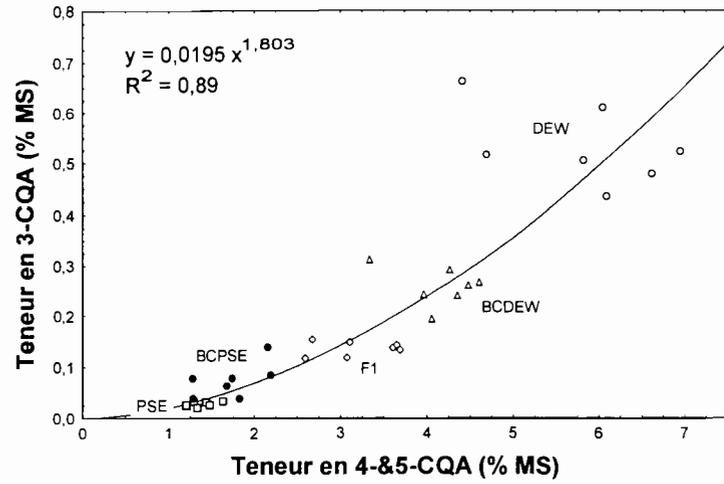
##### Chez les acides dicaféylquiniques

Trois relations ont été mises en évidence entre les trois isomères de la classe des DQA.

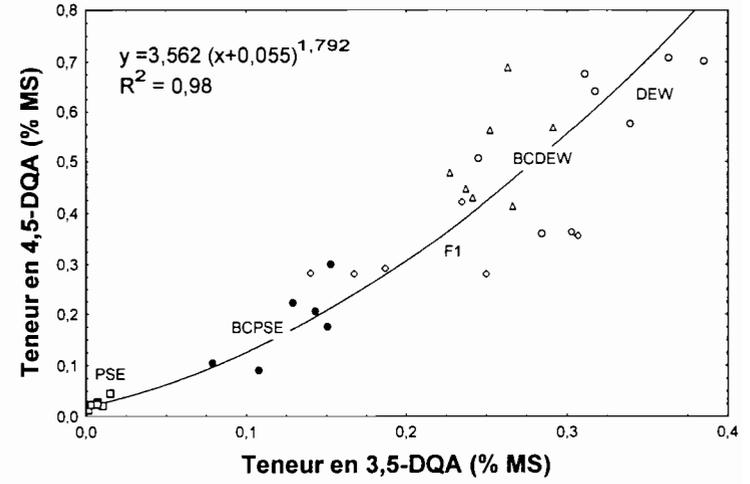
La teneur en 3,5-DQA est liée de façon curvilinéaire à la teneur en 3,4-DQA d'une part (Fig. 5.3.b.) et à la teneur en 4,5-DQA d'autre part (Fig. 5.3.c.). Les courbes de ces deux relations correspondent à une fonction de type :  $y = a(x + c)^b$  avec une valeur du paramètre  $b$  voisine de 2 ( $b = 2,18$  en moyenne pour les deux relations). Ainsi, ces relations seraient de type quadratique. Notons que le paramètre  $c$  montre la même valeur dans les deux relations. Statistiquement, la fonction permet d'expliquer 97 % de la variance. Le modèle suggère ainsi que les isomères 3,4-DQA et 4,5-DQA peuvent être théoriquement présents, même si l'isomère 3,5-DQA est absent des grains.

La troisième relation est linéaire et est observée entre les teneurs en 3,4-DQA et en 4,5-DQA (Fig. 5.3.d.). La linéarité est facilement expliquée lorsque l'on considère les deux relations précédentes comme quadratiques avec  $b = 2$ .

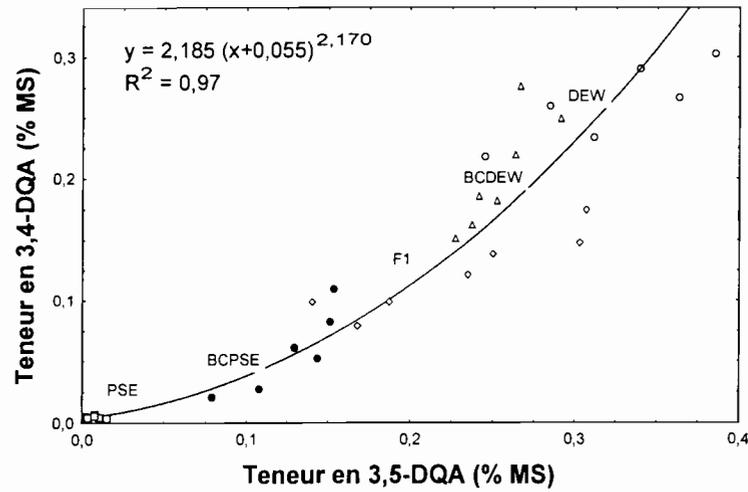
Figure 5.3. a) à f) : Relations entre les différents isomères au sein des 3 classes d'acide chlorogéniques.



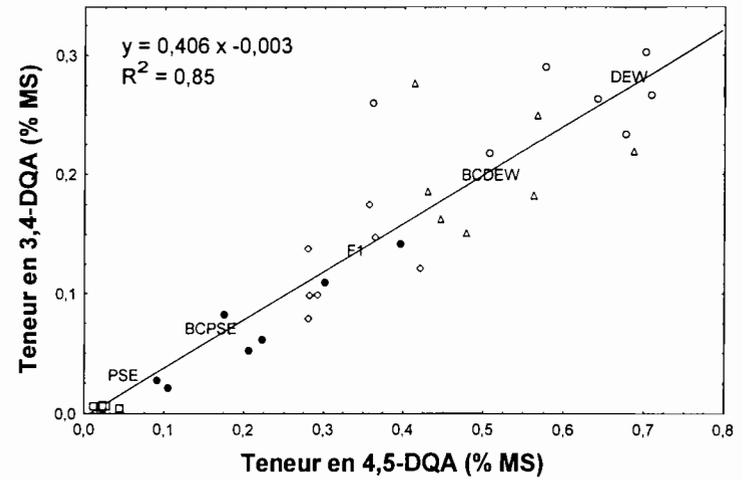
a)



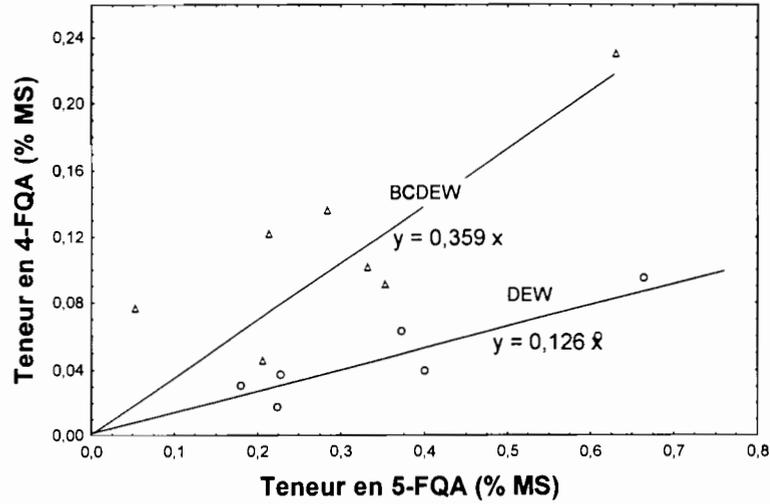
c)



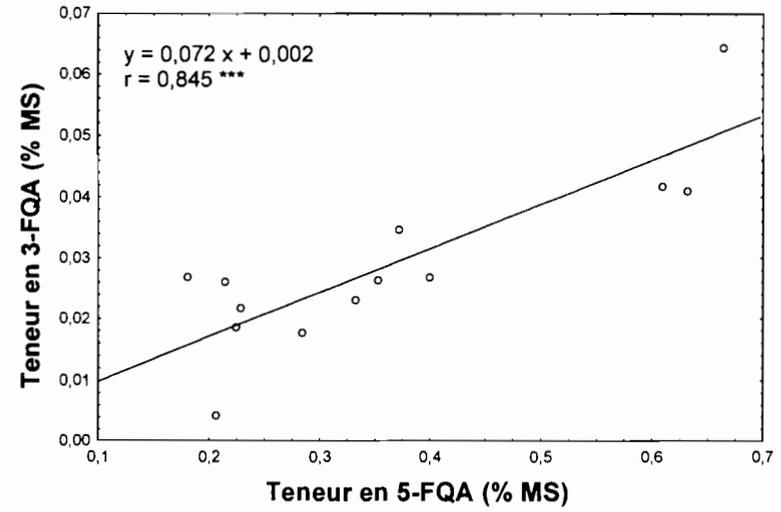
b)



d)

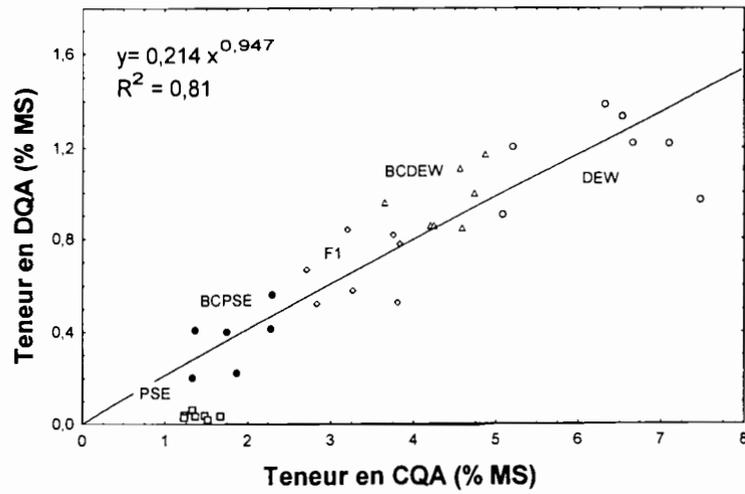


e)

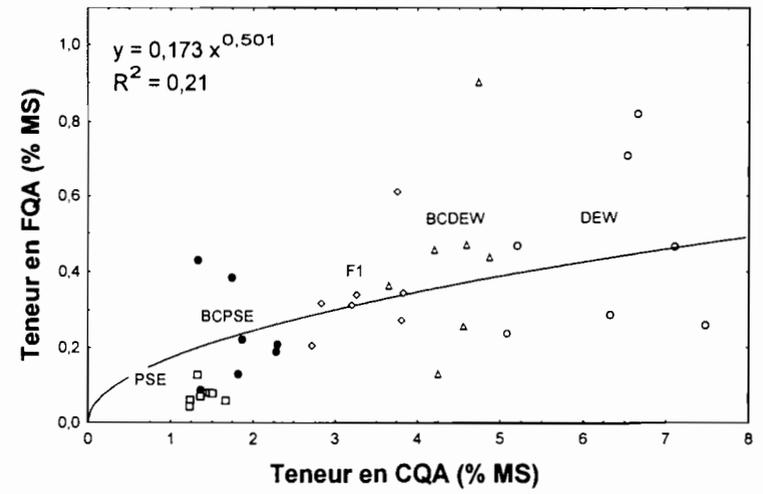


f)

Figure 5.4. a) et b) : Relations entre classes d'acides chlorogéniques.



a)



b)

Les relations quadratiques : i) expliquent pourquoi les trois variables diffèrent entre les espèces et ii) sont en accord avec nos résultats sur l'additivité de la teneur en 3,5-DQA après transformation.

### Teneurs en acides férulylquiniques

Une relation a été détectée entre les teneurs en 5-FQA et 4-FQA. Cette relation est linéaire, mais faible et seulement 37 % des variations de teneurs en 4-FQA sont statistiquement expliquées par celles en 5-FQA. En fait, il y a une absence de relation au sein des groupes PSE, BCPSE et F1 et la présence de deux relations différentes chez les hybrides BCDEW et chez les espèces DEW, groupes présentant les teneurs les plus élevées en 5-FQA. La relation est trois fois plus importante dans le groupe BCDEW que dans le groupe DEW (Fig. 5.3.e.).

Une relation linéaire est aussi observée entre les teneurs en 3-FQA et 5-FQA chez les arbres produisant de l'isomère 3-FQA, c'est à dire pour les groupes DEW et BCDEW. Les deux relations intra-groupes ne diffèrent pas pour la valeur de la pente ( $F_{1,9} = 0,11$  ;  $p = 0,75$ ) ou pour la valeur d'intersection ( $F_{1,10} = 3,00$  ;  $p = 0,11$ ). Une relation commune peut donc alors être estimée (Fig. 5.3.f.), dans laquelle la valeur d'intersection peut être considérée comme égale à zéro ( $t = 0,484$  ;  $ddl = 11$ ,  $p = 0,637$ ).

### **5. 3. 2. Relations entre classes d'acides chlorogéniques**

Une relation est observée entre les teneurs en DQA et CQA à l'intérieur des différents groupes, excepté pour PSE (Fig. 5.4.a.). Le modèle d'allométrie [ $y = ax^b$ ] explique statistiquement 81 % de la variance et le paramètre b est proche de 1 ( $b = 0,95$ ). Le modèle linéaire semble donc approprié.

Une seconde relation est mise en évidence entre les teneurs en FQA et CQA (Fig. 5.4.b.). Deux ajustements ont été appliqués : i) avec seulement les groupes BCPSE, F1, BCDEW et DEW en utilisant la fonction  $y = a(x)^b$  et ii) avec l'ensemble des cinq groupes en utilisant la fonction  $y = a(x + c)^b$ . Ces deux ajustements donnent une valeur du paramètre b similaire et proche de 0,5 (0,437 et 0,501 respectivement). La relation entre FQA et CQA semble être de type « racine carrée ». Ceci est en accord avec le fait que la teneur en FQA doit être transformée en utilisant la fonction  $y = x^2$  pour être additive et que la teneur en CQA est additive sans aucune transformation mathématique.

## 5. 4. Discussion

### 5. 4. 1. Importance relative des variations génotypiques et inter-annuelles

#### La teneur en saccharose

##### *Présence d'une interaction « arbre x année »*

La teneur finale en saccharose des grains verts résulte d'un long processus biochimique, où plusieurs paramètres externes sont impliqués : i) du fait que le saccharose est un composant majeur issu de la photosynthèse, sa teneur devrait varier lorsque les conditions climatiques présentes lors de la fructification changent (Huber, 1989); ii) du fait que l'accumulation est maximale chez la graine mature, la teneur devrait varier lorsque la récolte a lieu à différents niveaux de maturité (Petnga, 1986) ; et iii) la teneur peut aussi baisser après la récolte en fonction des conditions de séchage et de stockage (les températures élevées favorisent le développement des microorganismes qui dégradent le saccharose) (Guyot *et al.* 1995; Pokorny, 1975). Ces différents facteurs externes pourraient expliquer des variations de teneurs entre années, lorsque les périodes de fructification sont synchrones pour tous les arbres et lorsque les traitements après récolte sont similaires pour une même année, mais différents entre années.

En réalité, nous n'observons pas de différences entre années, mais une interaction signifiant que les facteurs externes influençant la teneur en saccharose changent entre arbres. En fait, la durée de fructification des espèces parentales est respectivement de trois et dix mois pour PSE et DEW, et une large variation existe pour ce caractère chez les hybrides de seconde génération. En conséquences, les conditions climatiques, de la floraison au stockage de la graine, diffèrent pour chaque hybride expliquant l'interaction.

##### *Importance des effets génotypiques*

Les différences entre espèces implantées sur plusieurs parcelles pourraient être interprétées comme résultant de variations de fertilité du sol. Néanmoins, la régression linéaire, que nous avons utilisé pour mettre en évidence l'additivité, ne peut être obtenue avec une hypothèse environnementale, les parcelles de PSE, de DEW, des F1 et des rétrocroisements étant réparties au hasard sur la station. Ceci souligne l'importance des effets génétiques dans les différences interspécifiques.

A l'intérieur des groupes (PSE, DEW, F1, BCPSE et BCDEW), les différences enregistrées entre les arbres constituent la principale source de variation. Des résultats

similaires ont été notés chez l'abricot, *Prunus armeniaca* L., où une teneur constante a été observée pendant trois ans pour six parents et leurs descendants (Bassi *et al.*, 1996).

Chez *Prunus*, comme chez *Coffea*, l'importance de la variation notée entre les arbres de la seconde génération est due à la grande différence de teneur des espèces parentales. Nous attendons une diminution de cette contribution lors des générations ultérieures de rétrocroisement.

#### La teneur en trigonelline

Les variations de teneurs en trigonelline ne sont expliquées que par les différences entre hybrides. Aucun effet environnemental et interaction n'influence cette teneur. En d'autres termes, les conditions climatiques qui siègent lors de la maturation des grains, ainsi que les conditions des traitements post-récolte, ne modifient pas la teneur en trigonelline. Cette très forte héritabilité *s.l.* rend possible la sélection clonale des hybrides pour une teneur donnée en trigonelline.

#### La teneur en acides chlorogéniques

A l'inverse des deux précédents composés biochimiques étudiés, les effets environnementaux et les interactions sont relativement importants chez les acides chlorogéniques. En effet, les effets génétiques ne représentent au mieux que 78 % de la variation.

La fertilisation et la nature des sols jouent un rôle important sur les variations de teneurs en acides chlorogéniques, comme cela a pu être constaté chez les plants de pomme de terre ayant reçu un fort apport en engrais potassique (Mondy *et al.*, 1967). Les acides chlorogéniques sont aussi très sensibles aux températures élevées et des pertes importantes peuvent être relevées durant la phase de séchage et de stockage des grains. C'est en particulier le cas avec les isomères de la classe des DQA (Guyot, comm. pers.). Ceci explique la part relativement importante de l'effet des traitements post-récolte (« année ») sur les teneurs en DQA par rapport aux isomères monosubstitués (CQA et FQA) (Tab. 5.3.).

### **5. 4. 2. Hérité quantitative des composés biochimiques**

Jusqu'à présent, aucune donnée n'a été présentée sur l'hérité des teneurs en saccharose, en trigonelline et en acides chlorogéniques au sein d'un croisement interspécifique de caféiers. En revanche, les paramètres génétiques concernant ces mêmes

caractères biochimiques ont été évalués au sein d'un croisement intraspécifique chez *C. canephora* entre arbres d'origine congolaise et guinéenne (Montagnon *et al.*, 1998).

#### Teneur en saccharose

L'hérédité additive de la teneur en saccharose a été notée sur le croisement intraspécifique étudié par Montagnon *et al.* (1998). L'exception observée chez les hybrides BCPSE peut être expliquée par une sélection des hybrides BCPSE à faible teneur en saccharose. Les hybrides BCPSE à fortes teneurs (Fig. 5.1.) seraient contre sélectionnés à la suite de distorsions de ségrégation.

#### Teneur en trigonelline

Le déterminisme génétique de la teneur en trigonelline pourrait être de type cytoplasmique (gènes mitochondriaux ou chloroplastiques) à hérédité maternelle. Cette hypothèse est confortée par le fait que la teneur en trigonelline présente des valeurs similaires entre les groupes PSE, BCPSE, F1 et BCDEW (Tab. 5.6.), qui ont tous un cytoplasme de type PSE. Pour obtenir confirmation, il faudrait disposer soit i) d'hybrides F1 générés à partir de DEW comme parent femelle, mais ce croisement réciproque DEW x PSE ne peut être obtenu, soit ii) d'hybrides BCDEW issus du croisement DEW x F1 ; ces arbres existent et devraient produire des graines au cours de l'année 2000.

Avec une hérédité exclusivement cytoplasmique, les teneurs observées chez les hybrides F1 devraient être homogènes, ce qui n'est pas le cas. L'amplitude de variation de 0,87 à 1,44 % MS suggère l'existence d'effets génétiques de type nucléaire découlant de l'hétérozygotie des espèces parentales. Dans le croisement intraspécifique, Montagnon *et al.* (1998) observe une héritabilité *s.s.* de 34 % impliquant une additivité et, *ipso facto*, l'existence de gènes nucléaires.

#### Teneurs en acides chlorogéniques

L'hérédité quantitative des acides chlorogéniques montrent trois types de comportements génétiques :

- i) l'hérédité est de type additif sans transformation mathématique des données. Les caractères concernés sont : le CGA, le CQA, le DQA, le 5-CQA, le 4,5-DQA et le 3,4-DQA ;
- ii) l'hypothèse d'additivité a été rejetée, mais une transformation mathématique des données rend le caractère additif. C'est le cas du 3-CQA avec la transformation  $y = \sqrt{x}$ , mais aussi du 3,5-DQA, du FQA et du 5-FQA avec la transformation  $y = x^2$ . Notons qu'un modèle additif après transformation des données a été observé dans ce même croisement pour la racine

carrée de la teneur en caféine (Barre *et al.*, 1998). Dans le cas des acides chlorogéniques, la nécessité d'une transformation est directement liée à l'existence de relations curvilinéaires entre isomères, dont l'interprétation sera discutée dans le paragraphe correspondant (5.5.3.) ;  
 iii) l'additivité n'est pas ou que partiellement observée. C'est le cas de la majorité des taux relatifs qui ne présentent pas d'additivité, à l'exception du taux de 4-&5-CQA dans les CQA. C'est aussi le cas pour l'isomère 3-FQA, où l'additivité est vérifiée pour les hybrides F1, BCDEW et l'espèce parentale DEW et où l'absence est gouvernée par un gène majeur dominant.

L'interprétation physiologique de cette absence d'accumulation pourrait être liée à : i) l'absence d'enzyme responsable de sa biosynthèse ; ii) l'absence d'un système de transport vers la graine ; et iii) la présence d'enzymes responsable de sa transformation. Le caractère dominant de l'absence est en faveur de la troisième hypothèse.

### 5. 4. 3. Relations entre isomères d'acides chlorogéniques

Les relations mises en évidence entre les différents isomères d'acides chlorogéniques permettent de préciser les voies de biosynthèse présentées dans le chapitre 2.

#### Relations à l'intérieur des classes d'acides chlorogéniques

Ces relations semblent résulter de processus d'isomérisation. Pour les classes CQA et DQA, nos résultats sont en accord avec les chaînes connues de biosynthèse. Il en est de même pour les relations observées entre le 4-FQA et le 5-FQA chez BCDEW et DEW et dont l'absence chez PSE, BCPSE et F1 peut être expliquée par une moindre précision dans les dosages en raison de la faible teneur en 4-FQA chez ces arbres. Ceci est aussi vérifié pour la relation entre le 3-FQA et le 5-FQA (l'isomère 3-FQA n'est absent dans les groupes PSE, BCPSE et F1).

L'existence de relations linéaires et de relations curvilinéaires au sein d'une même classe d'acides chlorogéniques est un point important de discussion. Dans la classe des DQA, les isomères 4,5-DQA et 3,4-DQA sont reliés au 3,5-DQA par une relation quadratique, tandis que l'isomère 3,4-DQA est lié au 4,5-DQA par une fonction linéaire. Ces dernières sont caractéristiques des réactions enzymatiques à l'équilibre : le rapport substrat/produit y est en effet constant. Une déviation à la linéarité suggère que l'isomère 3,5-DQA serait impliqué dans d'autres réactions et c'est effectivement le cas d'après les données de la littérature : il réagit avec la caféine afin de former un complexe (Horman et Viani 1971 ; Kappeler *et al.*, 1987 ; Baumann et Röhrig, 1989). La relation curvilinéaire entre le 3-CQA et le 4- & 5-CQA

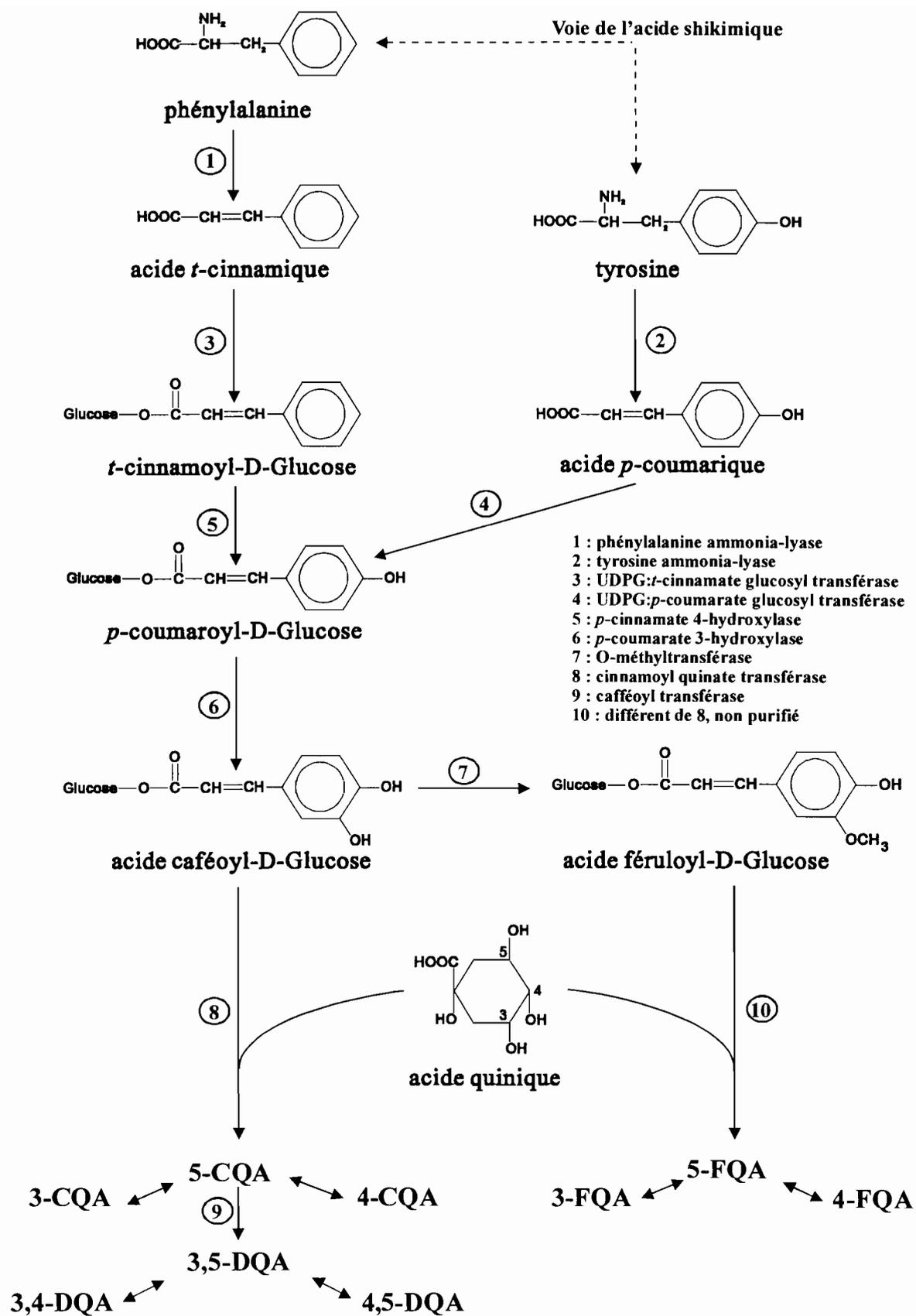


Figure 2.2.5.: Chaîne de biosynthèse conduisant à la formation des acides chlorogéniques.

peut être interprétée de la même façon : l'isomère 5-CQA majoritaire dans le pic 4-& 5-CQA (Clifford 1985 b) se complexe aussi avec la caféine (Baumann *et al.*, 1993). A l'opposé, la relation linéaire entre le 4,5-DQA et le 3,4-DQA suggère que ces derniers ne se complexent pas avec la caféine ou seulement à faible dose. Cette interprétation explique aussi l'existence d'une additivité après transformation pour les teneurs et l'absence d'additivité pour les taux relatifs

#### Relations entre classes d'acides chlorogéniques

Une relation entre CQA et DQA est attendue d'après les voies de biosynthèses. En effet, la totalité des isomères de DQA proviennent des CQA par ajout d'une unité caféate. En revanche, la relation entre CQA et FQA est indirecte : les FQA ne dérivent pas des CQA, mais la synthèse du 5-CQA et du 5-FQA nécessite un substrat commun, l'acide quinique. Cet effet indirect est confirmé par le fait que la relation CQA – FQA est plus faible que celle du CQA – DQA. L'absence de relation entre FQA et DQA est en accord avec leurs voies plus éloignées de biosynthèse.

Ces deux relations inter-classes ne sont pas linéaires. Comme pour les relations internes à une classe de CGA, la concavité suggère que les DQA et FQA soient impliqués dans d'autres voies de biosynthèses. Effectivement, les FQA interviennent dans la biosynthèse des lignines (Hahlbrock et Grisebach, 1979), tandis que les DQA jouent un rôle de détoxification de l'acide caféique (Colonna, 1979).

### **5. 5. En résumé**

Cette approche génétique, suivant un modèle équilibré à sept génotypes pour chacun des cinq groupes PSE, BCPSE, F1, BCDEW et DEW a permis à la fois : i) de déterminer les paramètres de génétique quantitative (effet année, additivité) et ii) de préciser les voies de biosynthèse des acides chlorogéniques, en relation avec les corrélations phénotypiques établies.

Les résultats majeurs qui se dégagent de cette étude sont les suivants :

- i) La teneur en saccharose est additive (exceptée pour les hybrides BCPSE) (Fig. 5.1.). Les variations de teneur en saccharose sont essentiellement dues aux différences entre hybrides. L'interaction « année de récolte x génotype » ne représente que 6,3 % de la variance totale. Aucune différence n'a été détectée entre les deux années de récolte ;

- ii) La teneur en trigonelline n'est pas un caractère additif. Chez l'espèce sauvage, la teneur en trigonelline présente une teneur (1 % MS) double de celle de DEW. Une hypothèse d'hérédité de type cytoplasmique reste à confirmer. Mais des gènes nucléaires sont sans doute impliqués à un niveau moindre ;
- iii) Les teneurs en acides chlorogéniques sont les plus sensibles aux conditions environnementales. La part des variations expliquée par les différences entre hybrides varie de 43 à 78 % ;
- iv) L'absence en isomère 3-FQA est gouvernée par un gène majeur dominant ;
- v) L'additivité, avec ou sans transformation mathématique des variables, est vérifiée pour l'ensemble des isomères et les trois classes (CQA, DQA et FQA), excepté pour l'isomère 4-FQA. En revanche, les taux relatifs ne suivent pas un modèle additif ;
- vi) Les relations observées entre isomères sont attendues des voies connues de biosynthèse et de réaction avec la caféine.

Les analyses préalables d'hérédité quantitative sont importantes dans les schémas d'amélioration : l'additivité d'un caractère permet de prédire la teneur moyenne attendue dans une descendance, à partir de géniteurs. Le fait que les variations sont en majorité expliquées par les différences entre hybrides permet une sélection de type clonale et une micro propagation de clones élites. L'existence d'interaction de type « année x génotype » suggèrent que les observations doivent être effectuées sur plusieurs années de récoltes pour une bonne estimation des valeurs phénotypiques. Ceci est surtout valable pour les isomères de la classes des DQA.

La cartographie génétique et la recherche de QTLs pour ces caractères constituent les étapes suivantes à cette étude et font l'objet des chapitres 6 et 7.

**Chapitre 6 : Cartographie génétique du  
croisement interspécifique PSE x DEW**

---

## Chapitre 6 : Cartographie génétique du croisement interspécifique PSE x DEW

Dans ce chapitre, nous présenterons la carte de liaison génétique de marqueurs majoritairement de type AFLP obtenue à partir de la descendance BCDEW (62 hybrides). L'origine des distorsions de ségrégation observées, caractéristiques des croisements interspécifiques sera discutée.

*N.B.* : Ces résultats ont été valorisé (Ky *et al.*, 2000) et la publication présentée en annexe D.

### 6. 1. Résultats

La construction de la carte repose sur une cartographie de données simulées reprenant les caractéristiques majeures de notre population : faible taille de population (62 hybrides) et présence d'une part non négligeable de marqueurs distordus. Cette simulation a été entreprise afin d'adopter une stratégie de cartographie minimisant au maximum les risques de détecter de fausses liaisons entre marqueurs.

#### 6. 1. 1. Simulation de cartographie

Afin de connaître la démarche et les valeurs de *LOD score* à appliquer pour la cartographie du croisement PSE x DEW, un fichier simulé a été construit (voir Chp. 3).

##### Cartographie des marqueurs groupe par groupe

Les 40 marqueurs sans distorsions correspondant aux quatre groupes GF-1 à GF-4 se cartographient sur quatre groupes de liaison pour des valeurs de *LOD score* inférieures à 14,9. Chaque groupe de liaison compte les dix marqueurs attendus. La longueur de ces groupes de liaison est de 29,1 cM et la distance entre deux loci de 3,2 cM. Pour une valeur de *LOD score* supérieure ou égale à 15, l'ensemble des marqueurs se retrouve non liés.

Deux groupes de liaisons ont aussi été obtenus à partir des groupes simulés de marqueurs distordus GF-5 et GF-6 avec des valeurs de *LOD score* inférieures 14,9. La longueur des groupes de liaison et la distance entre loci est la même que pour les groupes de marqueurs non-distordus. De même, les marqueurs se retrouvent dissociés avec un *LOD score* supérieur à 14,9.

Le groupe GF-7 doit théoriquement aboutir à dix marqueurs indépendants. Un tel résultat est vérifié pour des valeurs de *LOD score* supérieures à 5,4. De fausses liaisons sont mises en évidence en dessous de 5,5.

Enfin pour GF-8, c'est à une valeur de *LOD score* bien supérieure à 14,9 que les marqueurs du groupe se retrouvent complètement dissociés. Avec une valeur de *LOD score* de 5,5, de pseudo-groupes de 2 à 3 marqueurs apparaissent. Ceci pose le problème du codage des bandes superposées qui conduisent à de pseudo-distorsions de type 3 : 1.

#### Cartographie des marqueurs mixant des groupes non-distordus avec des groupes distordus

Cinq groupes de liaison de dix marqueurs ont été trouvés, comme attendu, à partir des 50 marqueurs appartenant à GF-1, GF-2, GF-3, GF-4 et GF-5 et ceci pour des valeurs de *LOD score* strictement inférieures à 14,9. Ils correspondent respectivement aux cinq groupes simulés. La longueur de chaque groupe de liaison est de 29,1 cM et la distance entre deux loci consécutifs de 3,2 cM.

Lorsque les marqueurs distordus du groupe GF-6 sont ajoutés aux 40 marqueurs non-distordus des groupes GF-1 à GF-4, quatre groupes de liaisons sont obtenus comme attendu (GF-6 est théoriquement lié à GF-1) pour des valeurs de *LOD score* comprises entre 0 et 4,2. Le groupe de liaison GL-1 est constitué de 20 marqueurs (longueur totale de 82,5 cM) distants de 3,2 cM, exceptés pour les deux marqueurs liant GF-1 et GF-6, où la distance est de 24,3 cM. Pour des valeurs de *LOD score* comprises entre 4,3 et 14,8, le groupe GL-1 redonne deux groupes de liaison indépendants. Au-delà d'une valeur de 14,8, tous les marqueurs sont indépendants.

La cartographie concernant la combinaison des groupes GF-1 à 4 avec les marqueurs du groupe GF-7 redonnent les quatre groupes de liaison et les dix marqueurs indépendants attendus lorsque le *LOD score* est compris entre 5,5 et 14,8. Bien évidemment au-delà d'une valeur de *LOD score* de 14,8, tous les marqueurs deviennent indépendants.

Les résultats obtenus pour la combinaison des groupes GF-1 à 4 avec GF-8 correspondent à une combinaison des résultats obtenus lors de la cartographie des quatre

groupes non distordus, d'une part, et du groupe GF-7, d'autre part. C'est au-delà de 4,9 que les marqueurs de GF-8 commencent à se dissocier des quatre groupes de liaison constitués par les marqueurs non distordus. A partir de 5,5, les quatre groupes de dix marqueurs non-distordus sont retrouvés avec à leur côté, de petits groupes de 2 à 3 marqueurs venant de GF-8, ainsi que quelques marqueurs indépendants. Encore une fois, au-delà d'une valeur de *LOD score* de 14,9, l'ensemble de tous les marqueurs sont non liés.

### En résumé

Cette cartographie sur données simulées montre que :

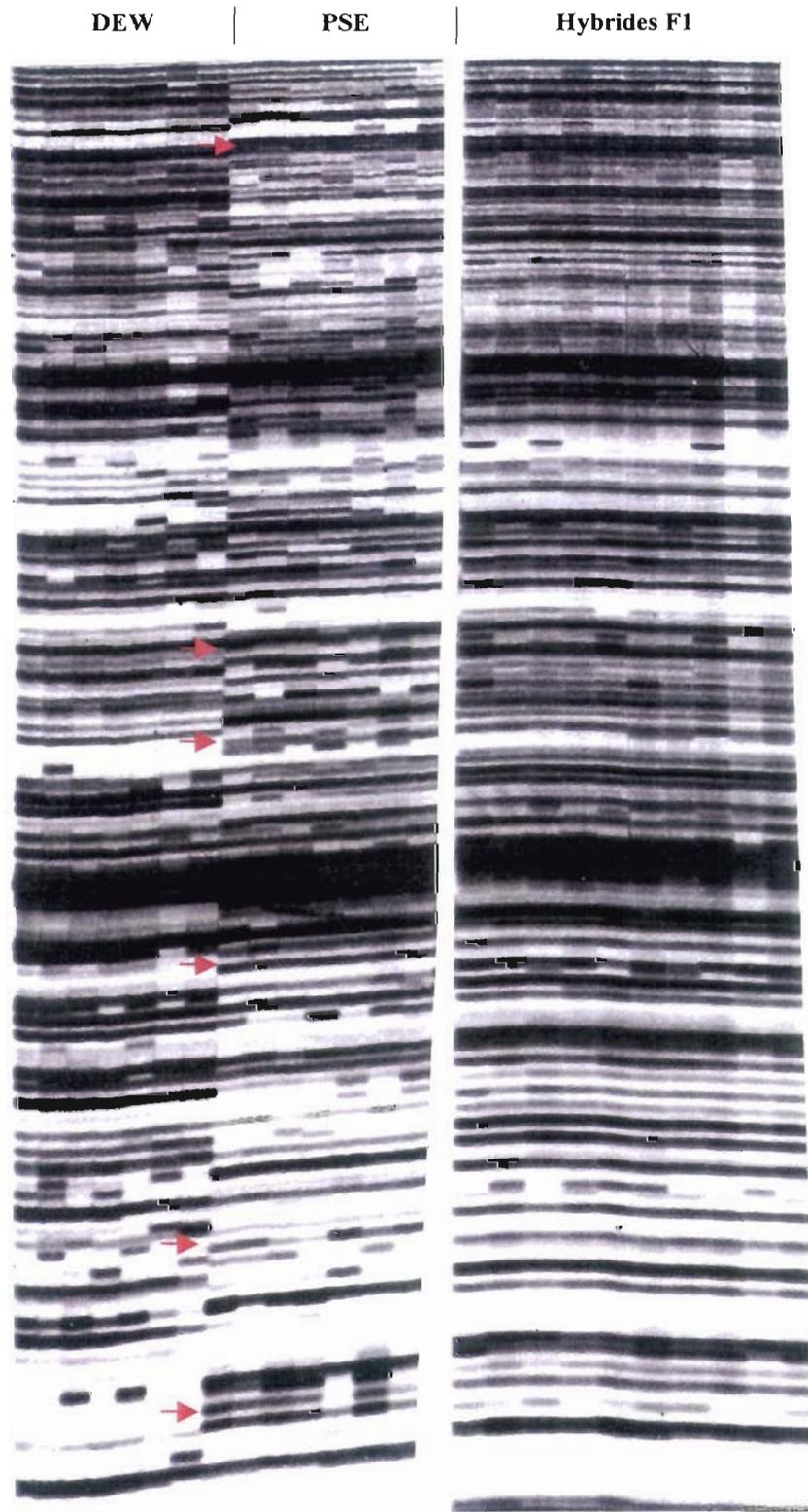
- i) La cartographie doit être effectuée dans un premier temps en séparant les marqueurs non distordus des marqueurs distordus ;
- ii) la valeur seuil de *LOD score* est alors fixée à 5,5 afin de minimiser la présence de fausse liaison. Néanmoins, les petits groupes de deux à trois marqueurs distordus de type 3 : 1 devront être interprétés comme de pseudo-groupes résultant du codage de bandes superposées. Cette valeur seuil permet d'obtenir un premier ensemble de groupes de liaison provisoires de forte vraisemblance. Une deuxième valeur seuil (4,8) est ensuite utilisée pour compléter l'ensemble.
- iii) Les différents segments obtenus à l'étape précédente sont ensuite agrégés à une valeur seuil de 3,1.

## **6. 1. 2. Cartographie génétique du croisement PSE x DEW**

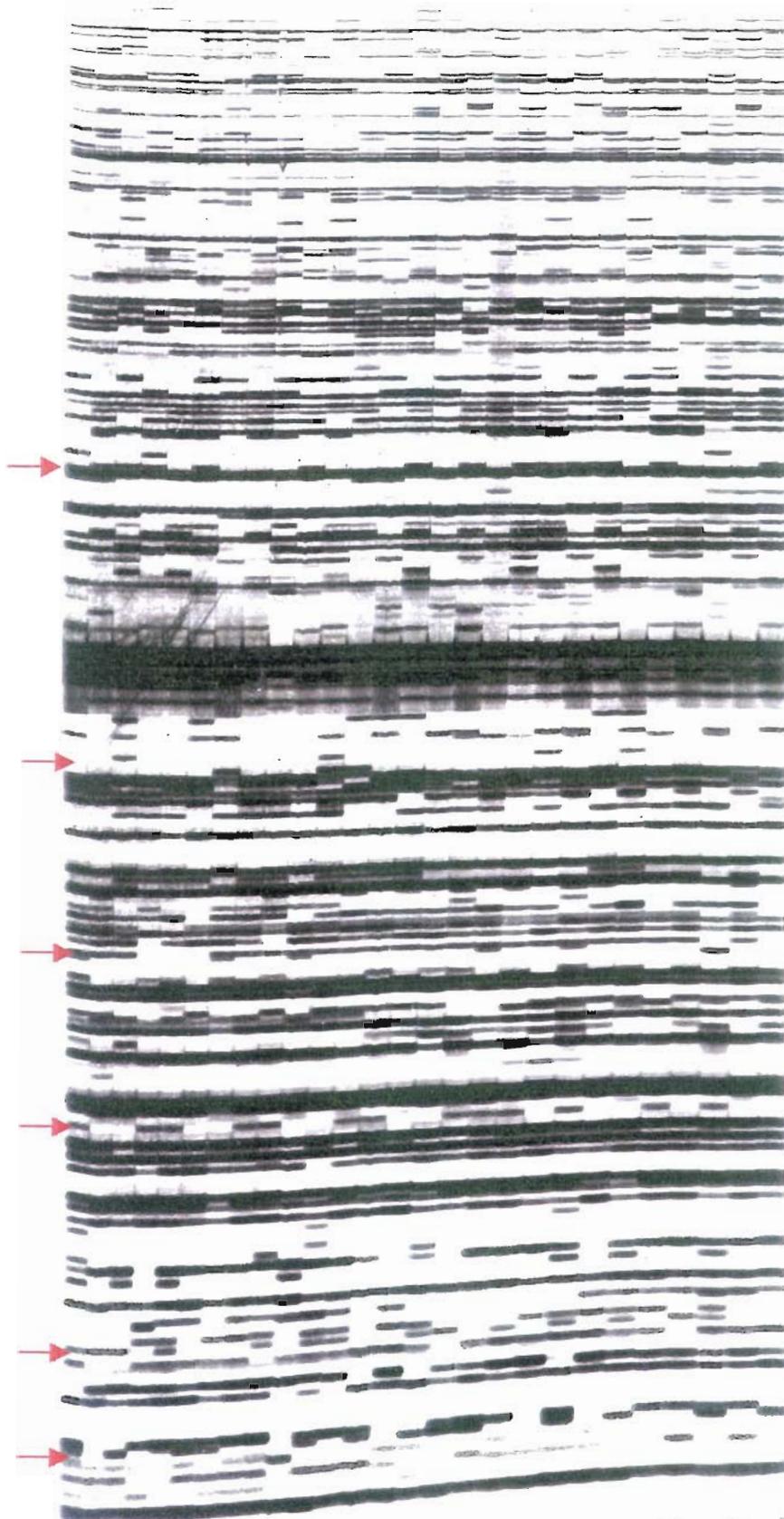
### Polymorphisme des marqueurs AFLP

Un total de 195 bandes AFLP polymorphes, spécifiques de l'espèce PSE a été retenu sur les 60 gels analysés. L'annexe 6.1. illustre les 30 combinaisons d'amorces qui ont permis de générer ces bandes. Le polymorphisme varie de 1 à 12 bandes en fonction du couple d'amorces considérées. De ces 195 marqueurs AFLP, trois ont été écartés des analyses en raison d'une lecture difficile de leur ségrégation, qui aurait pu introduire des erreurs de cartographie et nuire à une analyse ultérieure de QTL.

Les figures 6.1. et 6.2. illustrent respectivement le choix des bandes polymorphes : présentes chez PSE et les hybrides F1 et absentes chez DEW (voir Chp. 3), et la ségrégation de ces mêmes bandes chez les hybrides BCDEW.



**Figure 6.1.** : Autoradiogrammes illustrant la sélection de 6 marqueurs AFLP spécifiques de l'espèce *C. pseudozanguebariae*, absents des 7 géotypes de l'espèce *C. liberica* var. *dewevrei* et présents chez les 10 hybrides F1.



**Figure 6.2.** : Autoradiogramme montrant la ségrégation de 6 marqueurs AFLP chez 27 hybrides BCDEW (combinaison d'amorce *Eco* ACC / *Mse* CTT). Les flèches rouge localisent les bandes prises en considération pour la cartographie.

### Carte génétique

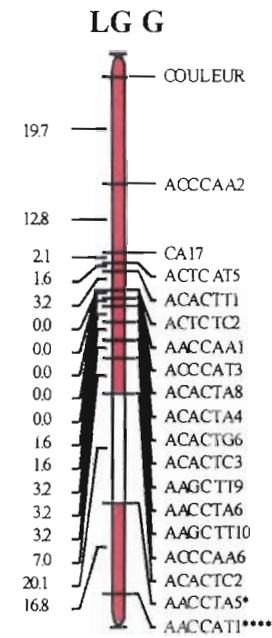
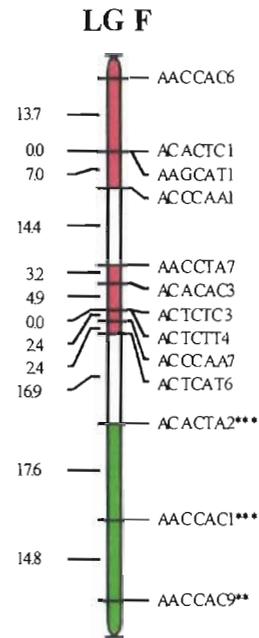
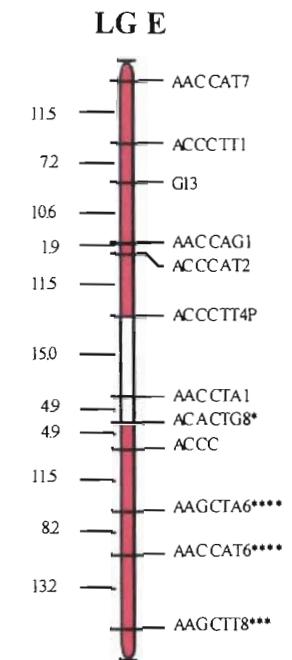
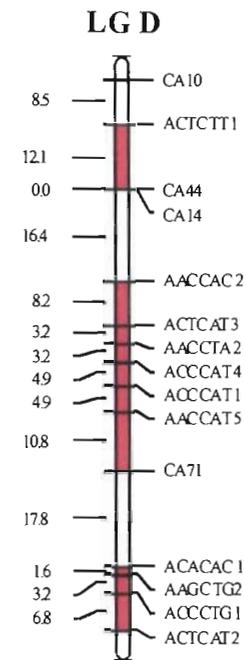
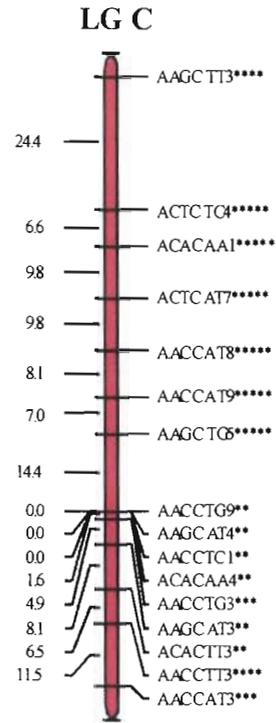
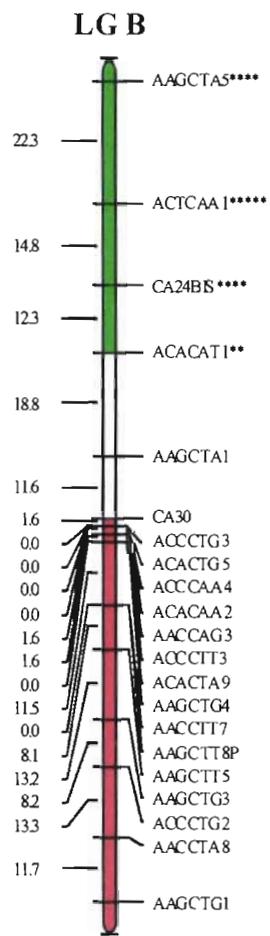
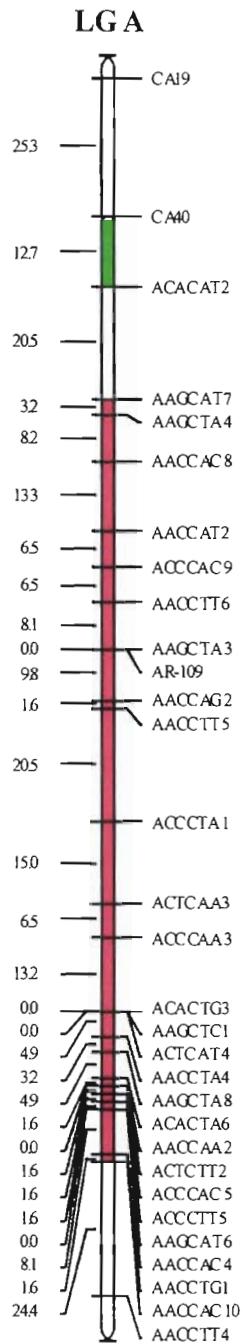
Cent quatre-vingt douze marqueurs AFLP ont servi pour la cartographie des 62 hybrides BCDEW. Des travaux antérieurs sur cette même population ont permis d'y adjoindre 16 marqueurs RFLP (Barre, 1997) et un marqueur morphologique : la couleur du fruit à maturité (Akaffou, 1999).

Avec une valeur de *LOD score* égale à 5,5 et appliqués aux marqueurs non distordus, 15 groupes provisoires de liaison (segments rouges non étoilés \*, sur la figure 6.3.) et 14 marqueurs indépendants sont détectés. En ce qui concerne les marqueurs distordus, la même valeur du *LOD score* conduit à quatre groupes provisoires de liaison (segments rouges étoilés \*, sur la Fig. 6.3.) et 24 marqueurs indépendants. En abaissant la valeur du *LOD score* à 4,8 (valeur indiquée Fig. 6.3.), de nouveaux groupes provisoires de liaison apparaissent (segments verts étoilés et non étoilés \*, sur la Fig. 6.3). Enfin, la fusion des groupes provisoires a lieu avec la valeur 3,1 de *LOD score*. En définitive, parmi les 192 marqueurs AFLP, vingt-cinq demeurent indépendants. De même, sur les 16 marqueurs RFLP, trois ne sont pas cartographiés.

La carte génétique obtenue est constituée de 14 groupes de liaison, comprenant 4 à 31 locus (Fig. 6.3.). Le tableau 6.1. résume les principales caractéristiques de chaque groupe. Ceux-ci, nommés par des lettres (de A à N) en fonction de leurs longueurs, se répartissent ainsi en trois catégories : i) les groupes de liaison de longueur comprise entre 150 et 225 cM et présentant de 21 à 31 marqueurs (groupes A et B) ; ii) les groupes de longueur comprise entre 48 et 113 cM et comprenant 11 à 19 marqueurs (groupes C à J) ; et iii) les groupes de petite taille (18 à 27 cM) avec 4 à 6 marqueurs (groupes K à N).

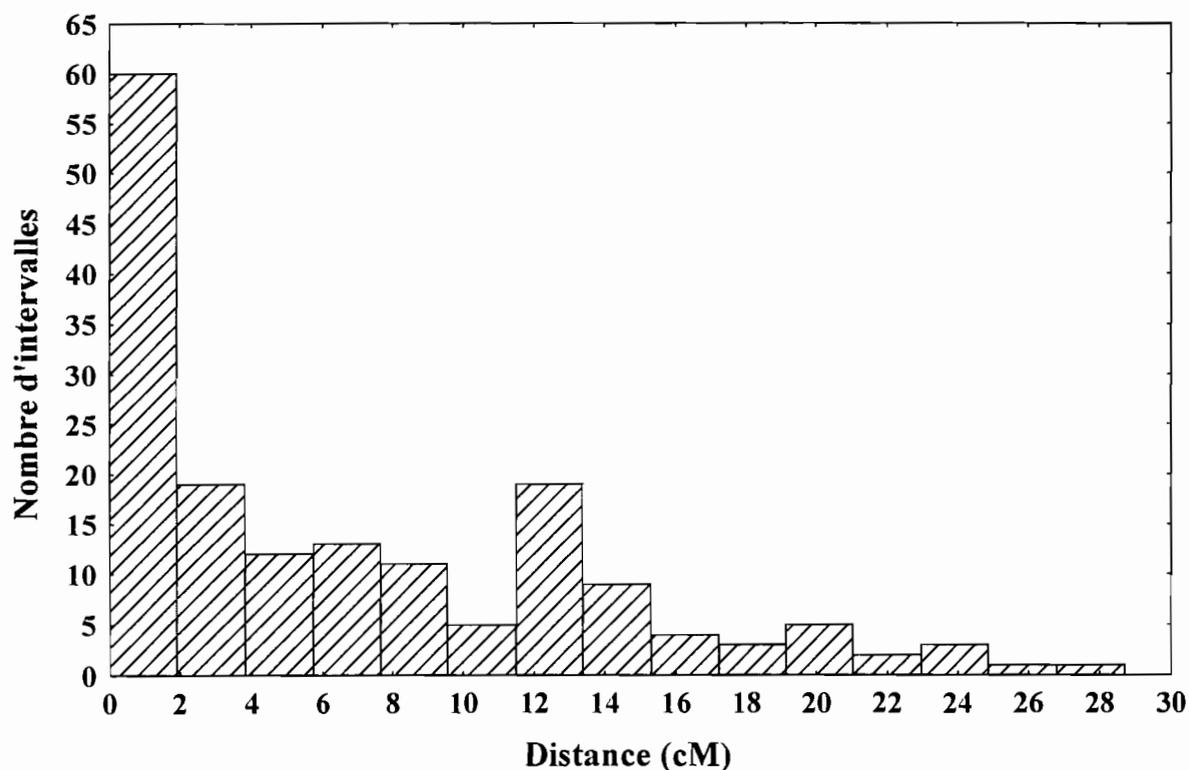
La répartition des marqueurs AFLP sur les groupes de liaison révèle des zones de haute densité (cluster), dans lesquelles la liaison est parfois totale (0 cM). On distingue aisément huit clusters sur la carte génétique (Fig. 6.3.) situés sur les groupes de grande et de moyenne taille (excepté les groupes D et E). La densité de marqueurs au niveau de ces clusters varie de 6 à 15 pour une zone limitée inférieure à 20 cM, soit environ un marqueur tous les 0,5 cM.

La taille de la carte, construite à partir de la population BCDEW, est de 1144 cM (addition des distances obtenues pour chaque groupe de liaison) et la distance entre marqueurs va de 0 à 28,7 cM. L'espacement moyen entre marqueurs est de 7 cM. La distribution des distances entre marqueurs est représentée sur la figure 6.4.

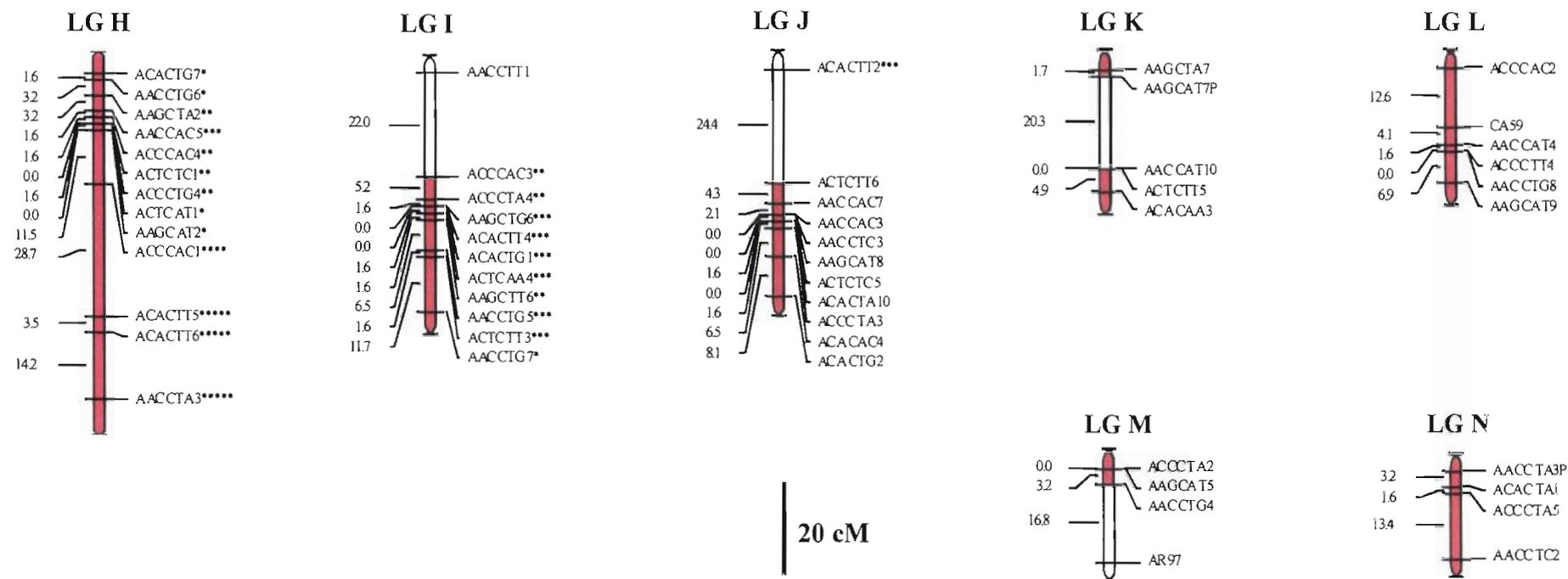


**Tableau 6.1.** : Principales caractéristiques (taille, nombre et espacement moyen des marqueurs) des groupes de liaison de la carte génétique obtenue à partir de la descendance BCDEW.

Groupe de liaison	Taille du groupe (cM)	Nombre de marqueurs	Espacement	
			Moyen (cM)	Min-max (cM)
A	224,4	31	7,5	0,0 - 25,3
B	150,6	21	7,5	0,0 - 22,3
C	112,7	16	7,5	0,0 - 24,4
D	101,6	15	7,2	0,0 - 17,8
E	100,4	12	9,1	1,9 - 15,0
F	97,3	13	8,1	0,0 - 17,6
G	96,1	19	5,3	0,0 - 20,1
H	70,7	13	5,9	0,0 - 28,7
I	51,8	11	5,2	0,0 - 22,0
J	48,6	11	4,9	0,0 - 24,4
K	26,9	5	6,7	0,0 - 20,3
L	25,2	6	5,0	0,0 - 12,6
M	19,7	4	6,7	0,0 - 16,8
N	18,2	4	6,1	1,6 - 13,4



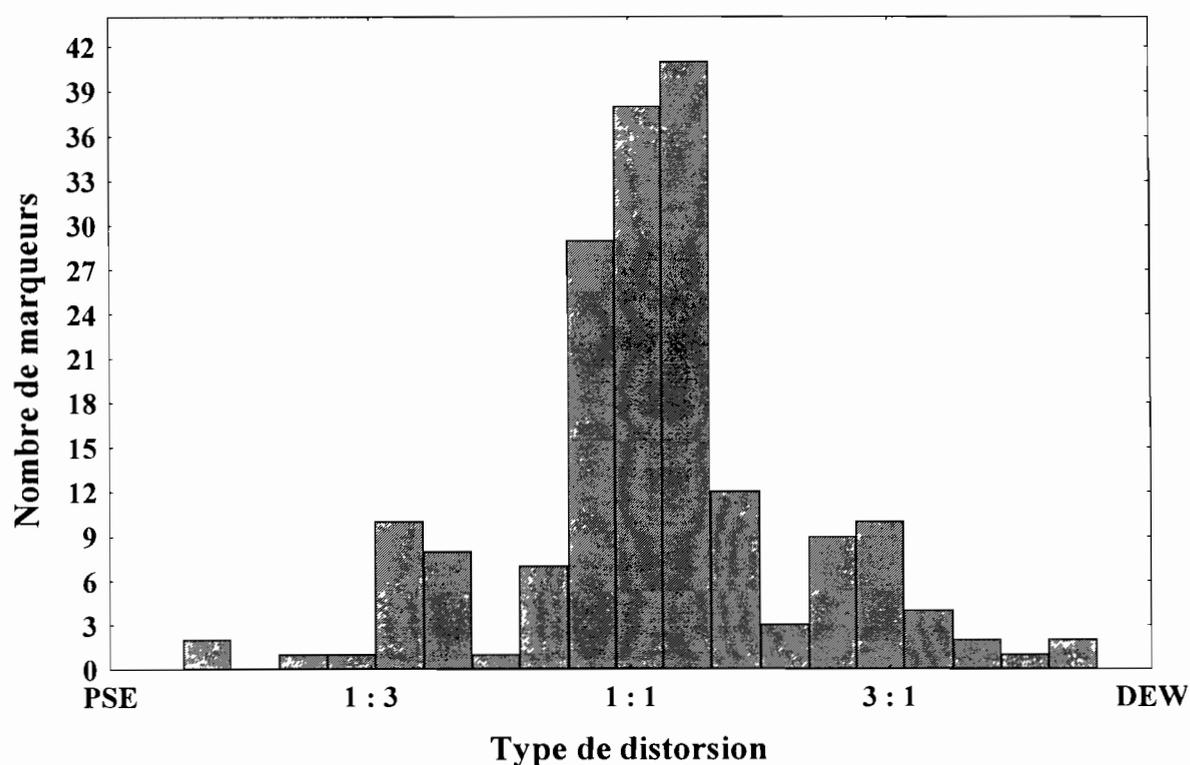
**Figure 6.4.** : Répartition des distances entre locus sur la carte génétique construite à partir de la descendance BCDEW.



**Figure 6.3.** : Carte de liaison génétique de la descendance BCDEW. Les marqueurs cartographiés figurent à droite des 14 groupes de liaison (167 AFLP, 13 RFLP et 1 morphologique). Les marqueurs AFLP sont représentés suivant l'exemple : "AACCTG8", les trois premières bases (AAC) correspondent à l'amorce *EcoRI*, les trois suivantes (CTG) à l'amorce *MseI* et le chiffre au numéro de bande. Les sondes RFLP proviennent d'ADNc d'un clone Arabusta (préfixe : "G" ou "AR") et d'ADN génomique de *C. Arabica* (préfixe : "CA"). Les distances entre marqueurs (représentées à gauche des groupes de liaison et exprimées en cM) ont été calculées en utilisant la fonction de Kosambi. Les fragments rouges et verts correspondent respectivement à des segments détectés pour des valeurs de *LOD score* de 5,5 et 4,8. L'agglomération de ces différents fragments a été effectuée pour une valeur du *LOD score* de 3,1 (segments blancs). Les loci marqués par \*, \*\*, \*\*\*, \*\*\*\* et \*\*\*\*\* correspondent aux distorsions de ségrégation significatives à  $p < 0,01$ ,  $p < 0,001$ ,  $p < 0,0001$ ,  $p < 0,00001$  et  $p < 0,000001$  respectivement.

**Tableau 6.2.** : Types de ségrégation obtenues à partir des 181 marqueurs cartographiés (167 marqueurs AFLP, 13 marqueurs RFLP et 1 marqueur morphologique) dans la descendance de BCDEW.

Marqueurs cartographiés	Ségrégation			Autres distorsions
	1 : 1	1 : 3	3 : 1	
181	127	26	21	7
100 %	70 %	14 %	12 %	4 %



**Figure 6.5.** : Distribution des 181 marqueurs cartographiés sur le descendance BCDEW en fonction des types de distorsion. La figure illustre clairement une majorité de marqueurs présentant une ségrégation mendélienne 1 : 1, et deux catégories de distorsion de ségrégation suivant un ratio 3 : 1 en faveur de PSE et un ration 1 : 3 en faveur de DEW.

## Distorsion de ségrégation

L'analyse de la ségrégation de chaque marqueur AFLP révèle des déviations par rapport à l'hypothèse Mendélienne (rapport 1 : 1 pour un marqueur dominant). Des tests de  $\chi^2$  montrent que parmi les 181 marqueurs cartographiés (167 AFLP, 13 RFLP et un marqueur morphologique), 54 sont des marqueurs distordus (53 AFLP et un RFLP), soit environ 30 % de distorsion (Annexe 6.2.). Ces 54 marqueurs distordus se répartissent en trois groupes :

- i) 26 (14 %) ségrègent suivant un rapport 1 : 3 en faveur de l'espèce PSE ;
- ii) 21 (12 %) ségrègent selon un rapport 3 : 1 en faveur de l'espèce DEW ;
- iii) 7 présentent d'autres types de ségrégation (Tab. 6.2.).

Il y a autant de distorsions en faveur de PSE que de DEW. La figure 6.5. illustre la distribution tri-modale de l'ensemble.

La répartition des 54 marqueurs distordus sur la carte génétique est illustrée sur la figure 6.6. Ils sont regroupés au niveau des extrémités présumées des groupes de liaison de grande et de moyenne taille et ne sont pas présents chez les quatre groupes de petite taille. Les fragments distordus en faveur de l'espèce DEW (en jaune sur la figure 6.6.) affectent deux groupes de liaison : la totalité du groupe H et la presque-totalité du groupe I. Les fragments distordus en faveur de l'espèce PSE (en orange sur la figure 6.6.) sont localisés au niveau de six groupes de liaison : sur la totalité du groupe C, sur environ la moitié des groupes B, E et F et pour un et deux marqueurs distordus respectivement sur les groupes J et G.

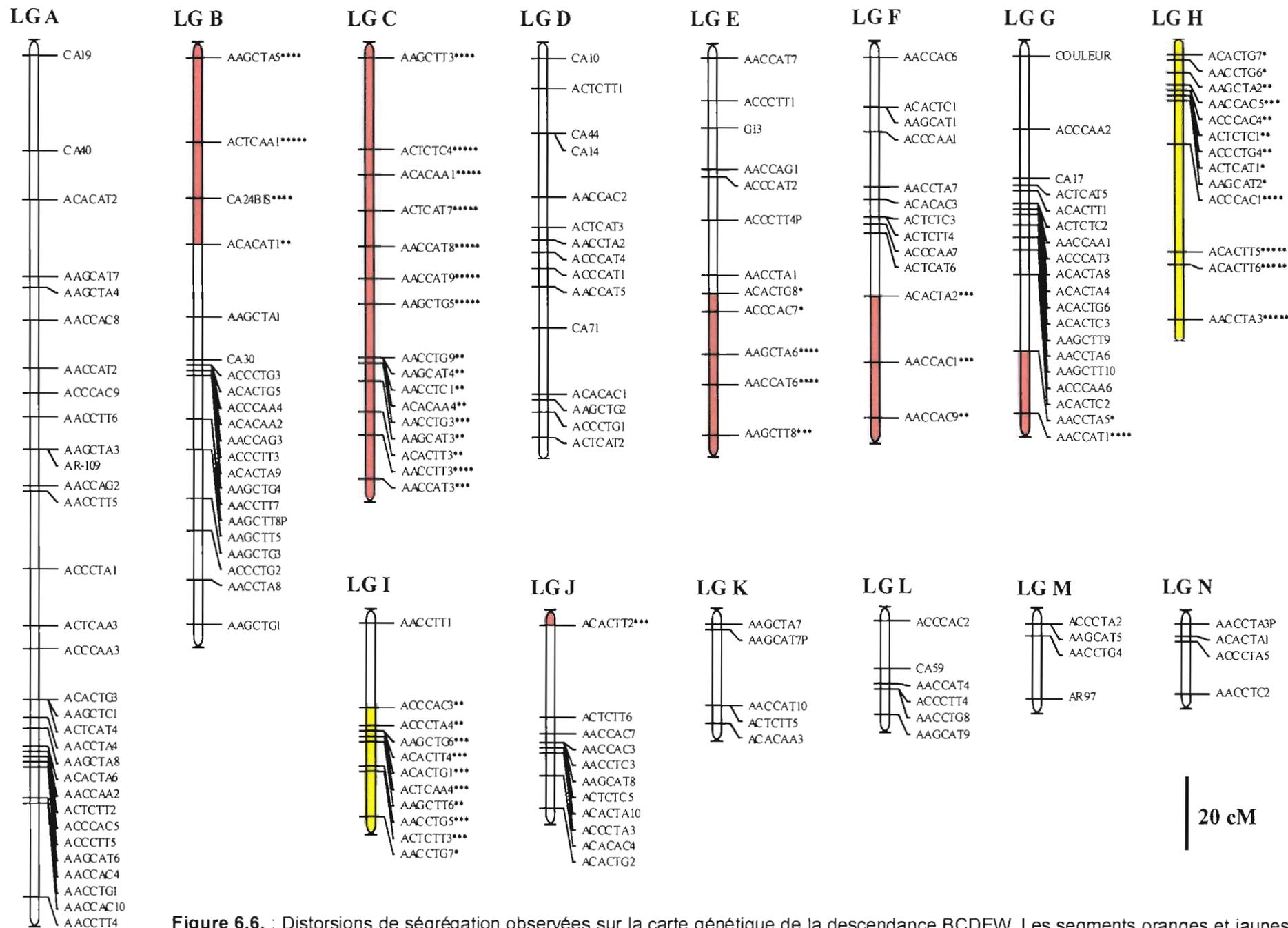
## **6. 2. Discussion**

### **6. 2. 1 Carte de liaison génétique**

#### Carte de liaison génétique

La simulation de cartographie a permis de définir, sur la carte de liaison génétique, des segments de marqueurs où la probabilité de détection de fausses liaisons est faible (valeur de LOD score de 5,5 en rouge sur la figure 6.3.). Sur un même groupe de liaison, ces zones rouges sont reliées par des zones « ponts » en blanc et/ ou vert.

L'obtention des 14 groupes de liaison de la carte génétique par agglomération des groupes provisoires se produit pour des valeurs variables du *LOD score*, mais toutes supérieures à 3,1. L'affectation d'un QTL donné à un groupe de liaison sera d'autant plus fiable que ce QTL sera localisé sur une zone rouge.



**Figure 6.6.** : Distorsions de ségrégation observées sur la carte génétique de la descendance BCDEW. Les segments oranges et jaunes correspondent respectivement aux distorsions en faveur de PSE et DEW.

Le nombre de groupes (14) supérieur au nombre de chromosomes du génome haploïde des caféiers ( $x = 11$ ) et l'existence de marqueurs non liés indiquent clairement que la carte n'offre qu'une représentation partielle du génome. L'augmentation du nombre de marqueurs devrait aboutir à une saturation de la carte actuelle par obtention des 11 groupes de liaison attendus. Cependant, trois groupes de liaison entièrement distordus existent (C en faveur de PSE, H et I en faveur de DEW) et il se pourrait que leur indépendance soit liée à leur distorsion (une distance de 10 cM entre deux marqueurs non-distordus devient égale à 24 cM lorsqu'elle regroupe un marqueur distordu à un marqueur non distordu). Dans ce cas, on parviendrait effectivement aux 11 groupes attendus.

Chez les caféiers diploïdes, espèces pérennes allogames, on ne dispose pas de lignées fixées, mais de géniteurs fortement hétérozygotes pour constituer les populations de cartographie. Cette hétérozygotie entraîne une ségrégation génétique des allèles dès la première génération. La construction de la carte génétique est donc plus complexe en raison de la ségrégation de deux à quatre allèles par locus et de la difficulté à observer la phase des allèles des parents (couplage ou répulsion). Pour remédier à ce problème, la cartographie du croisement PSE x DEW a été effectuée sur une population backcross avec des marqueurs spécifiques de PSE. Ceci s'apparente à un croisement entre un hybride F1 et un parent DEW génétiquement fixé, malgré le fait que nous soyons dans une situation de pollinisation libre entre plusieurs hybrides F1 et plusieurs géniteurs DEW. La distinction entre des locus de types parentaux et recombinés est possible, puisque la présence de la bande AFLP (marqueurs dominant) ne provient que du parent hybride F1. De même, le problème de la phase de couplage ou de répulsion ne se pose pas.

#### Taille de la carte génétique

La taille de la carte du croisement interspécifique PSE x DEW est d'environ 18 % inférieure à celle de *C. canephora* développée par Paillard *et al.* (1996) (1144 cM vs. 1402 cM). Ce résultat est en fait attendu : dans le croisement interspécifique (PSE x DEW), la faible homologie des génomes contribue à réduire le nombre de recombinaisons et par la même la taille de la carte. L'exemple extrême se rencontre chez la pomme de terre où la carte génétique d'un croisement interspécifique est réduite de 65 % par rapport à celle d'un croisement intraspécifique (Bonierbale *et al.*, 1988 ; Gebhardt *et al.*, 1991).

### Régions présentant des regroupements (« clusters »)

Les zones de haute densité en marqueurs AFLP, ou « clusters », sont caractéristiques des régions centromériques des chromosomes (Qi *et al.*, 1998). Les marqueurs AFLP sont parfois même utilisés afin de localiser avec précision la position d'un centromère sur un chromosome ; comme par exemple le chromosome 3 chez *Arabidopsis thaliana* (Alonso-Blanco *et al.*, 1998).

Il y a donc autant de clusters AFLP que de centromères et donc de chromosomes chez une espèce donnée. Sur notre carte de liaison génétique, neuf des quatorze groupes de liaison présentent des clusters et correspondraient donc à des chromosomes. Parmi les 13 marqueurs RFLP cartographiés, le marqueur CA-30 se trouve sur la région centromérique du groupe de liaison n°4 de la carte développée par Paillard *et al.* (1996) et dans celle du groupe de liaison B de notre carte.

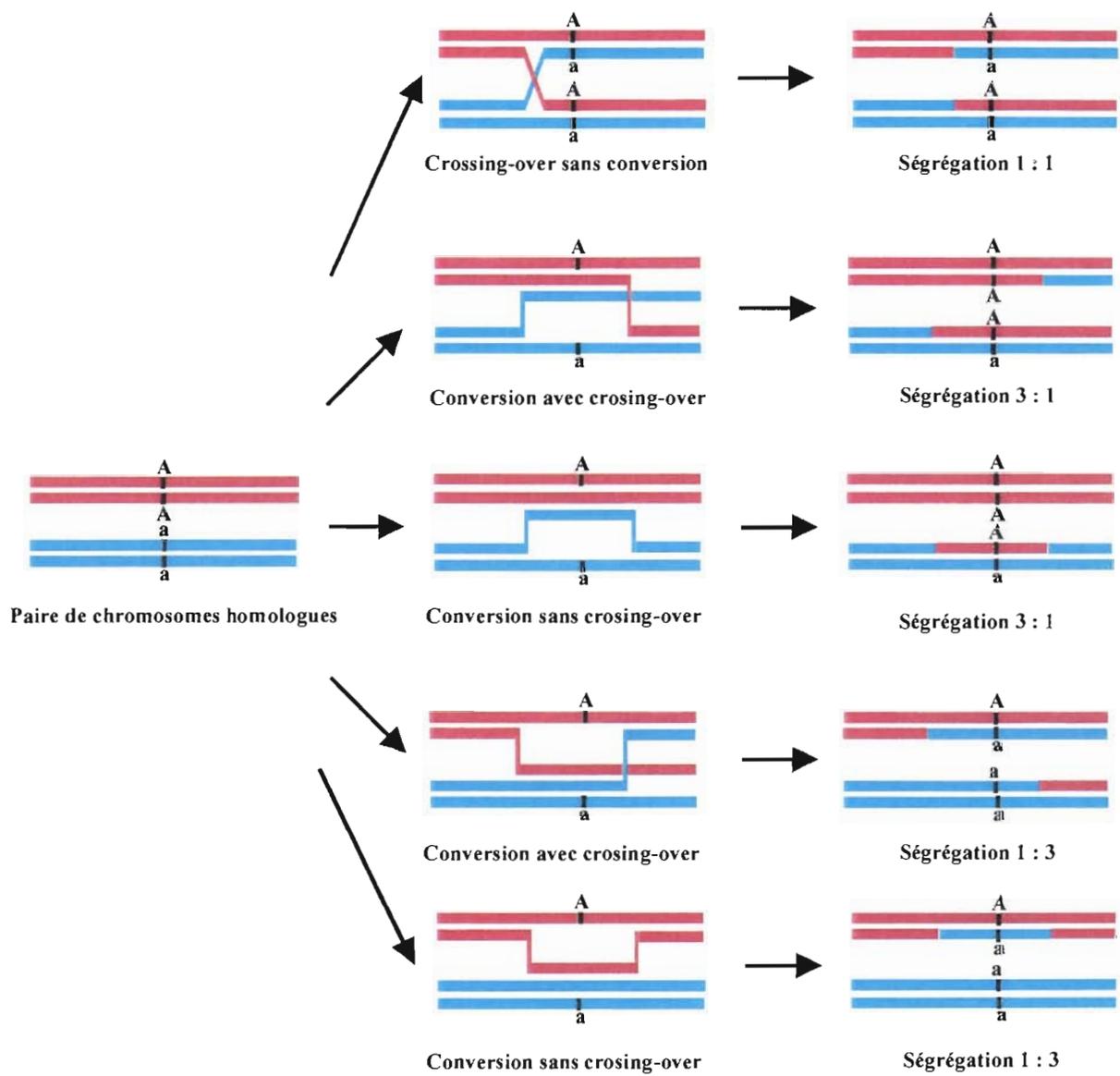
#### **6. 2. 2. Distorsions de ségrégation**

Sur l'ensemble des 14 groupes de liaison de la carte génétique, huit présentent des distorsions (Fig. 6.6.). Cette part de marqueurs distordus est variable en fonction des espèces considérées : 69 % chez *Cryptomeria japonica* (Nikaido *et al.*, 1999), 37 % chez *Citrus* (Luro *et al.*, 1994), 36 % chez *Oryza* (Virck *et al.*, 1998), 33 % chez *Prunus* (Foolad *et al.*, 1995), 23 % chez *Helianthus* (Quillet *et al.*, 1995) et 8,4 % chez *Lens* (Eujayl *et al.*, 1998). En comparaison, le taux de distorsion relevé au sein de notre croisement est intermédiaire (30 %).

L'existence de distorsions de ségrégation peut avoir une origine artéfactuelle ou biologique.

Les distorsions de ségrégation de type 3 : 1 en faveur de PSE pourraient résulter d'un artéfact, c'est à dire liées au codage de bandes superposées (double bande) (Cloutier *et al.*, 1997). Cette superposition de bandes ne peut expliquer alors l'existence d'une proportion équivalente de marqueurs distordus en faveur de DEW. L'observation de distorsions 3 : 1 et 1 : 3 tout à fait symétriques (Fig. 6.5.) aurait donc une origine biologique.

La sélection pourrait être gamétique (gamètes non viables), zygotique (échec de certaines fécondations) et post-zygotique (avortement sélectif des zygotes, faiblesse hybride) (voir Chp. 2. 3.). Néanmoins aucune de ces hypothèses ne permet d'expliquer avec vraisemblance les taux de distorsions suivant des ratios 1 : 3 ou 3 : 1, en faveur de l'un ou l'autre parent. Une telle symétrie ne peut refléter qu'un mécanisme génétique lié à la



**Figure 6.7.** : Représentation des échanges entre les chromatides lors de la méiose, rendant compte des ségrégations observées : mendélienne 1 : 1 et aberrantes : distorsions 3 : 1 et 1 : 3.

duplication des chromatides lors de la méiose. En fait, les distorsions de type 3 : 1 et 1 : 3 ne sont connues que dans les phénomènes de conversion génétique (Leblon et Rossignol, 1973; Nicolas et Rossignol, 1983; Nag *et al.*, 1989).

La conversion génétique est un phénomène connu chez les champignons ascomycètes (*Sordaria* sp., *Neurospora* sp.). Chez de tels organismes, lorsqu'on examine un nombre important d'asques, on en trouve toujours quelques uns qui font exception à la disjonction régulière 2 : 2 en présentant des disjonctions 1 : 3 et 3 : 1. Ces deux types d'asques se rencontrent avec des fréquences qui restent de même ordre de grandeur, mais qui peuvent être plus ou moins inégales, sans qu'aucune règle générale précise ne puisse être formulée à ce sujet (L'Héritier, 1975).

La conversion génétique n'est pas une mutation, mais un phénomène intimement lié au mécanisme de recombinaison et de réparation de l'ADN (Fig. 6.7.). Une conversion correspond à une zone où la duplication de l'un des chromosomes conduit non pas à 2, mais à 3 chromatides semblables, tandis que l'homologue n'est pas recopié. Tout crossing-over comporte, dans la région où les chromatides changent de partenaire, une zone de conversion plus ou moins étendue. Le mécanisme moléculaire à l'origine de ces événements implique la formation d'erreurs d'appariements sous forme d'hétéroduplex d'ADN, suivie de leur réparation (Nag *et al.*, 1989). La formation de ces hétéroduplex pourrait caractériser les méioses d'hybrides interspécifiques, lorsque la taille de génome des espèces parentales diffère, comme c'est le cas ici, où la taille du génome de DEW est 25 % plus élevée que celle de PSE (Barre *et al.*, 1996).

La symétrie des conversions génétiques n'est pas toujours retrouvée. Ainsi, dans le croisement inter générique entre *Citrus reticulata* et *Poncirus trifoliata* (mandarinier x pamplemoussier), il existe 37 % de locus distordus (marqueurs RFLP et RAPD confondus) (Luro *et al.*, 1994). Bien que les auteurs ne l'aient pas décrit, leurs données montrent clairement deux types de marqueurs distordus : 9 sont distordus en faveur de *Poncirus* avec une valeur de  $p = 0,30$  (distorsion 1 : 3) et 29 sont distordus en faveur du *Citrus* avec  $p = 0,75$  (distorsion 3 : 1).

### 6. 3. En résumé

Malgré la forte hétérozygotie, liée au mode de reproduction strictement allogame des caféiers diploïdes, une carte génétique du croisement interspécifique entre PSE et DEW a été réalisée. Cette cartographie repose sur une descendance backcross 1, qui s'apparente à un

croisement entre un géniteur hétérozygote (hybride F1) et un géniteur homozygote (DEW) respectivement pour la présence/ absence de bandes AFLP.

Les principales caractéristiques de la carte génétique obtenue sont :

- i) la présence de 14 groupes de liaison pour 181 marqueurs, couvrant une distance totale de 1144 cM ;
- ii) la présence de 30 % de marqueurs présentant une distorsion de ségrégation, au sein de laquelle des déviations de type 3 : 1 et 1 : 3 symétriques sont observées (26 %). L'origine probable de telles distorsions est la conversion génétique.

Les travaux complémentaires de cartographie sont nombreux :

- i) saturer la carte, car le nombre de groupes de liaison ne correspond pas encore au nombre chromosomique (14 vs 11) et des zones de plus de 20 cM sans marqueurs sont encore présentes (7 % des intervalles entre marqueurs). La densité de marqueurs AFLP pourra être accrue sachant que seulement 30 combinaisons d'amorces ont été utilisées sur une potentialité de  $64^2$  combinaisons possibles ;
- ii) faire une étude de cartographie comparée sur d'autres croisements de caféiers par cartographie réciproque de marqueurs codominants. Ces derniers peuvent être des sondes RFLP, des marqueurs microsatellites ou des sondes issues du séquençage de marqueurs AFLP. Des descendances backcross issues de croisements interspécifiques de *C. canephora* avec : *C. heterocalyx*, *C. pseudozanguebariae*, *C. liberica* et *C. eugenioides* sont en cours d'étude. La cartographie de ces descendances hybrides permettra d'étudier l'organisation génomique d'espèces différant par la taille de leur génome à partir des éléments de synténie et de colinéarité. Les informations recueillies devraient fournir des éléments de réflexion quant au schéma évolutif au sein du genre *Coffea* ;
- iii) utiliser la carte pour connaître les bases génétiques de caractères complexes comme les composantes de la qualité du café à la tasse. La première application de cette carte porte sur l'identification de tels QTLs et les premiers résultats sont développés dans le chapitre suivant.

## **Chapitre 7 : Hérité des caractères d'intérêt :**

### **Identification de QTLs**

---

## Chapitre 7 : Hérité des caractères d'intérêt : Identification de QTLs

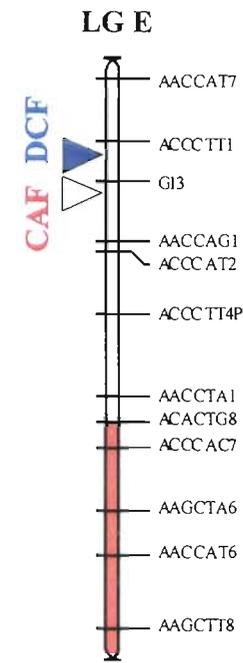
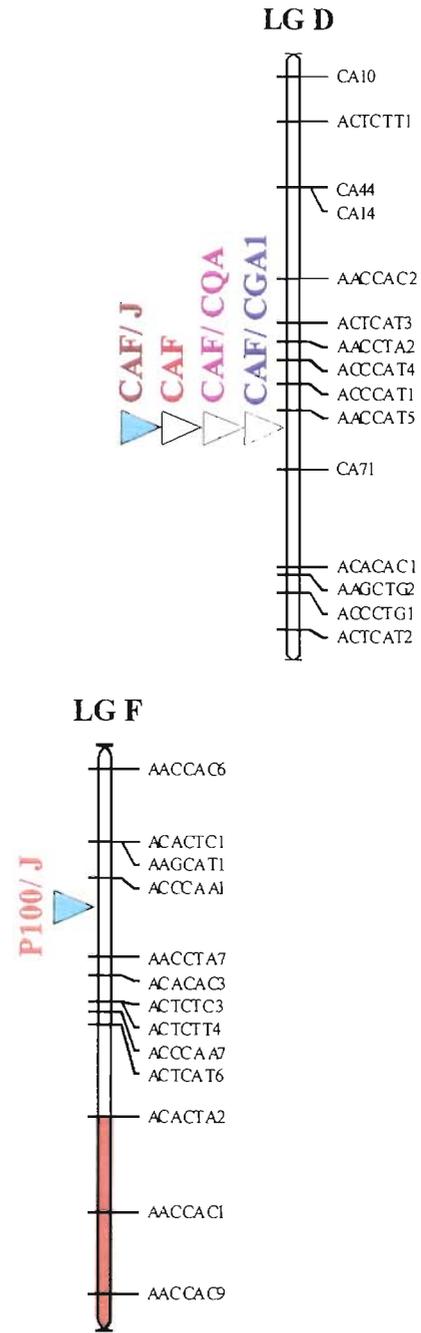
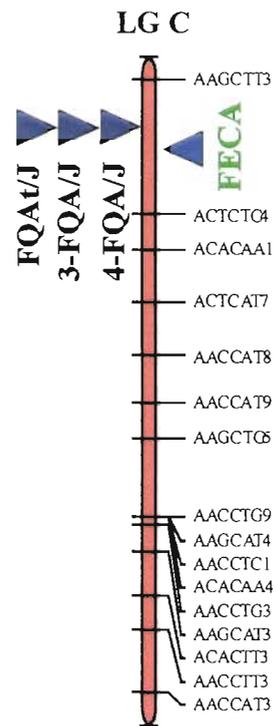
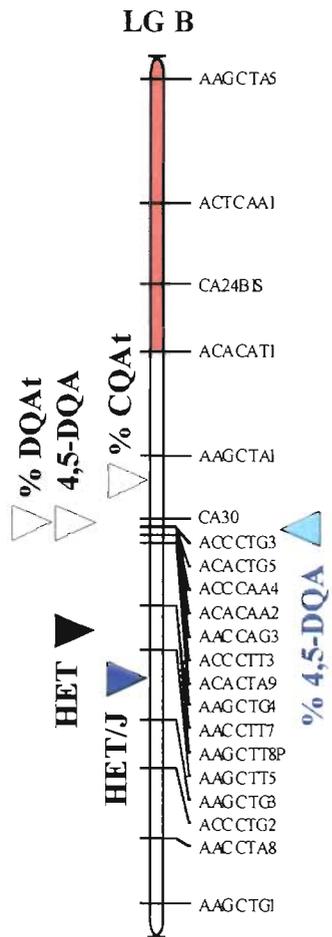
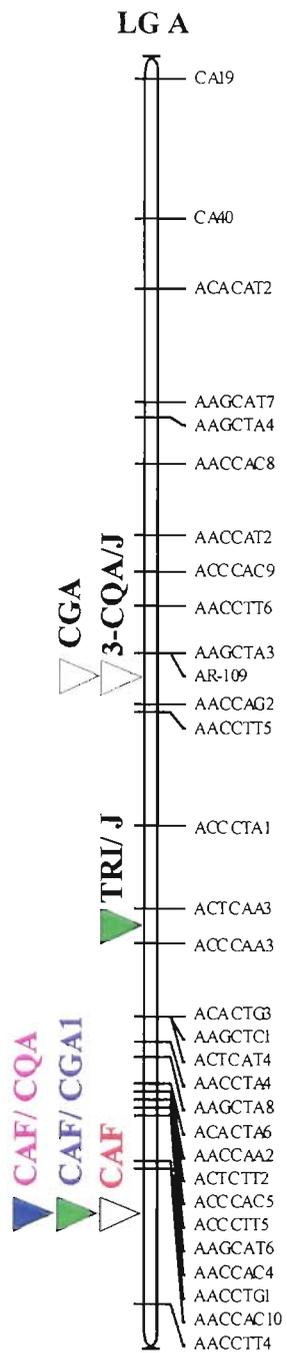
La descendance hybride BCDEW a permis d'établir une carte génétique (Chap. 6) et a fait l'objet de nombreuses évaluations concernant des caractères : biochimiques, phénologiques, agronomiques et de fertilité (voir Chap. 3). Les données de phénologie et de fertilité obtenues à la station de Man par S. Akaffou ont fait l'objet d'une thèse sur leur déterminisme génétique avec une approche quantitative classique (Akaffou, 1999). Les évaluations sur les teneurs en caféine et en hétéroside ont été réalisées dans le cadre de la thèse de P. Barre (Barre *et al.*, 1998).

Ce chapitre a pour but de présenter les différents QTLs correspondant à ces caractères, par la double approche : analyse de variance et utilisation du logiciel MAPMAKER/ QTL. La première partie correspond à une présentation générale des QTLs mis en évidence pour l'ensemble des 55 variables quantitatives présentées en annexe 3.3. La seconde partie correspond à l'étude des interactions entre QTLs pour les caractères dont les variations impliquent au moins deux QTLs.

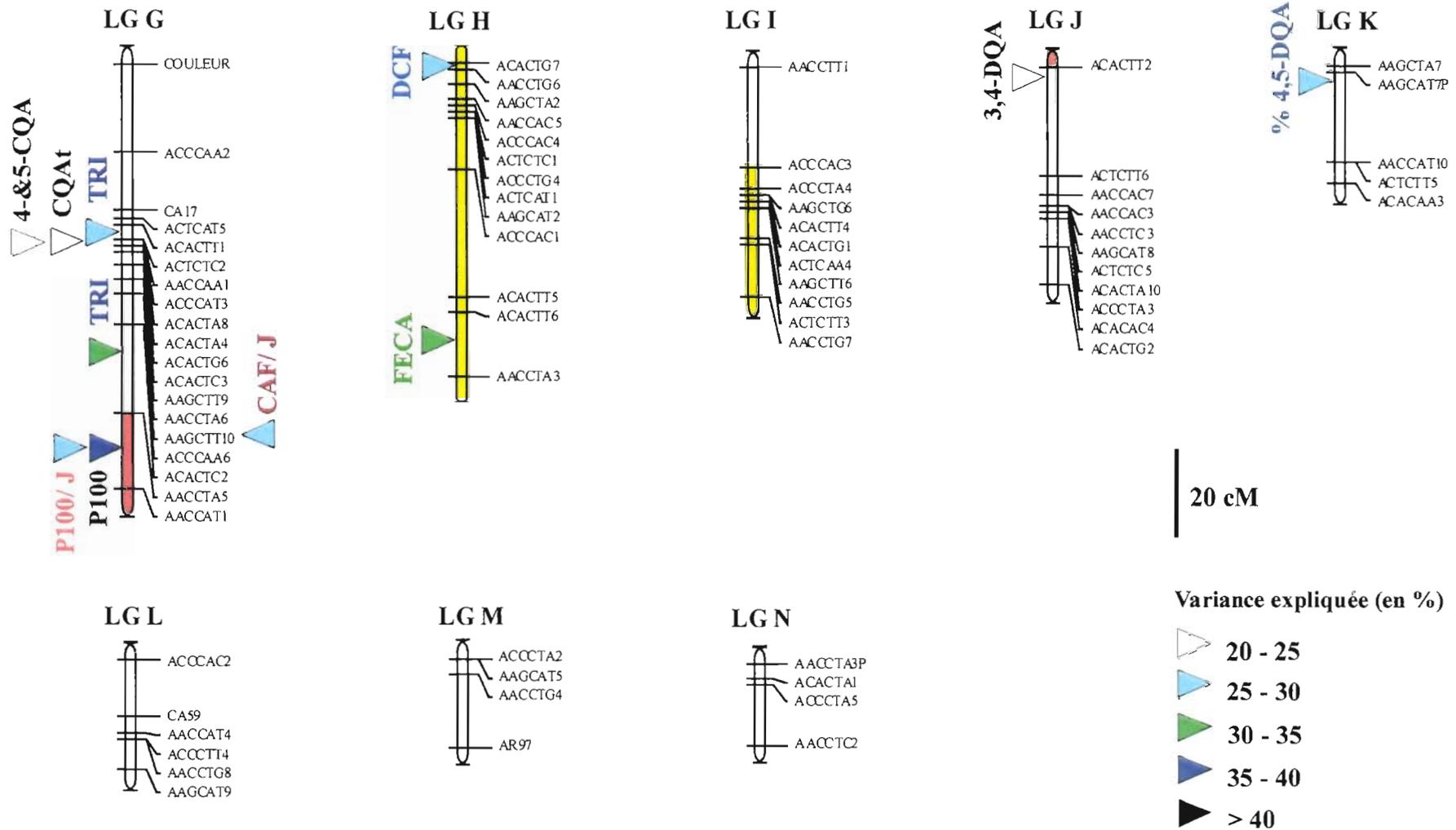
### 7. 1. Résultats

#### 7. 1. 1. Présentation de l'ensemble des QTLs détectés

L'objectif de cette partie est de présenter l'ensemble des QTLs mis en évidence au sein de la descendance BCDEW pour les 55 caractères quantitatifs (annexe 3.1.). L'annexe 7.1. résume, par groupe de liaison, les caractéristiques des QTLs mis en évidence pour un seuil de *LOD score* fixé à 2,0 (valeur peu restrictive permettant d'avoir un aperçu global des QTLs). La figure 7.1. permet de visualiser la position des QTLs sur le génome pour un seuil de *LOD score* plus restrictif fixé à 2,6.



20 cM



**Figure 7.1.** : Cartographie des QTLs révélés pour des valeurs de *LOD score* supérieures ou égales à 2,6 dans la descendance BCDEW. Chaque QTL est localisé le long des groupes de liaison par un triangle, dont l'intensité augmente avec l'importance de la variance expliquée (voir légende). Les caractères figurés en couleur et en noir mettent respectivement en évidence 2 QTLs et 1 QTL. Les codes des caractères figurent dans l'annexe 3.1. Les segments oranges et jaunes des groupes de liaison correspondent respectivement aux distorsions en faveur de PSE et DEW.

## Nombre de QTLs

La recherche de QTLs réalisée par ANOVA révèle que 35 associations variables quantitatives/ marqueurs moléculaires (sur les 9955 associations obtenues sous le logiciel Statistica) sont mises en évidence au seuil de  $p < 0,001$ , impliquant 27 caractères et 22 marqueurs. Pour un seuil plus sévère de  $p < 0,0001$ , vingt cinq associations sont retenues concernant 21 caractères et 18 marqueurs. Au seuil plus restrictif de  $p < 0,00001$ , il reste encore sept associations qui regroupent six caractères et six marqueurs. L'ANOVA présente l'avantage de détecter des QTLs avec des marqueurs non cartographiés. C'est le cas pour les deux caractères suivants : le poids de cent grains (P100) lié au marqueur ACCCAA5 ( $R^2 = 19,7\%$  ;  $p < 0,00005$ ) et la quantité de saccharose accumulée par jour (SUC/J) liée au marqueur ACCCTA1P ( $R^2 = 20,2\%$  ;  $p < 0,00008$ ).

Le nombre de QTLs mis en évidence par MAPMAKER/ QTL varie suivant les valeurs de *LOD score* choisies. Pour un *LOD score* de 2, cinquante quatre QTLs sont détectés. Ces derniers sont associés à des 55 caractères quantitatifs évalués. Pour un *LOD score* de 2,6, trente trois QTLs expliquant 24 des caractères sont mis en évidence. Treize QTLs restent en lice pour un *LOD score* compris entre 3 et 4. Deux QTLs correspondant à P100 et à l'accumulation journalière de l'hétéroside (HET/J) (variable obtenue en divisant la teneur en hétéroside par la durée du cycle de fructification) sont toujours présents pour des *LOD score* respectifs de 4,30 et 4,22. Un seul QTL est détecté à un *LOD score* de 6,5 et est associé à la teneur en hétéroside (HET).

Tous les QTLs mis en évidence par ANOVA correspondent à ceux mis en évidence par MAPMAKER/ QTL. En revanche, certains QTLs à effet mineur ont été mis en évidence avec MAPMAKER/ QTL, alors qu'ils n'étaient pas retenus au seuil  $p < 0,001$  avec l'ANOVA. Les six associations retenues au seuil de  $p < 0,00001$  en ANOVA ont été mis en évidence par MAPMAKER/ QTL pour des *LOD score* supérieurs à 3. Ils concernent la teneur en trigonelline (TRI), la teneur en hétéroside (HET), la teneur en hétéroside par jour (HET/J), le poids de cent grains (P100), la durée du cycle de fructification (DCF) et le taux de fruits à une écaille (FECA).

Les 24 caractères présentant un ou plusieurs QTL(s) (à *LOD* 2,6) peuvent être classés en deux groupes : ceux présentant un seul QTL (15 caractères) et ceux présentant deux ou trois QTLs (9 caractères).

**Tableau 7.1.** : Moyennes des caractères mono QTL détectés à un *LOD score* de 2,6 et effectifs (chiffre entre parenthèse), des classes définies par la présence de l'allèle apporté par PSE (colonne Htz DP) et de son absence spécifique de DEW (colonne Hmz DD). Les lettres entre parenthèse au côté des locus correspondent aux groupes de liaison. Les deux dernières colonnes correspondent aux valeurs moyennes retrouvées chez les espèces parentales PSE et DEW.

Caractère	Locus	Htz DP	Hmz DD	PSE	DEW
CQA45 (% MS)	ACTCAT5 (G)	3,83 (24)	4,48 (28)	1,37	5,80
DQA34 (% MS)	ACACTT2 (J)	0,230 (37)	0,177 (16)	0,005	0,262
DQA45 (% MS)	ACCCTT3 (B)	0,515 (26)	0,414 (26)	0,025	0,596
CQAt (% MS)	ACTCAT5 (G)	4,14 (24)	4,79 (28)	1,40	6,33
CGA2 (% MS)	AR109 (A)	5,33 (10)	6,60 (8)	1,51	7,97
% CQAt (%)	AAGCTA1 (B)	0,823 (30)	0,850 (19)	0,925	0,794
% DQAt (%)	ACCCTT3 (B)	0,167 (26)	0,138 (26)	0,025	0,149
FQA4/ J (% MS/ J)	AAGCTT3 (C)	0,158 (32)	0,238 (7)	0,112	0,150
FQA3/ J (% MS/ J)	AAGCTT3 (C)	0,033 (32)	0,044 (7)	0,016	0,015
FQA4/ J (% MS/ J)	AAGCTT3 (C)	0,007 (32)	0,013 (7)	0,000	0,011
CQA3/ J (% MS/ J)	AR109 (A)	0,108 (9)	0,132 (5)	0,039	0,172
HET (% MS)	AAGCTT5 (B)	0,150 (28)	0,010 (27)	1,000	0,000
HET/ J (% MS/ J)	AAGCTT5 (B)	0,060 (19)	0,003 (22)	0,015	0,000
TRI/ J (% MS/ J)	ACTCAA3 (A)	0,425 (19)	0,323 (19)	1,522	0,184
P100	AACCTA5 (G)	8,89 (38)	11,07 (18)	2,24	15,13

### Effet statistique des QTLs sur la variance phénotypique

Le pourcentage de variance expliqué par ces différents QTLs est variable (Annexe 7.1.). La figure 7.1. illustre cette variation en décomposant les 33 QTLs détectés en cinq classes : douze QTLs expliquent de 20 à 25 % de la variance du caractère (triangles blancs, figure 7.1.), huit QTLs expliquent de 25 à 30 % de la variance (triangles bleus ciel), cinq QTLs de 30 à 35 % (triangles verts), cinq QTLs de 35 à 40 % (triangles bleu foncé) et quatre QTLs expliquent plus de 40 % de la variance (triangles noirs). Finalement, les modèles multi-QTLs retenus expliquent de 22,4 % à 56,0 % de la variance (22,4 % pour la teneur en CQA totale et 56,0 % pour le taux de fruits à une écaille).

Pour chaque caractère à un seul QTL, les valeurs moyennes et les effectifs des classes définies par la présence ou l'absence de l'allèle PSE au locus marqueur nous renseignent respectivement sur les effets alléliques et les distorsions de ségrégation (Tab. 7.1.). Les effets des allèles sont estimés en comparant les moyennes phénotypiques des deux classes « 0 » et « 1 » chez les hybrides BCDEW avec celles des espèces DEW et PSE. Deux situations peuvent se présenter :

- i) La première correspond au cas où les valeurs des hybrides BCDEW sont cohérentes avec celles des espèces parentales. C'est le cas, par exemple, de la teneur en 4,5-CQA, où l'apport de l'allèle PSE chez l'hybride contribue à diminuer cette teneur et où son absence rend compte d'une teneur plus élevée ;
- ii) La seconde situation correspond au cas inverse et concerne les quatre variables suivantes : 3,4-DQA, 4,5-DQA, FQA/J et 4-FQA/J. Pour le 3,4-DQA par exemple, la présence de l'allèle PSE chez les hybrides BCDEW contribue à y élever la teneur.

### Localisation des QTLs

Les 33 QTLs détectés associés aux 24 caractères sont localisés dans 16 zones du génome, réparties sur dix des 14 groupes de liaison. Pour neuf caractères expliqués par deux ou trois QTLs, ceux-ci sont situés sur des groupes de liaison différents, à l'exception de la teneur en trigonelline, dont les deux QTLs sont localisés sur le groupe de liaison G.

Ce dernier groupe est aussi celui qui présente le plus de QTLs avec sept QTLs impliqués dans la variation de six caractères. En revanche, aucun QTL n'a été détecté au niveau des groupes de liaison de taille réduite I, L, M et N.

Huit zones du génome sont impliquées dans l'expression de plus de deux caractères, mais peuvent être considérées comme associées à un seul et même caractère. Par exemple, sur le groupe de liaison C, un point présente trois QTLs expliquant les variations de FQA/J, 4-FQA/J et 3-FQA/J. Ceci revient à dire que cette zone contrôle le flux journalier de production des FQA. De même, au niveau du groupe de liaison D, les quatre QTLs détectés sont liés à la production de la caféine.

Trois co-localisations de QTLs intervenant dans des caractères très différents, ont été mises en évidence. La première co-localisation implique le groupe de liaison C, où des QTLs contrôlant la production de FQA et le taux de fruits à une écaille co-existent. De même, au niveau du groupe G, un QTL de la teneur en CQA et un QTL de celle en trigonelline sont co-localisés. Il en est de même, sur ce groupe G, pour un QTL du poids de cent grains et un QTL de la teneur en caféine.

Enfin, neuf QTLs sont situés dans des régions présentant des distorsions de ségrégation (voir figure 7.1.).

### 7. 1. 2. Interprétation pour les caractères présentant plusieurs QTLs

Neuf caractères sont associés à plusieurs QTLs (TRI, DCF, %4,5-DQA, CAF/CQA, CAF/CGA1, CAF/J, P100/J, CAF et FECA). Dans ce paragraphe, nous allons exposer quelques cas d'interprétation, caractéristiques de ce qu'il est possible de rencontrer chez des variables quantitatives montrant plusieurs QTLs. Nous prendrons les exemples de la teneur en trigonelline (TRI), de la durée du cycle de fructification (DCF) et de la variable %4,5DQA.

Pour toutes ces analyses, nous avons écrit en caractère normal les allèles d'un locus AFLP, en majuscule pour DEW et en minuscule pour PSE. Exemples :  $T_1$  (DEW) et  $t_1$  (PSE). Nous avons écrit en caractère italique les allèles DEW et PSE du QTL correspondant :  $T_l$  (DEW) et  $t_l$  (PSE).

#### La teneur en trigonelline

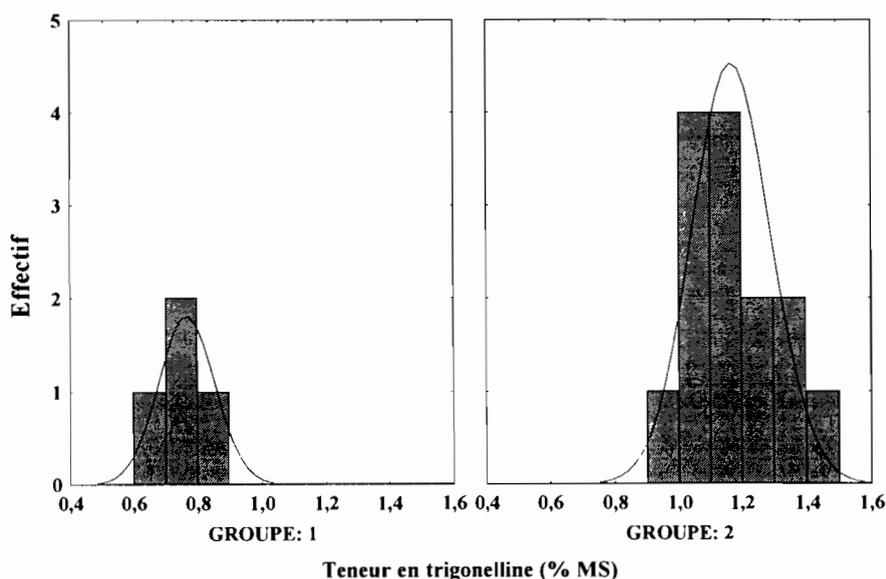
Nous avons formulé l'hypothèse (Chap. 5) d'une hérédité cytoplasmique pour la teneur moyenne en trigonelline des grains verts chez les hybrides BCPSE, F1 et BCDEW. Néanmoins, les variations observées entre les arbres à l'intérieur des espèces, des hybrides F1 et des hybrides BCPSE et BCDEW suggèrent l'existence de gènes nucléaires. Effectivement, la diversité observée pour ce caractère à l'intérieur de BCDEW est liée à deux marqueurs ACTCAT5 et ACACTC2 situés le long du même groupe de liaison G.

**Tableau 7.2.** : Moyennes des teneurs en trigonelline (% MS) et effectifs (chiffre entre parenthèse) des classes définies par les allèles  $T_1$ ,  $T_2$  de DEW et  $t_1$ ,  $t_2$  de PSE aux locus ACTCAT5 et ACACTC2. Les résultats des tests de Newman et Keuls sont indiqués par les lettres a et b en exposant.

		ACTCAT5	
		$T_1T_1$	$T_1t_1$
ACACTC2	$T_2T_2$	0,81 <sup>a</sup> (19)	1,21 <sup>b</sup> (2)
	$T_2t_2$	0,84 <sup>a</sup> (2)	1,07 <sup>b</sup> (18)

**Tableau 7.3.** : Moyennes des teneurs en trigonelline (% MS) et effectifs, des classes définies par les allèles  $T_1$ ,  $T_2$  de DEW et  $t_1$ ,  $t_2$  de PSE aux locus des QTL1 (proche de ACTCAT5) et QTL2 (proche de ACACTC2). Les résultats des tests de Newman et Keuls sont indiqués par les lettres a et b en exposant.

		QTL1		Moy.
		$T_1T_1$	$T_1t_1$	
QTL2	$T_2T_2$	0,81 <sup>a</sup> (19)	1,21 <sup>b</sup> (2)	0,85
	$T_2t_2$	0,79 <sup>a</sup> (6)	1,16 <sup>b</sup> (14)	1,05
Moy.		0,80	1,16	



**Figure 7.2.** : Distribution de la teneur en trigonelline (% MS) chez les hybrides BCDEW présentant la combinaison allélique [ $T_1t_1$   $T_2t_2$ ]. Deux groupes 1 et 2 distincts caractérisent les deux distributions normales.

Soient  $T_1$  et  $t_1$ , les allèles du locus ACTCAT5 et  $T_2$  et  $t_2$ , les allèles du locus ACACTC2. Dans ces conditions, le tableau 7.2 donne les quatre teneurs moyennes observées chez les génotypes  $[T_1T_1 T_2T_2]$ . Aucune interaction entre ces deux marqueurs n'a été détectée ( $F_{1,38} = 0,54$  ;  $p < 0,47$ ).

Une étude des distributions à l'intérieur de ces quatre classes montre que trois d'entre-elles sont homogènes. En revanche, la classe  $[T_1t_1 T_2t_2]$  présente une distribution de type bimodale (Fig. 7.2.) avec deux groupes, de 4 et 14 individus, caractérisés par des teneurs moyennes respectives de 0,77 et 1,16. En fait, la moyenne du groupe1 de 4 individus ne diffère pas de celle des génotypes  $[T_1T_1]$ . Ce groupe1 correspond à des plants pour lesquels il y a eu recombinaison entre le QTL et le locus ACTCAT5.

Le tableau 7.3 est obtenu en appelant  $T_1$ ,  $T_2$ , les allèles DEW et  $t_1$  et  $t_2$ , ceux de PSE aux des QTL1 et QTL2. En fait, le QTL2 apparaît être un artéfact (il n'existe aucune différence entre  $[T_2T_2]$  et  $[T_2t_2]$ ). La présence de l'allèle  $t_1$  apporte un gain de teneur de 0,31 % MS. Ceci nous permet d'estimer, sous une hypothèse d'additivité, la valeur phénotypique d'un génotype  $[t_1t_1]$  : 1,48 % MS. Nous remarquons que les valeurs phénotypiques de DEW et PSE sont nettement inférieures à celles des génotypes  $[T_1T_1]$  et  $[t_1t_1]$ , alors que celle de la F1 (1,13 % MS) est proche de celle des génotypes  $[T_1t_1]$ .

#### La durée du cycle de fructification

La durée moyenne du cycle de fructification (DCF) chez les espèces parentales PSE et DEW est respectivement de 67 et 309 jours. Chez les hybrides F1, cette durée est égale à 180 jours (valeur attendue de 188 jours) (Akaffou, 1999). Les deux QTLs expliquant les variations du caractère DCF au sein des hybrides BCDEW seraient liés aux locus ACCCTT1 et ACACTG7, localisés respectivement sur les groupes de liaison E et H.

Soient  $F_1$  et  $f_1$ , les allèles du locus ACCCTT1 et  $F_2$  et  $f_2$ , les allèles du locus ACACTG7. Le tableau 7.4. présente les durées moyennes de fructification pour les quatre génotypes  $[F_1F_1 F_2F_2]$ ,  $[F_1f_1 F_2F_2]$ ,  $[F_1F_1 F_2f_2]$  et  $[F_1f_1 F_2f_2]$ . Un point important à souligner est l'existence d'une interaction entre les deux locus ( $F_{1,37} = 5,19$  ;  $p < 0,03$ ). Cette interaction est due au fait que les génotypes  $[F_1F_1 F_2F_2]$  et  $[F_1F_1 F_2f_2]$  ont le même phénotype, alors que les génotypes  $[F_1f_1 F_2F_2]$  et  $[F_1f_1 F_2f_2]$  diffèrent pour DCF. L'étude des distributions à l'intérieur des quatre classes génotypiques montre, pour chacune d'entre-elles, l'existence de recombinants entre le locus AFLP et le QTL.

**Tableau 7.4.** : Moyennes de la durée du cycle de fructification (exprimée en jours) et effectifs, des classes définies par les allèles  $F_1$ ,  $F_2$  de DEW et  $f_1$ ,  $f_2$  de PSE aux locus ACCCTT1 et ACACTG7. Les résultats des tests de Newman et Keuls sont indiqués par les lettres a et b en exposant.

		ACACTG7	
		$F_2F_2$	$F_2f_2$
ACCCTT1	$F_1F_1$	282 <sup>a</sup> (14)	283 <sup>a</sup> (4)
	$F_1f_1$	265 <sup>a</sup> (11)	228 <sup>b</sup> (12)

**Tableau 7.5.** : Moyennes de la durée du cycle de fructification (exprimée en jours) et effectifs, des classes définies par les allèles  $F_1$ ,  $F_2$  de DEW et  $f_1$ ,  $f_2$  de PSE aux locus des QTL1 (proche de ACCCTT1) et QTL2 (proche de ACACTG7). Les résultats des tests de Newman et Keuls sont indiqués par les lettres a et b en exposant.

		QTL2	
		$F_2F_2$	$F_2f_2$
QTL1	$F_1F_1$	281 <sup>a</sup> (18)	297 <sup>a</sup> (4)
	$F_1f_1$	235 <sup>b</sup> (4)	225 <sup>b</sup> (12)

**Tableau 7.6.** : Moyennes du pourcentage en isomère 4,5-diCQA (% DQA45 = DQA45/ DQA4t) et effectifs, des classes définies par les allèles  $D_1$ ,  $D_2$  de DEW et  $d_1$ ,  $d_2$  de PSE aux locus AAGCAT7P et AAGCTG4. Les résultats des tests de Newman et Keuls sont indiqués par les lettres a et b en exposant.

		AAGCTG4	
		$D_2D_2$	$D_2d_2$
AAGCAT7P	$D_1D_1$	0,504 <sup>a</sup> (12)	0,531 <sup>c</sup> (13)
	$D_1d_1$	0,532 <sup>c</sup> (10)	0,551 <sup>b</sup> (13)

Appelons  $F_1$ ,  $F_2$ ,  $f_1$  et  $f_2$  les allèles DEW et PSE des QTL1 et QTL2 ; les classes génotypiques du tableau 7.5 montrent que :

- i) les effectifs des génotypes  $[F_2F_2]$  et  $[F_2f_2]$  ne sont pas significativement différents (la distorsion ne touche pas le QTL2) ;
- ii) les valeurs phénotypiques de ces deux classes génotypiques ne diffèrent pas ;
- iii) QTL1 et QTL2 sont liés.

En fait, il est fort probable que :

- i) QTL1 et QTL2 ne constituent qu'un seul QTL ;
- ii) le groupe de liaison H, distordu, est en réalité lié au groupe E ;
- iii) l'interaction résulte de la présence de distorsions. Dans ce cas, nous n'aurions que le QTL1 lié au locus ACCCTT1. Les deux classes génotypiques sont  $[F_1F_1]$  et  $[F_1f_1]$  de valeurs phénotypiques respectives 284,2 et 227,4 jours.

#### Le taux relatif en isomère 4,5-DQA

Les variations du %4,5-DQA seraient liées à deux locus : AAGCAT7P et AAGCTG4, situés respectivement sur les groupes de liaison K et B. Aucune interaction n'a été détectée entre ces deux locus ( $F_{1,44} = 0,306$  ;  $p < 0,59$ ).

Soient  $D_1$  et  $d_1$ , les allèles du locus AAGCAT7P et  $D_2$  et  $d_2$ , les allèles du locus AAGCTG4. Le tableau 7.6. présente les moyennes du %4,5-DQA pour les quatre classes génotypiques  $[D_1D_1 D_2D_2]$ ,  $[D_1d_1 D_2D_2]$ ,  $[D_1D_1 D_2d_2]$  et  $[D_1d_1 D_2d_2]$ . Contrairement au cas de la teneur en trigonelline et de la durée du cycle de fructification, les distributions de cette variable dans les classes génotypiques sont très étalées et n'ont pas permis d'identifier des recombinaisons entre le locus AFLP et le QTL correspondant. De ce tableau, on peut constater que les effets liés à la présence des allèles  $d_1$  et  $d_2$  sont identiques et que les valeurs des génotypes  $[D_1d_1 D_2D_2]$  et  $[D_1D_1 D_2d_2]$  correspondent à la valeur moyenne des classes  $[D_1D_1 D_2D_2]$  et  $[D_1d_1 D_2d_2]$ .

#### **7. 1. 3. Identification des QTLs expliquant les variations de teneur en caféine**

L'espèce parentale PSE est caractérisée par l'absence de caféine de ces grains. En revanche, l'espèce DEW et les hybrides F1 possèdent respectivement une teneur en caféine de 1,22 et 0,12 % MS.

La teneur en caféine implique directement ou indirectement quatre QTLs.

Le caractère « teneur en caféine » a été mis en relation avec :

- i) la durée du cycle de développement du fruit grâce à la nouvelle variable CAF/J, qui représente le flux d'accumulation journalière en caféine ;
- ii) l'association sous forme d'un complexe de la caféine avec certains isomères d'acide chlorogénique, évalué par les rapports CAF/CQA et CAF/CGA1. Le choix de ces deux rapports a été dicté par le fait que la caféine se complexe, d'après la littérature, avec le 5-CQA et le 3,5-DQA.

Un QTL est impliqué directement dans la teneur en caféine lorsque, des cinq variables étudiées (CAF, CAF/J, CAF/CQA, CAF/CGA1 et DCF), ce sont les variations de la variable CAF qui sont les mieux expliquées par le QTL candidat. Par contre, son implication est indirecte lorsque ce sont les variations des taux tels que CAF/J qui sont les mieux expliquées.

#### La complexation de la caféine avec les acides chlorogéniques (CAF/CQA et CAF/CGA1)

Il se trouve que le(s) QTL(s) de ces deux rapports se trouvent toujours associés, mais ce sont toujours les variations de CAF/CQA qui sont les mieux expliquées : le chlorogénate de caféine n'impliquerait pas d'acides dicaféylquiniques et la détection de (s) QTL(s) de CAF/CGA1 est la conséquence indirecte de la présence majoritaire du CQA dans CGA1.

Les variations de CAF/CQA sont expliquées par les QTL1.1 et QTL1.2 localisés respectivement aux locus AACCAC10 et AACCAT5 (groupes de liaison A et D). Il n'y a pas d'interaction entre les deux locus ( $F_{1,41} = 0,051$  ;  $p < 0,83$ )

Soit  $C_1$  et  $c_1$ , les allèles du locus AACCAC10 et soit  $C_2$  et  $c_2$ , les allèles du locus AACCAT5. Le tableau 7.7. présente les moyennes du rapport CAF/CQA pour les quatre classes génotypiques [ $C_1C_1 C_2C_2$ ], [ $C_1c_1 C_2C_2$ ], [ $C_1C_1 C_2c_2$ ] et [ $C_1c_1 C_2c_2$ ]. Les classes [ $C_1c_1 C_2C_2$ ], et [ $C_1C_1 C_2c_2$ ] montrent en fait une structure bimodale caractéristique de l'existence de recombinants entre les locus AFLP et les QTLs.

Le tableau 7.8 est obtenu en désignant  $C_1$ ,  $C_2$ ,  $c_1$  et  $c_2$ , les allèles DEW et PSE aux QTL1.1 et QTL1.2. Les génotypes [ $C_1C_1 C_2C_2$ ] ont un rapport proche de celui observé chez DEW (0,19). La présence de l'allèle  $c_1$  diminue le rapport de 0,040 et celle de l'allèle  $c_2$ , de 0,032

#### Le flux journalier de caféine (CAF/J)

Sa variation est expliquée par deux QTLs localisés respectivement aux locus AACCAT5 et AACCTA5 (groupe de liaison D et G). Le second locus se situe dans une

**Tableau 7.7.** : Moyennes de la variable CAF/CQA et effectifs (chiffre entre parenthèse), des classes définies par les allèles  $C_1$ ,  $C_2$  de DEW et  $c_1$ ,  $c_2$  de PSE aux locus AACCAT5 et AACCAT5. Les résultats des tests de Newman et Keuls sont indiqués par les lettres a, b et c en exposant.

		AACCAT5	
		$C_2C_2$	$C_2c_2$
AACCAT10	$C_1C_1$	0,167 <sup>c</sup> (13)	0,140 <sup>b</sup> (8)
	$C_1c_1$	0,133 <sup>b</sup> (7)	0,101 <sup>a</sup> (17)

**Tableau 7.9.** : Moyennes du caractère CAF/J (% MS/ J x 10<sup>-2</sup>) et effectifs, des classes définies par les allèles  $C_2$ ,  $P_1$  de DEW et  $c_2$ ,  $p_1$  de PSE aux locus AACCAT5 et AACCTA5. Les résultats des tests de Newman et Keuls sont indiqués par les lettres a, b et c en exposant.

		AACCAT5	
		$C_2C_2$	$C_2c_2$
AACCTA5	$P_1P_1$	0,296 <sup>c</sup> (6)	0,251 <sup>b</sup> (5)
	$P_1p_1$	0,235 <sup>b</sup> (10)	0,158 <sup>a</sup> (20)

**Tableau 7.8.** : Moyennes de la variable CAF/CQA et effectifs, des classes définies par les allèles  $C_1$ ,  $C_2$  de DEW et  $c_1$ ,  $c_2$  de PSE au QTL1.1 (proche de AACCAT10) et au QTL1.2 (proche de AACCAT5). Les résultats des tests de Newman et Keuls sont indiqués par les lettres a, b et c en exposant.

		QTL1.2	
		$C_2C_2$	$C_2c_2$
QTL1.1	$C_1C_1$	0,173 <sup>c</sup> (15)	0,134 <sup>b</sup> (15)
	$C_1c_1$	0,127 <sup>b</sup> (13)	0,101 <sup>a</sup> (17)

**Tableau 7.10.** : Moyennes du caractère CAF/J (% MS/ J x 10<sup>-2</sup>) et effectifs, des classes définies par les allèles  $C_2$ ,  $P_1$  de DEW et  $c_2$ ,  $p_1$  de PSE au QTL1.2 (proche de AACCAT5) et au QTL2.1 (proche de AACCTA5). Les résultats des tests de Newman et Keuls sont indiqués par les lettres a, b et c en exposant.

		QTL1.2	
		$C_2C_2$	$C_2c_2$
QTL2.1	$P_1P_1$	0,422 <sup>c</sup> (1)	0,195 <sup>a</sup> (8)
	$P_1p_1$	0,283 <sup>b</sup> (14)	0,144 <sup>a</sup> (18)

région du génome distordue en faveur de PSE. Enfin, ces deux locus ne présentent aucune interaction ( $F_{1,37} = 0,605$  ;  $p < 0,45$ ).

Le locus AACCAT5 étant déjà mis en évidence pour le caractère CAF/CQA, nous conserverons ici sa codification (QTL1.2) et nommerons QTL2.1 le QTL lié au marqueur AACCTA5.

Soient  $C_2$  et  $P_1$  les allèles de DEW et  $c_2$  et  $p_1$  les allèles de PSE aux locus AACCAT5 et AACCTA5 respectivement. Le tableau 7.9. présente les quatre classes génotypiques reconnues. Chacune des classes peut en fait être scindée en deux groupes.

Ceci conduit au tableau reconstitué 7.10 pour QTL1.2 et QTL2.2. Un seul génotype [ $C_2C_2 P_1P_1$ ] existe avec une valeur de CAF/J de 0,0042 proche de celle observée chez DEW (0,0040). Vis-à-vis du QTL1.2, on remarquera que les valeurs phénotypiques observées reflètent l'effet d'un gène avec deux allèles codominants et où l'un des allèles,  $c_2$ , est un allèle nul (l'homozygote  $c_2c_2$  aura une valeur attendue de zéro). Ce gène est un candidat pour expliquer l'absence de caféine de PSE. Pour l'effet de l'allèle  $p_1$  sur CAF/J, on constate que le rapport des valeurs phénotypiques des classes [ $P_1p_1$ ] et [ $P_1P_1$ ] est égal à 0,70.

#### La durée du cycle de fructification et son effet sur la teneur en caféine.

Le fait qu'un QTL soit impliqué dans CAF/J sous-entend que la durée du cycle intervient aussi dans la teneur finale. En conséquence, il est fort probable que le QTL détecté à côté du marqueur G13 pour la teneur en caféine, et donc à environ 10cM du QTL observé pour DCF, ne concerne en fait qu'un seul et même gène ou linkat gouvernant la durée du cycle de fructification (Akaffou, 1999). Ceci est confirmé par le fait que l'effet de G13 sur CAF disparaît lorsque DCF est utilisée comme covariable ( $F_{1,30} = 0,96$  ;  $p = 0,34$ ).

La recherche de QTL pour DCF a été abordée lors d'un précédent paragraphe. Il est apparu, pour résumer, qu'un QTL lié au marqueur ACCCTT1 explique 37 % de la variation observée pour DCF dans le BCDEW.

L'étude de son impact sur la teneur en caféine est plus complexe. Il faut en effet s'attendre à trouver deux groupes à l'intérieur des classes [ $F_1F_1$ ] et [ $F_1f_1$ ] qui ne refléteraient que les effets majeurs des allèles  $C_2$  et  $c_2$  de QTL1.2. C'est exactement ce que nous observons : non seulement il existe deux groupes par classe génotypique, mais les moyennes de ces sous-classes diffèrent, comme attendu, du simple au double. Dans ces conditions, la recherche de plants recombinants entre le locus RFLP et le QTL de DCF est hasardeuse. Les valeurs phénotypiques moyennes des classes [ $F_1F_1$ ] et [ $F_1f_1$ ] sont respectivement de 0,66 et de

0,44 % MS par jour, montrant que le QTL de DCF aurait pour effet de réduire la teneur en caféine de 33 %, lorsque l'allèle  $f_l$  est présent chez l'hybride.

### En résumé

Quatre QTLs situés sur des groupes de liaison indépendants expliqueraient la teneur finale en caféine :

- i) un QTL à effet majeur (un allèle nul) ayant des effets pléiotropiques sur CAF/J, CAF/CQA et CAF ;
- ii) un second QTL agissant sur le flux journalier ;
- iii) un troisième QTL modifiant le rapport CAF/CQA ;
- iv) et un QTL déterminant la durée du cycle de fructification.

## **7. 2. Discussion**

Après un bref récapitulatif des principaux résultats obtenus, nous aborderons :

- i) des limites d'acceptabilité des QTLs détectés, en soulevant le problème des seuils statistiques, du manque de saturation de la carte et de l'existence des segments présentant les distorsions,
- ii) des effets des QTLs sur la variance et la moyenne phénotypique,
- iii) de trois exemples de caractères d'intérêt biochimique et phénologique : les teneurs en acides chlorogéniques et en caféine et la durée du cycle de fructification.

### **7. 2. 1. Rappel des principaux résultats**

La cartographie de QTL des caractères biochimiques, phénologiques et de fertilité dans la population backcross issue du croisement PSE x DEW nous a permis de préciser les régions du génome impliquées. Parmi les 55 variables quantitatives définies dans l'annexe 3.1., vingt quatre ont été retrouvées liées à 33 marqueurs de la carte de liaison génétique pour une valeur du *LOD score* de 2,6. Parmi ces 24 variables, 40 % sont expliquées par plus d'un QTL. Les 33 marqueurs sont localisés dans 16 régions de la carte génétique, réparties sur 10 des 14 groupes de liaison. Aucun QTL n'est détecté sur les groupes I, L, M et N. Enfin, douze des 16 régions correspondent à des segments non distordus.

Les effets alléliques aux 33 QTLs ont été analysés pour l'ensemble des 24 caractères. La présence ou l'absence de phénomène de transgression a été mise en évidence. Pour les variables présentant deux et trois QTLs, l'ensemble des combinaisons alléliques a été analysé.

D'importants résultats ont été mis en évidence en particulier pour les variables DCF et CAF d'intérêt agronomique. Seul l'apport simultané des deux allèles PSE contribue à diminuer DCF de 50 jours. Pour CAF, les quatre QTLs mis en évidence n'interagissent pas entre eux, et ont des effets différents. Un QTL à effet majeur et pléiotropique a été détecté (locus AACCAT5).

### 7. 2. 2. Les limites de l'identification des QTLs

#### Choix des seuils de signification et risques d'erreur

La détection de QTL est basée sur des tests statistiques, que ce soit le F de Snedecor lorsque l'analyse de variance est utilisée, ou le maximum de vraisemblance (*LOD score*) avec la méthode de *l'interval mapping*. La prise de décision sur l'existence ou la non-existence d'un QTL est donc associée à un risque d'erreur : le risque  $\alpha$  de première espèce lorsqu'on accepte l'existence d'un QTL et de seconde espèce  $\beta$  lorsqu'on décide de rejeter son existence (acceptation de l'hypothèse nulle). La situation idéale est de diminuer les deux risques, ce qui revient à augmenter l'effectif de la population étudiée vers l'infini. Pour une variable quantitative donnée et des risques définis, la taille de l'échantillon nécessaire à la détection de l'effet d'un marqueur peut être déterminée classiquement.

Le problème particulier à la détection de QTL est lié au grand nombre de tests effectués. Pour une variable quantitative donnée, la recherche de QTL à partir de 100 marqueurs revient à effectuer 100 analyses de variance. Sur les 100 décisions prises à la suite de ces tests, on s'attend à un taux de mauvaises décisions équivalent au seuil de signification accepté. Il s'agit d'un problème similaire à la comparaison multiple de moyennes. Il suffirait alors d'appliquer au risque observé une correction de Ryan (1960). Le problème est en fait plus complexe, du fait de la dépendance entre marqueurs sur les groupes de liaison.

Lander et Botstein (1989) préconisent d'utiliser un seuil de 0,1 %, mais sans préciser l'effectif de la population, ni le nombre de marqueurs utilisés. En revanche, pour des seuils faibles (0,01 %), ou très faibles (0,001 %), on met en évidence de réelles associations. D'autres équipes (Edwards *et al.*, 1992 ; Zehr *et al.*, 1992) estiment que limiter le risque d'apparition de faux positifs augmente la probabilité de ne pas détecter un effet réel. Ces auteurs ont utilisé un seuil de 5 % dans leur recherche de QTL. Pour la même raison, d'autres travaux ont été conduits en utilisant un seuil de 1 % (Tanksley et Hewitt 1988 ; Keim *et al.* 1990 ; Diers *et al.* 1992).

Dans l'analyse présentée ici, une variable quantitative simulée a été créée. Trente-trois marqueurs ont été détectés comme QTLs potentiels au seuil de 5 %. Trois marqueurs l'ont été au seuil de 1 %, mais aucun à 0,1 %. Cette simulation confirme que pour des valeurs seuils relativement élevées (5 %), de fausses associations sont mises en évidence. Aussi seules les liaisons significatives au seuil de 0,1 % ont été considérées. Dans notre étude, sur les 35 décisions que nous avons prises en faveur de l'existence d'un QTL au seuil de 0,1 % en ANOVA, on attend quasiment aucune erreur. En revanche, la simulation que nous avons réalisée ne nous donne aucune information sur le risque de seconde espèce. Néanmoins, en prenant comme exemple le groupe de liaison F, trois QTLs peuvent être détectés au seuil de 1 % impliquant 16 variables, alors que nous n'avons identifié qu'un QTL ne concernant qu'une seule variable au seuil de 0,1 %. Sur les 15 décisions aboutissant à l'absence de QTL, le risque  $\beta$  doit en concerner la grande majorité. Nous conclurons qu'une analyse exhaustive d'identification de QTLs devrait classer les décisions en plusieurs catégories : les QTLs stochastiquement certains, les QTLs candidats et les QTLs à confirmer. La simulation permet de définir ces classes. Néanmoins, l'association de variables à ces QTLs doit aussi tenir compte des connaissances *a priori*. Ainsi, une relation au seuil de 1%, voire de 5%, d'un marqueur avec le 3-CQA, lorsque ce marqueur est aussi lié au 5-CQA au seuil de 0,1%, doit être prise en considération.

Une relation entre le risque  $\alpha$  et la valeur du *LOD score* existe. En effet, Ott (1985) montre que :  $4,6 \times LOD$ , suit asymptotiquement une distribution d'un  $\chi^2$  à un d.d.l. Au cours de notre recherche de QTL sous MAPMAKER/ QTL, deux valeurs de *LOD score* ont été utilisées : 2,0 et 2,6. La première valeur correspond à un risque de 0,2 %, tandis que la seconde à 0,05 %. Pour un *LOD* de 2,6 on se positionne dans une situation où les associations détectées correspondent quasiment à de vraies liaisons. Le *LOD* de 2,0 moins sévère, correspond au cas où la détection de fausse liaison demeure encore possible, mais très peu probable.

### L'importance de l'héritabilité *s.l.*

La détection de QTLs est d'autant plus facile que l'héritabilité *s.l.* est élevée. Celle-ci définit la part de la variance génotypique dans la variance phénotypique (Knight, 1948). L'écart à 100 % représente inversement l'effet de l'environnement. Classiquement, l'impact de l'environnement est estimé par des répétitions sur plusieurs arbres. Nous l'avons estimé sur plusieurs années pour le même arbre. Il n'y a pas indépendance totale entre années et l'influence relative de l'environnement s'en trouve diminuée. En un mot, la contribution

génotypique est théoriquement surévaluée. Néanmoins, elle permet la comparaison des caractères entre-eux pour leur sensibilité à l'environnement et nous pouvons vérifier que les QTLs détectés présentent tous de fortes héritabilités *s.l.* L'inverse n'est pas forcément vrai : la teneur en saccharose n'a pas été associée à un marqueur.

Le choix d'un environnement restreint favorise la localisation des QTLs majeurs. En revanche, elle ne permet pas d'identifier les gènes impliqués dans la plasticité des caractères.

La différence parentale explique les fortes valeurs d'héritabilité *s.l.* que nous avons observées. Ainsi, le choix d'un croisement entre deux espèces très différenciées pour les caractères abordés a permis la localisation de QTL malgré la faible taille de la population étudiée. Ceci est un avantage, mais aussi un inconvénient, car les QTLs mis en évidence concernent uniquement un polymorphisme interspécifique. Les gènes impliqués dans un polymorphisme intraspécifique ne peuvent être localisés.

### La complexité du caractère

Une bonne héritabilité *s.l.* n'est pas la seule condition à l'identification de QTLs. L'hérédité simple du caractère, c'est à dire le faible nombre de gènes impliqués dans la réalisation du phénotype, est fondamentale. La décomposition du caractère en quelques composantes structurelles améliore grandement la détection de QTLs (De Vienne et Cause, 1998). Le cas de la teneur en caféine illustre parfaitement cette situation : l'utilisation des variables CAF/J, CAF/CQA et DCF a permis la détection de quatre QTLs et d'en interpréter les effets. Il en est de même de la teneur en saccharose pour laquelle un QTL (non-cartographié) a pu être mis en évidence pour le taux SUC/J et pas pour la teneur finale.

### La saturation de la carte et l'effectif de la population

Théoriquement, l'efficacité de la détection et la précision de la localisation des QTLs dépend plus de l'effectif de la population que de la densité de marqueurs lorsque celle-ci atteint un espacement moyen de 20 à 25 CM (Darvasi *et al.*, 1993). Cette densité moyenne a été largement dépassée dans notre cas. Néanmoins, l'existence de 14 groupes de liaison en lieu et place des 11 attendus montre que la carte n'est pas entièrement saturée. En fait, la distance moyenne ne tient pas compte de la distribution des distances qui dépend fortement du type de marqueurs utilisés : les marqueurs de type EST, qui représentent des gènes codants, devraient être mieux répartis que les AFLP. L'absence de saturation de la carte nous a conduit à détecter des QTLs qui restent non localisés sur la carte génétique : c'est le cas du caractère SUC/J.

Il est aussi évident qu'avec 62 individus le taux minimal de recombinaison que l'on peut déceler entre deux marqueurs est de  $1/62^{\text{ème}}$  ce qui a conduit à un intervalle minimum de 1,6 cM. Pour doubler la précision sur les marqueurs, il faudrait multiplier par quatre la taille de la population.

### Les distorsions de ségrégation

Les distorsions de ségrégation ont pour effet d'aboutir à des classes très déséquilibrées en effectif, voire à des comparaisons non-orthogonales lorsque l'effectif d'une des classes est nul. L'impact d'une distorsion de ségrégation sur l'identification de QTLs est le même que l'effet d'un dispositif non-équilibré sur le test F : il dépend des conditions de normalité et surtout d'hétéroscédasticité de la variable quantitative : les variances inégales auront un impact d'autant plus important que la forte variance concernera la classe de plus faible effectif. Il est aussi prévisible que le biais apporté par les distorsions augmente avec l'ordre d'interaction. En cas d'inégalité des variances, le test F, relativement robuste pour les facteurs principaux, devient peu fiable pour les interactions. L'utilisation de tests non-paramétriques seraient préférable.

L'effet des distorsions de ségrégation sur l'identification d'un QTL diffère selon que les deux marqueurs qui encadrent le QTL ou lorsque seulement un des marqueurs sont distordus. Ceci est valable avec les deux méthodes d'identification (ANOVA et *interval mapping*). Avec la dernière méthode, le biais persiste lorsque les deux marqueurs qui encadrent le QTL sont proches d'un marqueur distordu.

### **7. 2. 3. La précision sur les effets des QTLs**

L'impact d'un QTL est estimé au travers de deux paramètres : la part de variance expliquée ( $R^2$ ) et effet additif sur la moyenne.

La part de variance globale expliquée ( $R^2$ ) pour chacun des caractères est comprise entre 22 et 56 % (Annexe 7.1.). Pour l'ensemble des caractères, la valeur de  $R^2$  globale est inférieure à la somme des valeurs de  $R^2$  individuelles. Cette situation reflète une co-linéarité qu'elle soit d'ordre structurelle (QTL appartenant à un même groupe de liaison) ou fonctionnelle (interactions épistatiques entre QTLs). Ceci est d'autant plus vrai, lorsque chez les caractères complexes, les effets indirects ne sont pas linéaires.

La valeur du  $R^2$  du QTL est différente du  $R^2$  du marqueur. Cette dernière valeur est <sup>inférieure</sup> supérieure à la première du fait de l'existence de recombinaisons entre le marqueur et le QTL.

Le biais sera d'autant plus grand que les distances génétiques entre le marqueur et le QTL seront grandes (De Vienne et Causse, 1998). Il est possible de corriger le  $R^2$  si l'on est capable de détecter de telles recombinaisons en regardant la forme des distributions à l'intérieur des classes génotypiques. Une telle méthode a été utilisée par Asins et Carbonell (1988), mais aussi lors de notre étude. Mais la présence de deux groupes à l'intérieur des classes génotypiques permettant d'identifier les individus recombinés n'est possible que lorsque les caractères parentaux sont suffisamment discriminants.

Il faudra aussi être très prudent lorsque des QTLs à effets majeurs existent parallèlement ; par exemple, dans le cas de la teneur en caféine, les deux populations à l'intérieur des classes génotypiques peuvent ne refléter que les effets majeurs du second QTL.

La discussion précédente conserve toute sa validité pour l'effet additif du QTL

#### **7. 2. 4. Les QTLs identifiés pour les acides chlorogéniques et la caféine**

##### La teneur en acides chlorogéniques

La répartition des différents QTL impliqués dans l'expression de la teneur en CGA confirme :

- i) L'indépendance génétique des voies de biosynthèse menant aux isomères de la classe des FQA, par rapport à celles menant aux deux autres classes (CQA et DQA). En effet, le QTL des quatre variables 3-FQA/J, 4-FQA/J, 5-FQA/J et FQA/J est uniquement localisé au niveau de la région (AAGCTT3 – ACTCTC4) du groupe de liaison C (Annexe 7.1. et Fig. 7.1.). Les QTLs des CQA et DQA sont localisés dans d'autres régions de la carte. Il est à noter qu'aucun QTL correspondant aux variables « brutes » (3-FQA, 4-FQA, 5-FQA et FQA) n'a été détecté.
- ii) La « parenté » dans les voies de biosynthèse entre les isomères CQA et DQA, qui ont en commun l'unité caféate. Les régions « centromériques » des groupes de liaison B et G présentent des QTLs pour le %DQA et le %CQA.
- iii) L'importance du 5-CQA pour la biosynthèse des isomères à unité caféate. Le groupe de liaison G possède un QTL correspondant à l'isomère 5-CQA (et donc aussi QTL du caractère CQA de part l'importance du 5-CQA).
- iv) L'indépendance structurelle des gènes menant à l'isomérisation du 3,5-DQA en 3,4-DQA et 4,5-DQA.

En revanche, aucun QTL expliquant les variations de la teneur en isomère 3,5-DQA n'a été détecté, alors que les différences entre espèces parentales sont du même ordre de grandeur.

### La teneur en caféine

Les variations de CAF sont expliquées par quatre QTLs et l'assignation de leur fonction peut être envisagée.

- i) Le QTL<sub>1</sub> (locus : AACAC10) serait impliqué dans la complexation de la caféine avec le 5-CQA. En effet, l'apport de l'allèle  $c_1$  de PSE au locus modifie les variables CAF/CQA et CAF/CGA1, mais plus fortement la première que la seconde.
- ii) Le QTL<sub>2</sub> (locus AACAT5) est impliqué dans la variation de CAF/J, c'est à dire dans la biosynthèse de la caféine; il agirait aussi indirectement sur le rapport CAF/CQA. Mais surtout, il a un effet majeur et permet d'expliquer l'absence de caféine chez PSE. De plus il est proche du marqueur CA71 que nous savons lié chez BCPSE au gène majeur contrôlant l'absence de caféine. Dans le programme d'introgression de l'absence de caféine, il s'agirait du locus ciblé.
- iii) Le QTL<sub>3</sub> (locus AACATA5) est aussi impliqué dans la variation de CAF/J, mais se trouve aussi lié au poids de 100 graines.
- iv) La durée du cycle de fructification à un effet direct sur la teneur finale. Les allèles provenant de PSE des deux derniers QTLs ont des effets non recherchés sur la granulométrie.

Deux voies d'introgression sont à considérer : QTL<sub>1</sub> et QTL<sub>2</sub>.

Lorsqu'un même marqueur est associé à plusieurs variables, on peut supposer la détection d'un QTL à effet pléiotropique ou l'existence de plusieurs QTLs différents, mais situés dans une zone de recombinaison réduite. Dans notre étude, le marqueur AACAT5 situé sur le groupe de liaison D est associé aux quatre variables : CAF, CAF/CQA, CAF/CGA1 et CAF/J. L'association entre CAF/J et CAF existe par définition et il s'agit d'un seul et même QTL pour les deux variables. Il en est de même entre CAF/CGA1 et CAF/CQA. La réponse est moins évidente entre CAF/J et CAF/CQA. Une cartographie fine de la région située entre AACAT5 et CA71, couvrant 10,8 cM, permettrait de préciser si l'effet du QTL sur CAF/J et CAF/CQA découle d'un effet pléiotropique ou d'une liaison génétique forte entre deux gènes. Ainsi, chez la tomate, Paterson *et al.* (1990) ont pu déterminer que l'association de QTLs pour des caractères distincts était due à une liaison étroite.

### 7. 3. En résumé

Plusieurs points importants ressortent de cette recherche de QTL :

- i) La double approche analyse de variance (ANOVA) et *interval mapping* (MAPMAKER/QTL), a permis de mettre en évidence des QTLs communs par les deux méthodes, auxquels s'ajoutent des QTLs à effets faibles, décelés uniquement par MAPMAKER/QTL. Cette dernière méthode permet de localiser les QTLs plus précisément, alors que l'ANOVA présente l'avantage de pouvoir détecter des QTLs de marqueurs non cartographiés (notamment pour une variable dépendante de la teneur en saccharose).
- ii) Trente trois QTLs expliquant 24 caractères ont été détectés pour un *LOD score* seuil de 2,6. Ces QTLs sont répartis sur 10 des 14 groupes de liaison et dans 16 régions différentes. Quinze caractères sont expliqués par un seul QTL, tandis que neuf sont multi-QTLs.
- iii) La teneur en CGA est complexe. Cette complexité est directement liée aux nombreux isomères caractérisant cette variable et aux nombreux QTLs correspondant détectés. Ces QTLs sont répartis sur plusieurs régions de la carte génétique.
- iv) La teneur en caféine est sous le contrôle de quatre QTLs. Deux présentent un grand intérêt agronomique.
- v) Des phénomènes de transgression pour les caractères : 3,4-DQA, 4,5-DQA, FQA/J et 4-FQA/J ont été mis en évidence. Ces résultats sont essentiels dans les schémas de sélection et montrent l'utilité des marqueurs moléculaires.

Plusieurs analyses complémentaires sont cependant nécessaires pour une localisation plus précise des locus impliqués dans l'expression des caractères analysés. Il faudrait augmenter le nombre de marqueurs moléculaires cartographiés sur la population BCDEW et le nombre d'individus analysés. Il serait aussi intéressant de pouvoir réaliser des études multilocales, afin de se rendre compte de l'expression différentielle des QTLs selon les milieux et sur d'autres croisements interspécifiques de caféiers, afin de varier le contexte génétique.

## **Chapitre 8 : Conclusions et perspectives**

---

## Chapitre 8 : Conclusions et perspectives

### 8. 1. Conclusions

L'amélioration de la qualité à la tasse de l'espèce cultivée *C. canephora* est une priorité pour les pays producteurs de basse altitude, dont l'économie est caféiculture dépendante. En effet, le point faible du café produit à partir de *C. canephora* est sa faible valeur gustative par rapport à celui de *C. arabica*. L'utilisation des ressources génétiques caféières rend possible cette amélioration. Parmi les espèces diploïdes, *C. pseudozanguebariae* (PSE) se présente comme un matériel de choix. Ses graines sont sans caféine et leur taux d'acides chlorogéniques est particulièrement bas. Elles ont aussi une teneur en saccharose et en trigonelline élevées. Ces caractéristiques biochimiques confèrent à *C. pseudozanguebariae* une forte ressemblance avec *C. arabica*. D'autres paramètres non étudiés ici, confirment cette tendance : richesse en acides aminés dicarboxyliques (acides aspartique et glutamique) et en lipides (30 %). En revanche, cette espèce contient un mozambioside (hétéroside diterpène) absent des caféiers cultivés et à l'origine d'une forte amertume (Prewo *et al.*, 1990). La teneur en ce composé est déterminée par un gène majeur à allèles codominants (Barre *et al.*, 1998a), détecté à un *LOD score* de 6,5. Il est indépendant des QTLs de la teneur en caféine et de la plupart des QTLs impliqués dans l'accumulation des acides chlorogéniques. L'exploitation d'une autre espèce diploïde de basse altitude, *C. liberica* var *dewevrei* (DEW) en croisement avec *C. pseudozanguebariae* devraient permettre :

i) d'appréhender le déterminisme génétique et les relations physiologiques de quelques composés biochimiques intervenant dans les qualités organoleptiques du café boisson, à l'aide de deux types de matériel végétal : des accessions spontanées de *C. arabica* et *C. canephora* et une population hybride (F1 et backcross 1) entre PSE et DEW.

ii) d'établir une carte de liaison génétique de ces croisement interspécifique à l'aide de marqueurs AFLP, afin de localiser les QTLs des caractères biochimiques, mais aussi les QTLs de caractères agro-phénologiques d'intérêt, évalués à la station IRD de Man en Côte-d'Ivoire.

Ces objectifs ont été atteints comme le montre le rappel des résultats suivants.

### **8. 1. 1. Relations entre composés biochimiques**

Les analyses effectuées sur les accessions spontanées de *C. arabica* et *C. canephora* et sur le matériel hybride (croisement PSE et DEW) ont permis de confirmer l'existence d'une grande variabilité en comparaison avec les cultivars. Ceci permet d'envisager une sélection de géniteurs intéressants (d'un point de vue biochimique) dans les schémas d'amélioration intraspécifique mais aussi interspécifique. Cette possibilité est confortée par l'indépendance physiologique et génétique observée entre les différentes familles de composés analysés, en particulier entre acides chlorogéniques et saccharose. Ces analyses ont aussi permis d'étudier et de préciser l'organisation de la variabilité en relation avec les voies de biosynthèse des acides chlorogéniques chez les caféiers, par une approche originale, l'analyse en composantes principales.

### **8. 1. 2. Déterminisme génétique des composés biochimiques**

L'héritabilité *s.l.* des caractères biochimiques, ainsi que les hypothèses d'additivité ont été respectivement estimées et testées pour les composés suivants : acides chlorogéniques, saccharose et trigonelline. Cette analyse sur une descendance hybride de caféiers est la première du genre pour des composés biochimiques.

La teneur en saccharose est un caractère additif (excepté chez les hybrides BCPSE) et très peu influencé par les effets environnementaux, puisque 86,5 % de la variance est d'origine génotypique.

La teneur en trigonelline n'est pas un caractère additif. Elle serait contrôlée par une hérédité cytoplasmique associée à des gènes nucléaires.

Les teneurs en acides chlorogéniques sont plus sensibles aux conditions environnementales : la part de variation d'origine génétique est de 43 à 78 %. L'additivité avec ou sans transformation mathématique est vérifiée pour les trois classes (CQA, DQA et FQA) et pour la majorité des isomères. La présence d'un gène majeur a été mise en évidence pour l'absence de la teneur en 3-FQA.

### 8. 1. 3. Cartographie génétique

Une carte génétique du croisement interspécifique PSE x DEW a été construite à partir d'une descendance backcross 1 de 62 hybrides : (PSE x DEW) x DEW.

La construction de la carte est basée sur une double stratégie utilisant deux valeurs seuils de *LOD score* (5,5 et 3,1), déterminées par une simulation de cartographie. La carte établie comprend 181 loci : 167 de type AFLP, 13 de type RFLP et le dernier de type morphologique. Ces marqueurs ont été assemblés en 14 groupes de liaison, chacun comprenant de 4 à 31 loci et couvrant 1144 cM. Les distorsions de ségrégation affectent 30 % des marqueurs. Ces distorsions sont de deux types 3 : 1 et 1 : 3 et favorisent équitablement l'un et l'autre des parents. L'origine d'une telle distorsion suggère l'implication d'évènements de conversion génétique.

### 8. 1. 4. Recherche de QTLs

La carte génétique de ce croisement a permis de localiser les QTLs impliqués dans les variations des teneurs des composés biochimiques étudiés. Pour les acides chlorogéniques (CGA), six régions de la carte génétique sont impliquées dans ses variations. Ceci montre la complexité du caractère, directement lié à l'interdépendance des isomères et classes d'isomères dans la chaîne de biosynthèse. Pour les alcaloïdes, les teneurs en caféine et en trigonelline sont respectivement expliquées par quatre et deux QTLs. Pour la teneur en saccharose, aucun QTL n'a été mis en évidence sur la carte. En revanche, la variable SUC/J présente un QTL correspondant à un marqueur non cartographié (marqueur : ACCCTA1P). Tous les QTLs « biochimiques » sont localisés sur des zones sans distorsions du génome, à l'exception des deux QTLs influençant des flux journaliers : un pour FQA et un autre pour CAF/J. D'autres QTLs intéressants ont été mis en évidence, pour la durée du cycle de fructification (DCF) et le poids de cent grains (deux QTLs).

### 8. 3. Perspectives

Trois pistes de recherche peuvent être envisagés :

- i) la poursuite des évaluations biochimiques,
- ii) la poursuite de la cartographie génétique,
- iii) le développement de la recherche de gènes candidats chez les caféiers.

L'existence à l'heure actuelle de 240 hybrides backcross 1 et productifs entre *C. pseudozanguebariae* et *C. canephora* (PSE x CAN), fait de ce croisement la base des futures recherches.

Le but final est le transfert de ces gènes pour l'obtention de variétés améliorées de *C. canephora*. Deux voies sont actuellement possibles : l'introgession ou la transformation génétique. Les informations génétiques et physiologiques - nombre de gènes impliqués, localisation du (ou des) gène(s) dans le génome, leur position par rapport au centromère, leur indépendance ou leur liaison génétique avec d'autres gènes, en particulier de stérilité, leur présence dans une zone du génome sujette à des distorsions de ségrégation - sont autant d'informations déterminantes dans le choix de l'une des voies et le succès de l'opération.

### 8. 3. 1. Evaluations biochimiques

Les évaluations biochimiques seront étendues à d'autres composés biochimiques impliqués dans la qualité, comme les acides aminés, les acides gras et les diterpènes ; ces nouvelles évaluations serviront à la recherche de nouveaux QTLs. Elles seront aussi poursuivies sur d'autres espèces du sous-genre *Coffea* (sur une base de quatre plants par espèce) et sur d'autres organes végétaux comme les feuilles. Dans ce dernier cas, la recherche donnera une idée des « flux » éventuellement mis en jeu entre les organes. Les évaluations concerneront aussi différents stades de maturation de la graine chez PSE et CAN. Cette analyse donnera une idée des rapports entre la synthèse et la dégradation (« turn-over ») des composés.

### 8. 3. 2. Cartographie de différents croisements interspécifiques

La cartographie moléculaire sera étendue à d'autres types de matériels hybrides et d'autres marqueurs moléculaires. Cinq matériels hybrides sont disponibles et maintenus au centre IRD à Man en Côte-d'Ivoire. Cinq des six croisements font intervenir l'espèce cultivée *C. canephora* avec : PSE (croisement modèle), *C. heterocalyx* (autogamie), *C. eugenioides* (parenté avec *C. arabica*) et *C. liberica* (granulométrie importante et fructification groupée).

La comparaison entre cartes génétiques nécessitera de disposer de marqueurs « ancre » (locus spécifique) régulièrement répartis sur le génome. De tels marqueurs pourraient être de type microsatellite et/ ou PCR ciblée (après séquençage de sondes RFLP et/ ou de bandes AFLP et définition d'amorces spécifiques). L'obtention de tels marqueurs aura l'avantage de faciliter toute constitution de nouvelle carte, tout en fournissant des données pour la

cartographie comparée. La comparaison des organisations génomiques donnera des informations d'ordre structurel (existence d'inversions, de translocations) et fonctionnel (localisation des locus présentant des distorsions de ségrégation, différence de taux de recombinaison). Elle permettra de comparer la position et la fréquence des distorsions de ségrégation en fonction de la « distance génétique » à l'espèce cultivée ou en fonction de la taille du génome.

### 8. 3. 3. Clonage et expression de gène d'intérêt

A partir de la carte PSE x CAN, il sera possible de vérifier chez les caféiers, si certains des gènes séquencés chez d'autres espèces et présents dans les banques de données sont liés de façon forte aux QTLs. Dans ce cas, les gènes « reconnus » deviendront de véritables gènes cibles (candidats) et donc étudiés de façon prioritaire. Cette approche découle du choix délibéré d'enzymes ou de protéines potentiellement impliquées dans la diversité biochimique.

Les résultats déboucheront sur une approche rationnelle de la valorisation des ressources génétiques, la recherche des gènes impliqués, l'analyse des possibilités de transfert, le choix d'une stratégie et l'obtention de variétés améliorées. Cependant, l'amélioration de *C. canephora* pour la qualité ne doit pas se faire au détriment de leur adaptation et de la résistance aux maladies. En effet, les acides chlorogéniques impliqués dans l'amertume du café (Leloup *et al.*, 1995) constituent aussi des agents de résistance aux maladies (Zuluaga *et al.*, 1971 ; Echandi et Fernandez, 1962 ; Macheix *et al.*, 1990).

## **Chapitre 9: Références bibliographiques**

---

## Chapitre 9: Références bibliographiques

- Akaffou D.S.** (1999) Etude des hybrides interspécifiques entre *Coffea pseudozanguebariae* Bridson et *C. liberica* var. *dewevrei* de Wild et Th. Dur. Et analyse des possibilités d'introgression. Thèse de l'Université de Cocody (Côte d'Ivoire), 164 pp.
- Alonso-Blanco C., Peeters A.J.M., Koornneef M., Lister C., Dean C., Van den Bosch N., Pot J. et Kuiper M.T.R.** (1998) Development of an AFLP based map of Ler, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a Ler/ Cvi recombinant inbred lines population. *Plant Journal*, 14 : 259-271.
- Anthony F.** (1992) Les ressources génétiques des caféiers. Collecte, gestion d'un conservatoire et évaluation de la diversité génétique. ORSTOM Editions, TDM n°81, 320 pp.
- Anthony F. et Le Pierrès D.** (1987) La diversité génétique de l'espèce *Coffea congensis* Froehner. Evaluation en Côte d'Ivoire de caféiers sauvages originaires de la république Centrafricaine. *Café Cacao Thé*, 4: 251-266.
- Anthony F., Berthaud J., Guillaumet J.L. et Lourd M.** (1987) Collecting wild *Coffea* species in Kenza and Tanzania. *Plant Genetic Resources News Letter*. 69 : 23-29.
- Anthony F., Clifford M.N. et Noirot M.** (1993) Biochemical diversity in the genus *Coffea* L.: chlorogenic acids, caffeine and mozambioside contents. *Genetic Resources and Crop Evolution*, 40 : 61-70.
- Arctander** (1969) *Perfume and flavors chemicals*. Monclair N.J.
- Armorin H.V. et Silva D.M.** (1968) The relationship between the polyphenol oxidase activity and the quality of the beverage. *Nature*, 219 : 381-382.
- Armorin H.V., Cruz A.R., St. Angelo A.J., Dias R.M., Melo M., Teixeira A.A., Gutierrez L.E. et Ory R.L.** (1977) Biochemical, physical and organoleptical changes during raw coffee quality deteriorations. *ASIC* 8, Paris : 183-186.
- Armorin H.V., Teixeira A.A., Guercio M.A., Cruz V.F. et Malavolta E.** (1974) Chemistry of Brazilian green coffee and the quality of the beverage II phenolic compounds. *Turrialba* 24: 217-221.
- Arnold U. et Ludwig E.** (1996) Analysis of free amino acids in green coffee beans. II. Changes of the amino acid content in arabica coffees in connection with post-harvest model treatment. *Z. Lebensm. Unters. Forsch.*, 203: 379-384.
- Arnold U., Ludwig E., Kühn R. et Möschwitzer U.** (1994) Analysis of free amino acids in green coffee beans. I. Determination of amino acids after precolumn derivatization using 9-fluorenylmethylchloroformate. *Z. Lebensm. Unters. Forsch.*, 199: 22-25.
- Asins M.J. et Carbonell E.A.** (1988) Detection of linkage between restriction fragments length polymorphism markers and quantitative traits. *Theor. Appl. Genet.*, 76: 623-626.

- Baltes W. et Bochmann G.** (1987a) Model reactions on roast aroma formation I: Reaction of serine et threonine with sucrose under the conditions of coffee roasting and identification of new coffee aroma compounds. *J. Agric. Food Chem.*, 35: 340-346.
- Baltes W. et Bochmann G.** (1987b) Model reactions on roast aroma formation II : Mass spectrometric identification of furans and furanones from the reaction of serine and threonine under the conditions of coffee roasting. *Z. Lebensm-unters. u. forsch.*, 184: 179-186.
- Baltes W. et Mevissen L.** (1988) Model reactions on roast aroma formation VI: Volatile reaction products from the reaction of phenylalanine with glucose during cooking and roasting. *Z. Lebensm-unters. u. forsch.*, 187: 209-214.
- Baltes, W.** (1990) Roast aroma formation. The role of amino acids during the Maillard reaction. In the maillard Reaction in Food Processing, Human Nutrition and Physiology, ed. Finot P.A., Aeschbacher H.U., Hurrell R.F. et Liardon R. Birkhäuser verlag, basel : 43-62.
- Balyaya K.J. et Clifford M.N.** (1995) Individual chlorogenic acids and caffeine contents in commercial grades of wet and dry processed Indian green *Robusta* coffee. *J. Food Sci. Technol.* 32: 104-108.
- Barre P.** (1997) Les transferts de gènes entre espèces de caféiers diploïdes : étude des hybrides F1 et BC1 entre *C. pseudozanguebariae* Bridson et *C. liberica* Bull ex Hiern (*C. dewevrei* De Wild et Th. Dur). PhD thesis, Ecole Nationale Supérieure Agronomique de Montpellier, France: 117 pp.
- Barre P., Akaffou S., Louarn J., Charrier A., Hamon S. et Noirot M.** (1998a) Inheritance of caffeine and heteroside contents in an interspecific cross between a cultivated coffee species *Coffea liberica* var 'dewevrei' and a wild species caffeine free *C. pseudozanguebariae*. *Theor. Appl. Genet.*, 96 : 306-311.
- Barre P., Layssac M., D'Hont A., Louarn J., Charrier A., Hamon S. et Noirot M.** (1998b). Relationship between parental chromosomal contribution and nuclear DNA content in the coffee interspecific hybrid : *C. pseudozanguebariae* x *C. liberica* var. *dewevrei*. *Theor. Appl. Genet.*, 96 : 301-305.
- Barre P., Noirot M., Louarn J., Duperray C. et Hamon S.** (1996) Reliable flow cytometric estimation of nuclear DNA content in coffee trees. *Cytometry*, 24 : 32-38.
- Bassi D., Bartolozzi F. et Muzzi E.** (1996) Patterns and heritability of carboxylic acids and soluble sugars in fruits of apricot (*Prunus armeniaca* L.). *Plant Breeding*, 115 : 67-70.
- Baumann T.W. et Röhrig L.** (1989) Formation and intracellular accumulation of caffeine and chlorogenic acid in suspension cultures of *Coffea arabica*. *Phytochemistry*, 28 : 2667-2669.
- Baumann T.W. et Wanner H.** (1972) *Planta*, 108 : 11-20.
- Baumann T.W., Möslé S.S., Schulthess B.H. et Aearts R.J.** (1993) Interdependence of caffeine and chlorogenic acid (5-CQA) metabolism in coffee. *Proc. Int. Congr. ASIC*, 15 : 134-140.
- Beckmann J.S. et Soller M.** (1990) Towards a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites. *Biotechnology*, 8: 930-932.
- Bennett M.D. et Leitch I.** (1995) Nuclear DNA amounts in Angiosperms. *Ann. Bot.*, 76 : 113-176.
- Berthaud J.** (1977) Caractéristiques comparées des hybrides interspécifiques tétraploïdes et hexaploïdes *Coffea arabica* L. x *C. canephora* Pierre. *Proc. coll. (Abidjan) ASIC* 8 : 393-397.

- Berthaud J.** (1986) Les ressources génétiques pour l'amélioration des caféiers africains diploïdes. Coll. Trav. & Doc., ORSTOM, Paris, 379 pp.
- Berthaud J. et Charrier A.** (1988) Genetic resources of *Coffea*. In *Coffee* Vol. 4, Clarke R.J. et Macrae R., Elsevier Applied Science, Londres, 1-40.
- Berthaud J. et Guillaumet J.L.** (1978) Les caféiers sauvages en Centrafrique. Résultats d'une mission de prospection (janvier-février 1975). *Café Cacao Thé*, 22 : 171-186.
- Berthaud J., Anthony F. et Lourd M.** (1983) Les caféiers sauvages de Tanzanie. Résultats d'une mission de prospection effectuée du 5 mars au 11 avril 1982. *Café Cacao Thé* 27 : 245-258.
- Berthaud J., Anthony F., Charrier A., Couturon E., Le Pierrès D. et Louarn J.** (1989) Les caféiers en Afrique : diversité génétique et amélioration des plantes. *Bull. Soc. bot. Fr.*, 136 : 239-250.
- Berthaud J., Le Pierrès D., Guillaumet J.L. et Lourd M.** (1980) Les caféiers sauvages du Kenya : prospection et mise en culture. *Café Cacao Thé*, 24 : 101-112.
- Bettencourt A.J.** (1984) características agronomicas de derivadas de cruzamentos entre híbrido de Timor e as variedades caturra, Villa sarchi e catuai. CIFIC 50eiras), Simposio sobre Ferrugens do Cafeeira: 353-373.
- Binelli G., Gianfranceschi L., Pe M.E., taramino G., Busso C., Stenhouse J. et Ottaviano E.** (1992) Similarity of maize and sorghum genomes as revealed by maize RFLP probes. *Theor. Appl. Genet.*, 84: 10-16.
- Bonierbale M.W., Plaisted R.L. et Tanksley S.D.** (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics*, 120: 1095-1103
- Botstein D., White R.L., Skolnick M. et Davis R.W.** (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.*, 32: 314-331.
- Bouharmont J.** (1959) Recherche sur les affinités chromosomiques dans le genre *Coffea*. Publication INEAC, Série scientifique, 77: 94 pp.
- Bouharmont J.** (1963) Somatic chromosomes of some *Coffea* species. *Euphytica*, 12 : 254-257.
- Braudeau J., Cambrony H.R., Capot J., Dublin P., Etasse C et Foury C.** (1962) Les principes de la sélection des caféiers canephoroides et libérico-excelsoïdes. Leur application aux travaux des centres de recherches de l'IFCC en Côte d'Ivoire, à Madagascar et en RCA. *Café Cacao Thé*, 6: 89-104, 169-186, 271-286.
- Breto M.P., Asins M.J. et Carbonell E.A.** (1994) salt tolerance in *Lycopersicon* species. III. Detection of QTLs by means of molecular markers. *Theor. Appl. Genet.*, 88: 395-401.
- Bridson D.M.** (1987) Nomenclatural notes on *Psilanthus*, including *Coffea* sect. *Paracoffea* (*Rubiaceae* tribe *Coffeae*). *Kew Bulletin*, 36 (4) : 817-860
- Bridson D.M.** (1994) Additional notes on *Coffea* (*Rubiaceae*) from tropical east Africa. *Kew Bulletin.*, 49 : 331-341.
- Bridson D.M. et Verdcourt B.** (1988) *Flora of tropical East Africa-Rubiaceae* (Part 2), Polhill R.M. (eds.), 727 pp.
- Bruneau J.M., Worell A.C., Cambou B., Lando D. et Voelker T.A.** (1991) Sucrose phosphate synthase, a key enzyme for sucrose biosynthesis in plants. *Plant Physiology*, 96 : 473-478.
- Buttler L.** (1977) Viability estimates for 60 tomato mutants. *Can. J. Genet. Cytol.*, 19: 8-14.

- Capot J.** (1962) Le caféiers Robusta dans la cuvette centrale congolaise. *Bulletin Information de l'INEAC*, 11: 33-40.
- Capot J.** (1972) L'amélioration du caféier en Côte-d'Ivoire : les hybrides « arabusta ». *Café Cacao Thé* 16 : 3-17.
- Carneiro M.F.** (1997) Coffee biotechnology and its application in genetic transformation. *Euphytica* 96: 167-172.
- Carvalho A.** (1982) melhoramento do cafeeiro cruzamentos entre *Coffea arabica* x *C. canephora*. *Proc. of the Int Congr. ASIC* 10: 363-368.
- Carvalho A. et Monaco L.C.** (1967) Genetic relationships of selected *Coffea* species. *Ciencia e cultura* 19 : 151-165.
- Causse M., Fulton T.M., Chi Y.G., Ahn S.N., Chunwongse J., Wu K., Xiao J., Yu Z., Ronald P.C., Harrington S.E., Second G., Mc Couch S.R. et Tanksley S.D.** (1994) Saturated molecular map of the rice genome based on an interspecific backcross population. *Genetics*, 138: 1251-1274.
- Causse M.A., Rocher J.P., Henry A.M., Charcosset A., Prioul J.L. et De Vienne D.** (1995) Genetic dissection of the relationship between carbon metabolism and early growth in maize, with emphasis on key-enzyme loci. *Molecular Breeding*, 1: 259-272.
- Charmettant P et Le Pierrès D.** (1991) Rapport de mission en Nouvelles-Calédonie. *Mult. ORSTOM – IRCC/ CIRAD*: 21 pp.
- Charrier A.** (1976) la structure génétique des caféiers spontanés de la région malgache (*Mascarocoffea*). Leur relations avec les caféiers d'origine africaine (*Eucoffea*). Thèse, Université paris-Sud: 223 pp.
- Charrier A.** (1978) La structure génétique des caféiers spontanés de la région malgache et leurs relations avec les caféiers d'origine africaine (*Eucoffea*). Thèse Doct., Mémoire ORSTOM, Paris, n°87, 223 p.
- Charrier A. et Berthaud J.** (1975) Variation de la teneur en caféine dans le genre *Coffea*. *Café Cacao Thé*. 19 : 251-264.
- Chrestin H.** (1985) La vacuole dans l'homéostasie et la sénescence des cellules laticifères d'hévéa. Ed. ORSTOM, Collection Etudes et Thèse: 575 pp.
- Chrestin H. et Noirot M.** (1982) Utilisation comparés de méthodes d'analyses des données dans l'étude de quelques composantes physiologiques et biochimiques du rendements chez *Hevea brasiliensis*. *Int. Biometrics Conf. (Toulouse)*, 11: 6 pp.
- Clifford M.N.** (1985a) Chemical and physical aspects of green coffee and coffee products. In : *Coffee, botany, biochemistry and production of beans and beverage*. Clifford et Wilson (eds), Avi publishing company, Westport, Connecticut, pp 305-374.
- Clifford M.N.** (1985b) Chlorogenic acids. In: *Coffee I Chemistry*. Clarke et Macrae (eds), Elsevier Applied Science Publisher, pp 153-202.
- Clifford M.N. et Jarvis T.** (1988) The chlorogenic acids content of green robusta coffee beans as possible index of geographic origin. *Food Chem.* 29: 291-298.
- Clifford M.N. et Ohiokpehai O.** (1982) Coffee as-tringency. *Royal Society of Chemistry Special Techniques*

- group Meeting*. Metal Box Ltd., Wantage.
- Clifford M.N. et Staniforth P.S.** (1977) A critical comparison of six spectrophotometric methods for measuring chlorogenic acids in green coffee beans. *Proc. Int. Congr. ASIC*, 8 : 109-113.
- Clifford M.N., Kellard B. et Ah-Sing E.** (1989a) Caffeoyl-tyrosine from green robusta coffee beans. *Phytochemistry*, 28 : 1989-1990.
- Clifford M.N., Williams T. et Bridson T.** (1989b) Chlorogenic acids and caffeine as possible taxonomic criteria in *Coffea* and *Psilanthus*. *Phytochemistry*, 28 : 829-838.
- Cloutier S., Cappadocia M. et Landry B.S.** (1997) Analysis of RFLP mapping inaccuracy in *Brassica napus* L. *Theor Appl Genet* 95: 83-91
- Colonna J.P.** (1979) L'acide chlorogénique et les depsides de divers caféiers africains et malgaches: leur participation au métabolisme et leur signification biologique. Ph.D. Thesis, Université Paul Sabatier, Toulouse, 207 pp.
- Colonna J.P.** (1986) Biosynthèse et renouvellement de l'acide chlorogénique et des depsides voisins dans le genre *Coffea*. *Café Cacao Thé*, 30 : 83-92 et 247-258.
- Comstock R.E., Robinson H.F. et Harvey P.H.** (1949) A breeding procedure designed to make maximum use of both general and specific combining ability. *Agronomy journal*, 41 : 360-367.
- Conagin C.H.T.M. et Mendes A.J.T.** (1981) Pesquisas citologicas e geneticas em tres especies de *Coffea*. Autoincompatibilidade em *Coffea cabephora*. *Bragantia*, 20 : 787-804.
- Cordier L.** (1961) Les objectifs de la sélection caféière en Côte d'Ivoire. *Café Cacao Thé*, 5: 147-158.
- Correia M.N.G., Leitao M.C.A. et Clifford M.N.** (1995) Caffeoyl-tyrosine and Angola II as characteristic markers for Angolan robusta coffees. *Food Chem.*, 53 : 309-313.
- Cramer P.J.S.** (1957) A review of literature of Coffee research in Indonesia. Interamerican Institute of agricultural Sciences, Turrialba, Costa-Rica 262 pp.
- Cronin D.A. et Ward M.K.** (1971) The characterisation of some mushroom volatile. *J. Sci. Food Agric.*, 22 : 477-479.
- Cros J.** (1994) Implications phylogénétiques des variations de l'ADN chloroplastique chez les caféiers (genre *Coffea* L. et *Psilanthus* Hook. f.). Thèse Doct., Univ. Montpellier II 160 pp.
- Cros J., Combes M.C., Chabrilange N., Duperray C., Monnot des Angles A. et Hamon S.** (1995) Nuclear DNA content in the subgenus *Coffea* (*Rubiaceae*): inter- and intra-specific variation in African species. *Can. J. Bot.* 73 : 14-20.
- Cros J., Combes M.C., Trouslot P., Anthony F., Hamon S., Charrier A. et Lashermes P.** (1997) Phylogenetic analysis of chloroplast DNA variation in *Coffea* L. *Molecular phylogenetics and evolution*. (sous presse).
- Czok G.** (1965) On the question of the share of chlorogenic acid, trigonelline and coffee oil in the physiological effects of coffee beverages. *Proc. of the Int Congr. ASIC*, 2 : 239-246.
- Dart, S.K. et Nursten H.E.** (1985) Volatile components. In: R.J. Clarke and R. Macrae (eds), *Coffee: Chemistry* (Vol. I), 223-265, Elsevier Applied Science Publishers Ltd, London, UK.

- Darvasi A., Weinreb A., Minke V., Weller J.I. and Soller M.** (1993) Detection marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics*, 134 : 943-951.
- De Lumen B.O., Stone E.J., Kazeniac S.J. et Forsythe R.H.** (1978) Formation of volatile flavor compounds in green beans from linoleic and linolenic acids. *J. Food Sci.*, 43 : 698-702.
- De Maria C.A.B., Trugo L.C., Aquino Neto F.R., Moreira R.F.A. et Alviano C.S.** (1996) Composition of green coffee water-soluble fractions and identification of volatiles formed during roasting. *Food Chemistry*, 55 : 203-207
- De Maria C.A.B., Trugo L.C., Moreira R.F.A. et Werneck C.C.** (1994) Composition of green coffee fractions and their contribution to the volatile profile formed during roasting. *Food Chemistry*, 50 : 141-145
- De Reffye P.** (1974) Le contrôle de la fructification et de ses anomalies chez les *Coffea arabica*, *robusta* et leur hybrides Arabustas. *Café Cacao Thé*, 28 : 237-254.
- De Ross B., Van der Weg G., Urgert R., Van de Bovenkamp P., Charrier A. et Katan M.B.** (1997) Levels of cafectol, kahweol, and related diterpenoids in wild species of the coffee plant *Coffea*. *J. Agric. Food Chem.*, 45 : 3065-3069.
- De Vienne D.** (1998) Les marqueurs moléculaires en génétiques et biotechnologies végétales. Ed. INRA, Mieux comprendre : 200 pp.
- De Vienne D. et Causse M.** (1998) la cartographie et la caractérisation des locus contrôlant la variation des caractères quantitatifs. Dans : Les marqueurs moléculaires en génétiques et biotechnologies végétales. Ed. INRA, Mieux comprendre : 89-118.
- Dellaporta S.J., Wood J. et Hicks** (1983) A plant DNA miniprep: Version II. *Plant MOL. Biol. Repr.* 1 : 19-21.
- Devos K.M. et Gale M.D.** (1997) Comparative genetics in the grasses. *Plant Mol. Biol.*, 35: 3-15.
- Devreux M., Vallayes G., Pochet P et Gilles A.** (1959) Recherches sur l'autostérilité du caféier *robusta* (*C. canephora* Pierre). Publication INEAC, Série scientifique, 78 : 44 pp.
- Diers B.W., Mansur L., Imsande J. et Shoemaker R.C.** (1992) Mapping Phytophthora resistance loci in soybean with restriction fragment length polymorphism markers. *Crop Science*, 32 : 377-383.
- Dirlewanger E., Isaac P.G., Ranade S., Belajouza M., Cousin R. et De Vienne D.** (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.
- Doebley J., Stec A et Gustus C.** (1995) Teosinte branched 1 and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics*, 141 : 333-346.
- Dolezel J.** (1991) Flow cytometric analysis of nuclear DNA content in higher plants. *Phytoche. Anal.*, 2 : 143-154.
- Dolezel J., Binarova P. et Lucretti S.** (1989) Analysis of nuclear DNA content in plant cells by flow cytometry. *Biologia Plantarum* 32 : 113-120.
- Dolezel J., Sgorbati S. et Lucretti S.** (1992) Comparaison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiol. Plantarum* 85 : 625-631.

- Dufour P., Deu M., Grivet L., D'Hont A., Paulet F., Bouet A., Lanaud C., Glaszmann J.C. et Hamon P.** (1997) Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. *Theor. Appl. Genet.*, 94 : 409-418.
- Dufour P., Grivet L., D'Hont A., Deu M., Trouche G., Glaszmann J.C. et Hamon P.** (1996) Comparative genetic mapping between duplicated segments on maize chromosomes 3 et 8 and homeologous regions in sorghum and sugarcane. *Theor. Appl. Genet.*, 92: 1024-1030.
- Dussert S., Lashermes P., Anthony F., Montagnon C., Trouslot P., Combes M.-C., Berthaud J., Noirot M. et Hamon S.** (1999) Le caféiers, *Coffea canephora*. Diversité génétiques des plantes tropicales cultivées. Eds. Hamon P., Seguin M., Perrier X. et Glaszmann C., CIRAD, 175-194.
- Echandi E. et Fernandez C.E.** (1962) Relacion entre el contenido de acido chlorogenico y la resistencia a la Haga macana o cancer de los cafetos causados por *Ceratocystis fimbriata*. *Turrialba*, 12 : 87-90.
- Edwards M.D., Helentjaris T., Wright S. et Stuber C.W.** (1992) Molecular-marker facilitated investigations of QTL in maize. VI. Analysis based on genome saturation and restriction fragment length polymorphism markers. *Theor. Appl. Genet.*, 83 : 765-774.
- Edwards M.D., Stuber C.W. et Wendel J.F.** (1987) Molecular-marker facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distributions and types of gene action. *Genetics*, 116: 113-125.
- Eujayl I., Baum M., Powell W., Erskine W. et Pehu E.** (1998) A genetic linkage map of lentil (*Lens* sp.) based on RAPD and AFLP markers using recombinant inbred lines. *Theor Appl Genet* 97: 83-89.
- Evans L et Tramontano WA** (1984) Trigonelline and promotion of cell arrest in G2 of various legumes. In: *Bioregulators chemistry and uses*, Ed. Ory RL and Rittig FR, ACS symposium series : 75-82.
- Fauré S.** (1993) Contribution à la cartographie génétique du génome des bananiers diploïdes (*Musa acuminata*) à l'aide de marqueurs moléculaires. Thèse de l'ENSA de Rennes, 65 pp.
- Feldman J.R., Ryder W.S. et Kung J.T.** (1969) Importance of nonvolatile compounds to the flavour of coffee. *J. Agric. Food Chem.*, 17 : 733-739.
- Ferwerda F. P.** (1948) Coffee breeding in Java. *Economic Botany*, 2 : 258-272.
- Ferwerda F. P.** (1969) Breeding of *Coffea canephora*. *Outlines of perennial crop breeding in the tropics*, eds. Ferwerda F. P. et Wit F., Wageningen, 222-235.
- Fiddler K., Parker W.E., Wasserman A.E. et Doerr R.C.** (1967) Thermal decomposition of ferulic acid. *J. Agric. Food Chem.*, 15 : 757-761.
- Fisher R.A.** (1936) *Statistical methods for research workers*. 11<sup>th</sup> Ed. (eds.) Oliver and Boyd, Edinburgh, London, 260 pp.
- Flament I.** (1991) Coffee, cocoa and tea. Volatile compounds in foods and beverages. Ed. Maarse H. et Dekkerf M., New-York: 617-653.
- Fleuriet A. et Macheix J.-J.** (1981) Quinyl esters and glucose derivatives of hydroxycinnamic acids during growth and ripening of tomato fruit. *Phytochemistry*, 20 : 667-671.
- Folstar P.** (1985) Lipids. I: Chemistry, ed. Clarke R.J. et Macrae R., Elsevier Applied Science (London) : 203-223.

- Folstar P., Pilnik W. et Van der Plas H.C.** (1977) Liquid chromatographic coffee wax analysis. Proc. of the Int Congr. ASIC 8: 121-124.
- Foolad M.R., Arulsekhar S., Becerra V. et Bliss F.A.** (1995) A genetic map of *Prunus* based on an interspecific cross between peach and almond. Theor Appl Genet 91: 262-269
- Forss D.A.** (1969) Role of lipids in flavors. J. Agric. Food Chem., 17 : 681-685.
- Fressanges R.** (1954) la sélection du caféiers en Côte d'Ivoire. Bulletin Scientifique de la France d'Outre-mer, 5 : 223-231.
- Friedman M.** (1997) Chemistry, biochemistry, and dietary role of potato polyphenols. A review. J. Agric. Food Chem., 45 : 1523-1540.
- Fujimaki M., Kato S. et Kurata T.** (1969) Pyrolysis of sulfur-containing amino acids. Agric. Biol. Chem., 33 : 1144-1151.
- Gadish I. et Zamir D.** (1986) Differential zygotic abortion in an interspecific *Lycopersicum* cross. Genome 29 : 156-159.
- Gales M.D. et Devos K.M.** (1998) Comparative genetics in the grasses. Proc. Natl. Acad. Sci. USA, 95 : 1971-1974.
- Gebhardt C., Ritter E., Barone A., Debener T., Walkemeier B., Schachtschabel U., Kaufmann H., Thompson R.D., Bonierbale M.W., Ganai M.W. et Tanksley S.D.** (1991) RFLP maps of potato and their alignment with the homologous tomato genome. Theor Appl Genet 83 : 49-57.
- Geldermann H.** (1975) Investigations on inheritance of quantitative characters in animals by gene marker. I. Methods. Theor. Appl. Genet., 46 : 319-330.
- Gianturco M.A., Giammarino A.S. et Pitcher R.G.** (1963) The structure of five cyclic diketones isolated from coffee. Tetrahedron, 19 : 2051-2059.
- Glaszmann J.C., Dufour P., Grivet L., D'Hont A., Deu M., Paulet F. et Hamon P.** (1997) Comparative genome analysis between several tropical grasses. Euphytica, 96 : 13-21.
- Grant V.** (1975) genetics of flowering plants. Columbia University Press, New-York.
- Grassias-Hubault M.** (1980) Etude de la fertilité et du comportement méiotique des hybrides interspécifiques tétraploïdes arabusta *Coffea arabica* x *C. canephora*. Thèse Doc. Univ. Paris-Sud (Centre d'Orsay) 98 pp.
- Grattapaglia D. et Sederoff R.** (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross mapping strategy and RAPD markers. Genetics 137 : 1121-1137.
- Grini P.E., Schnittger A, Schwarz H. et al.** (1999) Isolation of ethyl methanesulfonate-induced gametophytic mutants in *Arabidopsis thaliana* by a segregation distortion assay using the multimarker chromosome 1. Genetics, 151 : 849-863.
- Grodziger T., Williams J., Sharp P. et Sambrook J.** (1974) Physical mapping of temperature-sensitive mutations of Adenovirus. Cold Spring harbor Symp. Quant. Biol., 39 : 439-446.
- Guadagni D.G., Miers J.C. et Venstrum D.W.** (1969) Concentration effect on odor addition or synergism in mixtures of methyl sulfide. J. Food Sci., 34 : 630-632.

- Guadani D.G., Buttery R.G. et Harris J.** (1966) Odor intensities of hop oil components. *J. Sci. Food Agric.*, 17 : 142-144.
- Guillaumet J.-L. et Hallé F.** (1967) Etude de la variabilité du *Coffea arabica* dans son aire d'origine. Rapport de mission ORSTOM dans le Sud-Ouest de l'Ethiopie. Mult. ORSTOM Sadiopodoumlé), Abidjan, 54 pp.
- Guyot B., Christophe T. et Gueule D.** (1997) Effet inhibiteur de l'acide caféyl 5-quinique sur la formation des pyrazines dans les réactions de Maillard. *Proc. Int. Congr. ASIC 17* : 169-175.
- Guyot B., Gueule D., Manez J.C., Perriot J.J., Giron J. et Villain L.** (1996) Influence de l'altitude et de l'ombrage sur la qualité des cafés arabicas. *Plantation Recherche Développement 3* : 272-280.
- Guyot B., Petnga E. et Vincent J.C.** (1988) Analyse qualitative d'un café *Coffea canephora* var *robusta* en fonction de la maturité. Partie I. Evolution des caractéristiques physiques, chimiques et organoleptiques. *Café Cacao Thé*, 32 : 229-242.
- Guyot B., Gueule D., Assemat S., Tchana E. et Pomathios L.** (1995) Influence du mode de préparation du café vert Robusta sur sa composition chimique et ses qualités organoleptiques. *Proc. Int. Congr. ASIC 16*, 267-277.
- Hahlbrock K. et Grisebach H.** (1979) Enzymatic controls in the biosynthesis of lignin and flavonoids. *Ann. Rev. Plant Physiol.* 30 : 105-130.
- Haldane J.B.S** (1919) The combination of linkage values, and the calculation of distances between the loci of linked factors. *J. Genet.*, 8: 299-309.
- Hamon S., Anthony F. et Le Pierrès D.** (1984) La variabilité génétique des Caféiers spontanés de la section Mozambicoffea A. Chev. I. Précisions sur deux espèces affines : *Coffea pseudozanguebariae* Bridson et *C. sp. A* Bridson. *Bull. Mus. natn. Hist. nat., Paris.* 2 : 207-223.
- Hanson K.R.** (1965) Chlorogenic acid biosynthesis. Chemical synthesis and properties of the mono-*O*-cinnamoylquinic acids. *Biochemistry* 4 : 2719-2730.
- Hassan M.U.** (1970) The precursor of coffee aroma, Ph D Thesis, University of Strathclyde, Glasgow.
- Ho C.T., Hwang H.I., Yu T.-H. et Zhang J.** (1993) An overview of the Maillard reactions related to aroma generation in coffee. *ASIC 15*, Montpellier : 519-527.
- Ho C.T., Jin Q.Z., Lee K.N., Carlin J.T. et Chang S.S.** (1983) Synthesis and aroma properties of new alkyloxazoles and alkythiazoles identified in cocoa butter from roasted cocoa beans. *J. Food Sci.*, 48: 1570-1571.
- Horman I. et Viani R.** (1971) The caffeine-chlorogenate complex of coffee NMR study. *Proc. Int. Congr. ASIC 5* : 102-111.
- Hubart N.L., Huber S.C. et Pharr D.M.** (1989) Sucrose phosphate synthase and acid invertase as determinants of sucrose concentration in developing muskmelon (*Cucumis melo* L.) fruits. *Plant physiology*, 91 : 1527-1534.
- Huber S.C.** (1989) Biochemical mechanism for regulation of sucrose accumulation in leaves during photosynthesis. *Plant physiology*, 91, 656-662.
- Jin Q.Z., Hartman G. et Ho C.T.** (1984) Aroma properties of some oxazoles. *Perfum. Flavor.*, 9 : 25-29.

- Kappeler A.W., Baumann T.W. et Greutert P.D.** (1987) Complexation of purine alkaloids in the coffee plant. *Proc. Int. Congr. ASIC* 12: 247-253.
- Keim P., Diers B.W., Olson T.C. et Shoemaker R.C.** (1990) RFLP mapping in soybean: association between marker loci and variation in quantitative traits. *Genetics*, 126: 735-742.
- Keppler J.G.** (1977) Twenty-five years of flavor research in a food industry. *J. Am. Oil Chem. Soc.*, 54 : 474-477.
- Kianian S.F. et Quiros C.F.** (1992) Generation of a brassica oleracea composite RFLP map: linkage arrangements among various populations and evolutionary implications. *Theor. Appl. Genet.*, 84: 544-554.
- Klann E.M., Chetelat R.T., Bennett A.B.** (1993). Expression of acid invertase gene controls sugar composition in tomato (*Lycopersicon*) fruit. *Plant Physiol.* 103: 863-870.
- Knight R.L.** (1948) Dictionary of genetics. Waltham, Mass., Chronica Botanica Co.
- Kosambi D.D.** (1944) The estimation of map distances recombination values. *Ann Eugen* 12 : 172-175
- Kroplien U** (1971) Monosaccharides in coffees and coffee substitutes. *Proc. Int. Congr. ASIC* 5, 217-223
- Krug C.A et Carvalho A.** (1951) The genetics of coffee. *Advances in genetics*, 4: 127-168.
- Kuang H., Richardson T., Carson S., Wilcox P. et Bongarten B.** (1999) Genetic analysis of inbreeding depression in plus tree 850.55 of *Pinus radiata* D. Don. I. genetic map with distorted markers. *Theor. Appl. Genet.*, 98: 697-703.
- Ky C.-L., Barre P., Lorieux M., Trouslot P., Akaffou S., Louarn J., Charrier A., Hamon S. et Noirot M.** (2000) Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). *Theoretical and Applied Genetics* (accepté pour publication).
- Ky C.-L., Doubeau S., Guyot B., Charrier A., Hamon S., Louarn J., Noirot M.** (1999) Inheritance of sucrose content in the interspecific cross : *Coffea pseudozanguebariae* x *Coffea liberica* 'dewevrei'. *Plant Breeding* (sous-presse).
- Ky C.-L., Louarn J., Akaffou S., Chrestin H., Charrier A., Hamon S., et Noirot M** (1999) Qualitative improvement of coffee seeds by genetical and physiological approaches. *Proc. Coll. ASIC* 18 (sous-presse).
- Ky C.-L., Louarn J., Guyot B., Charrier A., Hamon S. et Noirot M.** (1999) Relations between and inheritance of chlorogenic acid content in an interspecific cross between *Coffea pseudozanguebariae* and *Coffea liberica* var 'dewevrei'. *Theoretical and Applied Genetics* 98 : 628-637.
- Ky C.-L., Noirot M. et Hamon S.** (1997) Comparison of five purification methods for chlorogenic acids in green coffee beans (*Coffea* sp.). *Journal of Agricultural and Food Chemistry* 45 : 786-790.
- Ky C-L, Hamon S, Louarn J, Akaffou S, Guyot B, Chrestin H, Charrier A et Noirot M.** (1999) The use of genetic resources for coffee cup taste improvement. *SIBAC* (sous presse)
- L'Héritier P.** (1978) Génétique – Biologie maitrise. Eds. Masson, Paris, 313 pp.
- Lander E.S. et Botstein D.** (1989) mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, 121: 185-199.

- Lander E.S., Green P., Abrahamson J., Barlow A., Daly M.J., Lincoln S.E. et Newburg L. (1987)** MAPMAKER: an interactive computer package for primary genetic linkage maps of experimental and natural populations. *Genomics* 1: 174-181
- Lashermes P., Combes M.-C., Robert J., Trouslot P., D'Hont A., Anthony F. et Charrier A. (1999)** Molecular characterisation and origin of the *Coffea arabica* L. genome. *Mol. Gen. Genet.*, 261: 259-266.
- Lashermes P., Combes M.C., Trouslot P. et Charrier A. (1997)** Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. *Theor. Appl. Genet.* 94 : 947-955.
- Lashermes P., Combes M.C., Trouslot P., Anthony F. et Charrier A. (1996 a)** Molecular analysis of the origin and genetic diversity of *Coffea arabica* L.: implications for coffee improvement. Proc. coll. Eucarpia, Meeting on tropical plants, Montpellier, 23-29.
- Lashermes P., Couturon E., Moreau N., Paillard M. et Louarn J. (1996 b)** Inheritance and genetic mapping of self-incompatibility in *Coffea canephora* Pierre. *Theor Appl Genet* 93 : 458-462.
- Lashermes P., Cros J., Combes M.C., Trouslot P., Anthony F., Hamon S. et Charrier A. (1996 c)** Inheritance and restriction fragment length polymorphism of chloroplast DNA in the genus *Coffea* L. *Theor. Appl. Genet.* 93 : 626-632.
- Le Pierrès D. (1987)** Influence des facteurs génétiques sur le contrôle de la teneur en caféine du café. Proc. coll. ASIC 12 Montreux 468-475.
- Le Pierrès D. (1995)** Etude des hybrides interspécifiques tétraploïdes de première génération entre *Coffea arabica* L. et les caféiers diploïdes. Thèse Doct. Univ. Paris XI (Orsay). 278 p.
- Le Pierrès D. et Anthony F. (1980)** Les hybrides interspécifiques hexaploïdes *Coffea arabica* x *C. canephora*. Influence du milieu et de la structure génétique sur les potentialités agronomiques. *Café Cacao Thé*, 24: 291-296.
- Le Pierrès D., Vincent J.C. et Guyot B. (1989)** Conséquences des variations biologiques de la teneur en caféine du café vert sur le choix d'une méthode d'échantillonnage pour son dosage. Proc. coll. ASIC 13 Païpa 211-217.
- Leblon G. et Rossignol J.L. (1973)** Mechanism of gene conversion in *Ascobolus immersus*. The interaction of heteroalleles in the conversion process. *Molec Gen Genet* 122: 165-182.
- Lefebvre V et Palloix A (1996)** Both epistatic and additive effects of QTLs are involved in polygenic induced resistance to disease: a case study, the interaction pepper-*Phytophthora capsici* Leonian. *Theor. Appl. Genet.*, 93: 503-511.
- Leloup V., Louvrier A. et Liardon R. (1995)** De-gradation mechanisms of chlorogenic acids during roasting. *Proc Int Congr ASIC* 16: 192-198.
- Lentner C. et Deatherage F.E. (1959)** Organic acids in coffee in relation to the degree of roast. *Food res.* 24: 483-492.
- Leroy J.F. (1980)** Evolution et taxogénèse chez les caféiers (*Coffea* L., *Psilanthus* Hook. f. et *Nostolachma* Durand). Hypothèse sur leur origine. *Comptes rendus de l'Académie des Sciences, Paris*, 291 : 593-596.

- Leroy T.** (1993) Diversité, paramètres génétiques et amélioration par sélection récurrente réciproque du caféier *Coffea canephora* P. Thèse Doct., ENSA de Rennes, 142 p.
- Leroy T.** (1997) Introduction de gènes d'intérêt agronomique dans l'espèce *Coffea canephora* Pierre par transformation par *Agrobacterium* sp. Proc. coll. ASIC Nairobi (sous-presse).
- Leroy T. et Charrier A.** (1990) Mise en place d'un programme de sélection récurrente réciproque chez *Coffea canephora* Pierre. Proc. of the Int Congr. ASIC 13: 457-465.
- Leroy T., Montagnon C., Cilas C., Yapo A., Charmetant P. et Eskes A.B.** (1997) reciprocal recurrent selection applied to *Coffea canephora* Pierre. III. Genetics gains and results of first cycle intergroup crosses. Euphytica, 95: 347-354.
- Lincoln S., Daly M. et Lander E.** (1992) Constructing genetic maps with MAPMAKER / EXP version 3.0. Whitehead Institute Technical Report, 3<sup>rd</sup> edn
- Lorieux M.** (1993) Cartographie des marqueurs moléculaires et distorsions de ségrégation : modèles mathématiques. Thèse Doct., Univ. Montpellier II, 126 p.
- Louarn J.** (1972) Introduction à l'étude génétique des Mascarocoffea. Nouvelles déterminations de leurs nombres chromosomiques. Café Cacao Thé, 16: 312-315.
- Louarn J.** (1976) Hybrides interspécifiques entre *C. canephora* Pierre et *C. eugenioides* Moore. Café Cacao Thé 20 : 33-52.
- Louarn J.** (1982) Bilan des hybridations interspécifiques entre caféiers africains diploïdes en collection en Côte-d'Ivoire. Proc Int. Congr ASIC 10: 375-383.
- Louarn J.** (1987) Possibilités d'amélioration de l'espèce cultivée *Coffea canephora* Pierre par hybridation interspécifique avec les caféiers africains diploïdes : données cytogénétiques sur les hybrides F1. Proc. coll. ASIC 12 Montreux 434-441.
- Louarn J.** (1992) La fertilité des hybrides interspécifiques et les relations génomiques entre caféiers diploïdes d'origine africaine (Genre *Coffea* L. sous-genre *Coffea*). Thèse Doct. Univ. Paris-Sud (Orsay) 200 p.
- Luro F., Lorieux M., Laigret F., Bové J.M. et Ollitrault P.** (1994) Cartographie du génome des agrumes à l'aide des marqueurs moléculaires et distorsions de ségrégation. In : Techniques et utilisations des marqueurs moléculaires, Montpellier, France, Ed. INRA, 72 : 69-82
- Macheix J.J., Fleuriet A. et Billot J.** (1990) fruit phenolics. CRC Press éd. (USA), pp 378.
- Maga J.A.** (1981) The chemistry of oxazoles and oxazolines in foods. CRC Crit. Rev. Food Technol., 14 : 295-307.
- Maier H.G.** (1987) The acids of coffee. Proc. Int. Congr. ASIC 12: 229-237.
- Maillard L.C.** (1912) Action of amino acids on sugar. Compt. Rend., 154: 66-68.
- Mazzafera P.** (1991) Phytochemistry 30 : 2309.
- Mazzafera P. et Carvalho A.** (1992) Breeding for low seed caffeine content of coffee (*Coffea* L.) by interspecific hybridization. Euphytica 59 : 55-60.
- Mc Grath J.M. et Quiros C.F.** (1992) Genetic diversity at isozyme and RFLP loci in *Brassica campestris* as related to crop type and geographical origin. Theor. Appl. Genet., 83: 783-790.
- Mialoundama F. et Paulet P.** (1975) C.R. Acad. Sci. Ser. D, 280: 1385-1387.

- Millot F.** (1974) Recherche en sélection caféière en Ouganda. Contribution à l'étude de la cacaoculture et de la caféiculture en Ouganda, Bulletin 12, Paris (France): 47-66.
- Miron D., Schaffer A.A.** (1991). Sucrose phosphate synthase, sucrose synthase, and invertase activities in developing fruit of *Lycopersicon esculentum* Mill. and the sucrose accumulating *Lycopersicon hirsutum* Humb. and Bonpl. *Plant Physiol.* 95: 623-627.
- Mondy N.I., Mobley E.O. et Geddedahl S.** (1967). *Journal of Food Science*, 32, 4, 378.
- Montagnon C., Guyot B., Cilas C. et Leroy T.** (1998) Genetic parameters of several biochemical compounds from *Coffea canephora* green coffee. *Plant Breeding* 117, 6, 576-578.
- Morigushi T., Villegas R.J.A., Kondo T. et Ko-jima M.** (1988) Isolation of 1-O-*trans-p*-coumaroyl- $\beta$ -D-glucopyranose from sweet potato roots and examination of its role in chlorogenic acid biosynthesis. *Plant Cell Physiol.* 29: 1221-1226.
- Morishita H., Iwahashi H. et Kido R.** (1989) Chlorogenic acids composition of green and roasted coffee beans. *Bull. Fac. Edu. Wakayama Univ. Nat. Sci.* 38: 33-39.
- Morton N.E.** (1955) Sequential tests for the detection of linkage. *Am. J. Hum. Genet.*, 7: 277-318.
- Nag D.K., White M.A. et Petes T.D.** (1989) Palindromic sequences in heteroduplex DNA inhibit mismatch repair in yeast. *Nature*, 340: 318-320
- Nicolas A. et Rossignol J.L.** (1983) Gene conversion: point-mutation heterozygosities lower heteroduplex formation. *Embo Journal* 2: 2265-2270
- Nienhuis J., Helentjaris T., Slocum M., Ruggero B. et Schaefer A.** (1987) Restriction fragment length polymorphisms analysis of loci associated with insect resistance in tomato. *Crop Sci.*, 27: 797-803.
- Nikaido A., Yoshimaru H., Tsumura Y., Suyama Y., Murai M. et Nagasaka K.** (1999) Segregation distortion for AFLP markers in *Cryptomeria japonica*. *Genes and genetic systems* 74: 55-59
- Noirot M.** (1990) Diversité des mises en place des structures reproductives chez *Panicum maximum*. Ed. ORSTOM, Collection Etudes et Thèse: 145 pp.
- Ohiokpehai O., Brumen G. et Clifford M.N.** (1982) The chlorogenic acids content of some peculiar green coffee beans and the implications for beverage quality. *Proc. Int. Congr. ASIC* 10: 177-185.
- Ohloff G.** (1978) Recent developments in the field of naturally-occurring aroma components, *Prog. Chem. Org. Nat. Prod.*, 35 : 431-527.
- Osborn T.C., Alexander D.C et Fobes J.F.** (1987) Identification of restriction fragment length polymorphisms linked to genes controlling soluble solids content in tomato fruits. *Theor. Appl. Genet.*, 73: 350-356.
- Ott J.** (1976) A computer program for linkage analysis of general human pedigrees. *Am. J. Hum. Genet.*, 28: 528-529.
- Paillard M., Lashermes P. et Pétiard V.** (1996) Construction of a molecular linkage map in coffee. *Theor. Appl. Genet.* 93 : 41-47.
- Paterson A.H., Damon S., Hewitt J.D., Zamir D., Robinowitch H.D., Lincoln S.E., Lander E.S. et Tanksley S.D.** (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. *Genetics* 127: 181-197

- Paterson A.H., Lander E.S., Hewitt J.D., Peterson S., Lincoln S.E. et Tanksley S.D.** (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335: 721-726.
- Paulino A.J., matiello J.B, Paulini A.E. et Braganca J.B.** (1978) Cultura do cafe conilon. Instruções tecnicas sobre a cultura de cafe no brasil, COTEC, Rio de Janeiro 5brésil): 43 pp.
- Payne R.C., Oliveira A.R. et Fairbrothers D.E.** (1973) Disc electrophoretic investigation of *Coffea arabica* and *Coffea canephora*. General protein and malate dehydrogenase of mature seeds. *Biochemical systematics*, 1 : 59-61.
- Petnga E.** (1986) Application de l'analyse multidimensionnelle à la classification qualitative du café Robusta en fonction de la maturité, Ph D Thesis, University de Montpellier, France, pp 135.
- Pittet A.O. et Hruza D.E.** (1974) Comparative study of flavor properties of thiazoles derivatives. *J. Agric. Food Chem.*, 22 : 264-269.
- Poisson J.** (1977) Aspects chimiques et biologiques de la composition du café vert. ASIC 8, Abidjan : 33-57.
- Prewo R., Guggisberg A., Lorenzi-Riatsch A., Baumann T.W. et Wettstein-Bättig M.** (1990) Crystal structure of mazambioside, a diterpene glycoside of *Coffea pseudozanguebariae*. *Phytochemistry*, 29 (3): 990-992.
- Qi X., Stam P et Lindhout P.** (1998) Use of locus-specific AFLP markers to construct a high density molecular map in barley. *Theoretical and Applied genetics*, 96 : 376-384.
- Quijano-Rico M. et Spettel B.** (1973) Analisis instrumental por activation con neutrone termicos de algunas muestras de café. ASIC 6, Paris : 102-105.
- Quillet M.C., Madjidian N., Griveau Y., Serieys, Tersac M.,Lorieux M. et Bervillé A.** (1995) Mapping genetic factors controlling pollen viability in an interspecific cross in *Helianthus* sect. *Helianthus*. *Theor. Appl. Genet.* 91 : 1195-1202.
- Radman M. et Wagner R.** (1993) Mismatch recognition in chromosomal interactions and speciation. *Chromosoma*, 102 : 369-373
- Rakotomalala J.J.** (1993) Diversité biochimique des caféiers. Thèse Doct., TDM, ORSTOM (eds), 216 p.
- Rakotomalala J.J., Cros J., Charrier A., Anthony F. et Noirot M.** (1993) Marqueurs biochimiques de la diversité des caféiers. Proc. coll. ASIC 15 Montpellier 47-55.
- Reineccius G.A.** (1995) The Maillard reaction and coffee flavor. ASIC 16, Kyoto, 249-257.
- Resnick M.A., Skaanild M. et Nilsson-Tillgren T.** (1989) Lack of DNA homology in a pair of divergent chromosomes greatly sensitizes them to loss by DNA damage. *Proc. Natl. Acad. Sci. U.S.A.* 86 : 2276-2280
- Rick C.M.** (1963) Differential zygotic lethality in a tomato species hybrid. *Genetics.* 48 : 1497-1507.
- Rick C.M.** (1966) Abortion of male and female gametes in the tomato determined by allelic interaction. *Genetics.* 53 : 85-96.
- Rick C.M.** (1969) Controlled introgression of chromosomes of *Solanum pennellii* into *Lycopersicon esculentum*: segregation and recombination. *Genetics.* 62 : 753-768.
- Ritter E., Gebhardt C. et Salamini F.** (1990) Estimation of recombination frequencies and construction of RFLP linkage maps in plant from crosses between heterozygous parents. *Genetics*, 125: 645-654.

- Ross H.A. et Davies H.V.** (1992) Sucrose metabolism in tubers of potato (*Solanum tuberosum* L.) *Plant Physiology*, 98 : 287-293.
- Ryan, T.A.** (1960) Significance tests for multiple comparison of proportions, variances and other statistics. *Psychological Bulletin* 57, 318-328.
- Sakaguchi M. et Shibamoto T.** (1978) Formation of sulfur-containing compounds from the reaction of D-glucose and hydrogen sulfide. *J. Agric. Food Chem.*, 26 : 1260-1262.
- Savitri Kumar N., Hewavitharanage P. et Adikaram N.K.B.** (1995) Attack on the tea by *Xyleborus fornicatus* : inhibition of the symbiote, *Manocrosporium ambrosium*, by caffeine. *Phytochemistry*, 40, 4 : 1113-1116.
- Schulthess B.H. et Baumann T.W.** (1995) Are xanthosine and 7-methylxanthosine caffeine precursors ? *Phytochemistry*, 39, 6 : 1363-1370.
- Schulthess B.H., Morath P. et Baumann T.W.** (1995) Caffeine biosynthesis starts with the metabolically channelled formation of 7-methyl-XMP, a new hypothesis. *Phytochemistry*, 41, 1 : 169-175.
- Shadaksharaswamy M. et Ramachandra G.** (1968) Changes in the oligosaccharides and the alpha galactosidase of coffee seeds during soaking and germination. *Phytochemistry*, 7 : 715-719.
- Shibamoto T.** (1980) Heterocyclic compounds found in cooked meat. *J. Agric. Food Chem.*, 27 : 237-243.
- Shibamoto T.** (1991) An overview of coffee aroma and flavor chemistry. ASIC 14, San Francisco : 107-116.
- Shibamoto T. et Russell G.F.** (1977) Study of the volatiles isolated from a D-glucose-hydrogen sulfide-ammonia model system. *J. Agric. Food Chem.*, 25 : 109-113.
- Shimizu Y., Matsuto S., Mizunuma Y. et Okada I.** (1970) Studies on the flavor of roast barley (mugi-cha). Separation and identification of 5-hydroxymaltol, maltol, 5-methylcyclopent-2-en-2-ol-1-one and other compounds. *Agric. Biol. Chem.*, 34 : 845-853.
- Smith A.W.** (1985) Introduction in coffee. I: Chemistry, ed. Clarke R.J. et Macrae R., Elsevier Applied Science (London): 1-41.
- Snoeck J.** (1968) la rénovation de la caféiculture malgache à partir de clones sélectionnés. *Café Cacao Thé*, 3 : 223-235.
- Sondheimer E.** (1964) Chlorogenic acids and related depsides. *Botanical reviews*, 30: 263-266.
- Speer K., Sehat N. et Montag A.** (1993) Fatty acids in coffee. *Proc. of the Int Congr. ASIC 15*: 583-592.
- Stam P. et Zeven A.C.** (1981) The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing. *Euphytica*, 30: 227-238.
- Stebbins G.L.** (1958) The inviability, weakness, and sterility of interspecific hybrids. *Advances in genetics*. 9 : 147-215.
- Stennert A. et Maier H.G.** (1993) Trigonellin in Bohnenkaffee. I. Vergleich der Dünnschicht- mit der Hochleistungsflüssigchromatographie. Gleichzeitige von Coffein. *Z. Lebensm. Unters Forsch.*, 196: 430-434.
- Stennert A. et Maier H.G.** (1994) Trigonelline in coffee. II. Content of green, roasted and instant coffee. *Z. Lebensm. Unters Forsch.*, 199: 198-200.

- Stöckigt J. et Zenk H.** (1974) Enzymatic synthesis of chlorogenic acid from caffeoyl coenzyme A and quinic acid. *FEBS Letters*, 42: 131-134.
- Stoll M., Winter M., Gautschi F, Flament I et Willhalm B** (1967) The coffee aroma. *Helv. Chem. Acta.*, 50 : 628-694.
- Stommel J.R.** (1991). Enzymic components of sucrose accumulation in the wild tomato species *Lycopersicon peruvianum*. *Plant Physiol.* 99 : 324-328.
- Stommel J.R., Haynes K.G.** (1993). Genetic control of fruit sugar accumulation in a *Lycopersicon esculentum* x *L. hirsutum* cross. *J. Amer. Hort. Sci.* 118 (6) : 859-863.
- Streuli H.** (1970) Alkaloidhaltige genussmittel, ge-würze, kochsalg. In: *Kaffee*. Schormüller (ed), Springer-Verlag, Berlin, 55p.
- Stuber C.W., Lincoln S.E., Wolff D.W., Helentjaris T. et Lander E.S.** (1992) Identifiocation of genetic factors contributing to heterosis in a hybrid from two elite maize onbred lines using molecular markers. *Genetics*, 132 : 823-839.
- Sun J., Loboda T., Sung S.-J. S. et Black C.C.** (1992) Sucrose synthase in wild tomato, *Lycopersicon chmielewskii*, and tomato fruit sink strenght. *Plant physiology*, 98 : 1163-1169.
- Suyama K. et Adachi S.** (1980) Origin of alkyl-substituted pyridines in foof flavor : formation of the pyridines from the reaction of alkanals with amino acids. *J. Agric. Food Chem.*, 28 : 546-549.
- Tanaka M. et Kojima M.** (1991) Purification and characterization of *p*-coumaryl-D-glucose hydroxy-lase of sweet potato (*Ipomoea batatas*) roots. *Arch. Biochem. Biophys.* 284: 151-157.
- Tanksley S.D.** (1983) Molecular markers in plants breeding. *Pl. Mol. Biol. Rep.*, 1 : 3-8.
- Tanksley S.D.** (1993) Mapping polygenes. *Annu. Rev. Genet.* 27 : 205-233.
- Tanksley S.D. et hewitt J.** (1988) use of molecular markers in breeding for soluble solids content in tomato -- a re-examination. *Theor. Appl. Genet.* 75, 811-823.
- Tanksley S.D. et Loaiza-Figueroa F.** (1985) Gametic self-incompatibility is controlled by a single majot locus on chromososome 1 in *Lycopersicon peruvianum*. *Proc. Natl. Acad. Sci. USA*, 82 : 5093-5096.
- Tanksley S.D., Bernatzky R., Lapitan N et Prince J.P.** (1988) Conservation of gene repertoire but not gene order in peper and tomato. *Proc. Natl. Acad. Sci. USA*, 85: 6419-6424.
- Tanksley S.D., Medina-Filho H. et Rick C.M.** (1982) Use of naturally-occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. *Heredity*, 49: 11-25.
- Tautz D.** (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research*, 17 : 6463-6471.
- Taylor A.O.** (1968) *Phytochemistry*, 7: 63-71.
- Thaler H.** (1975) Makromolekulare Strukturen in Kaffee. *ASIC 7, Paris* : 175-188.
- Thaler H. et Arneith W.** (1967) Hochpolymere Kolhenhydrate des rohen end gerösteten Kaffees. *Coll. Int. Chimie des Cafés* 3, 127-136.
- Thoday J.M.** (1961) Location of polygene. *Nature*, 191: 368-370.
- Thoquet P.** (1995) Cartographie des gènes de résistance à *Pseudomonas solanacearum* chez la tomate. Thèse,

- Université d'Orsay, 132 pp.
- Tramontano WA, Hartnett CM, Lynn DG, Evans L** (1982) *Phytochemistry* 21, 1201.
- Tressl R, Holzer M. et Kamperschoer H.** (1982) Bildung von Aromastoffen in Röstkaffee in Abhängigkeit vom Gehalt an Aminosäuren und Reduzierenden Zuckern. *Proc. Int. Congr. ASIC* 10, 279-292
- Tressl R.** (1989) Formation of flavor components in roasted coffee. Thermal generation of aromas. Ed. Parliment T.H., Mc Gorrin R.J. et Ho C., American Chemical Society, Whashington DC: 285-301.
- Tressl R., Grünewald K.G., Köppler et Silwar R.** (1978) Phenols in roasted coffees of different varieties. *Z. Lebensm. Unters. Forsch.*, 167 : 108-110.
- Trugo L.C.** (1985) Carbohydrates. Dans : Coffee ; Eds. : Clarke R.J. et Macrae R. ; Publ. : Elsevier Applied Science ; Volume I, Chemistry : 83-113.
- Trugo L.C. et Macrae R.** (1984) A study of the effect of roasting on the chlorogenic acid composition of coffee using HPLC. *Food Chem.* 15: 219-227.
- Trugo L.C. et Macrae R.** (1984) Chlorogenic acid composition of instant coffees. *Analyst* 109, 263-266.
- Ulbrich B. et Zenk M.H.** (1979) Partial purification and properties of hydroxycinnamoyl-CoA: quinate hydroxycinnamoyl transferase from higher plants. *Phytochemistry* 18 : 929-933.
- Viani R** (1985) Physiological active substances in coffee. Dans : Coffee ; Eds. : Clarke R.J. et Macrae R. ; Publ. : Elsevier Applied Science ; Volume I, Chemistry : 1-31
- Viani R. et Horman I.** (1974) Thermal behavior of trigonelline. *J. Food Sci.*, 39 : 1216-1217.
- Viani R. et Horman I.** (1975) Determination of trigonelline in coffee. *Proc. of the Int Congr. ASIC* 7 : 273-278.
- Villegaas R.J.A. et Kojima M.** (1986) Purification and characterization of hydroxycinnamoyl D-glucose. *J. Biol. Chem.* 261 : 8729-8733.
- Virk P.S., Ford-Lloyd B.V. et Newbury H.J.** (1998) Mapping AFLP markers associated with subspecific differentiation of *Oryza sativa* (rice) and an investigation of segregation distortion. *Heredity* 81 : 613-620.
- Vitzthum O.G., Barthels M. et Kwasny K.** (1974) Rapid gas-chromatographic determination of caffeine in caffeine-containing and decaffeinated coffees with the nitrogen-sensitive detector. *Z. Lebensm. Unters. Forsch.* 154 : 135-140.
- Voilley A., Sauvageot F. et Durand D.** (1977) *ASIC* 8 : 251-259.
- Vos P., Hogers R., Bleeker M., Reijans M., Van de Lee T., Hornes M., Frijters A., Pot J., Peleman J., Kuiper M. et Zabeau M.** (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 23 : 4407-4414.
- Wainwright T., McMahon J.F. et Mc Dowell J.** (1972) Formation of methional and methanethiol from methionine. *J. Sci. Food Agric.*, 23 : 911-914.
- Waite MJ, Reynolds SB et Friend J.** (1978) The metabolism of chlorogenic acid in tuber discs of a resistant and a susceptible potato cultivar after inoculation with *Fusarium solani* var. *caeruleum*. *Biochemical Society Transactions*, 6 : 441-442.
- Welsh J et Mc Clelland M.** (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research*, 18 : 7213-7218.

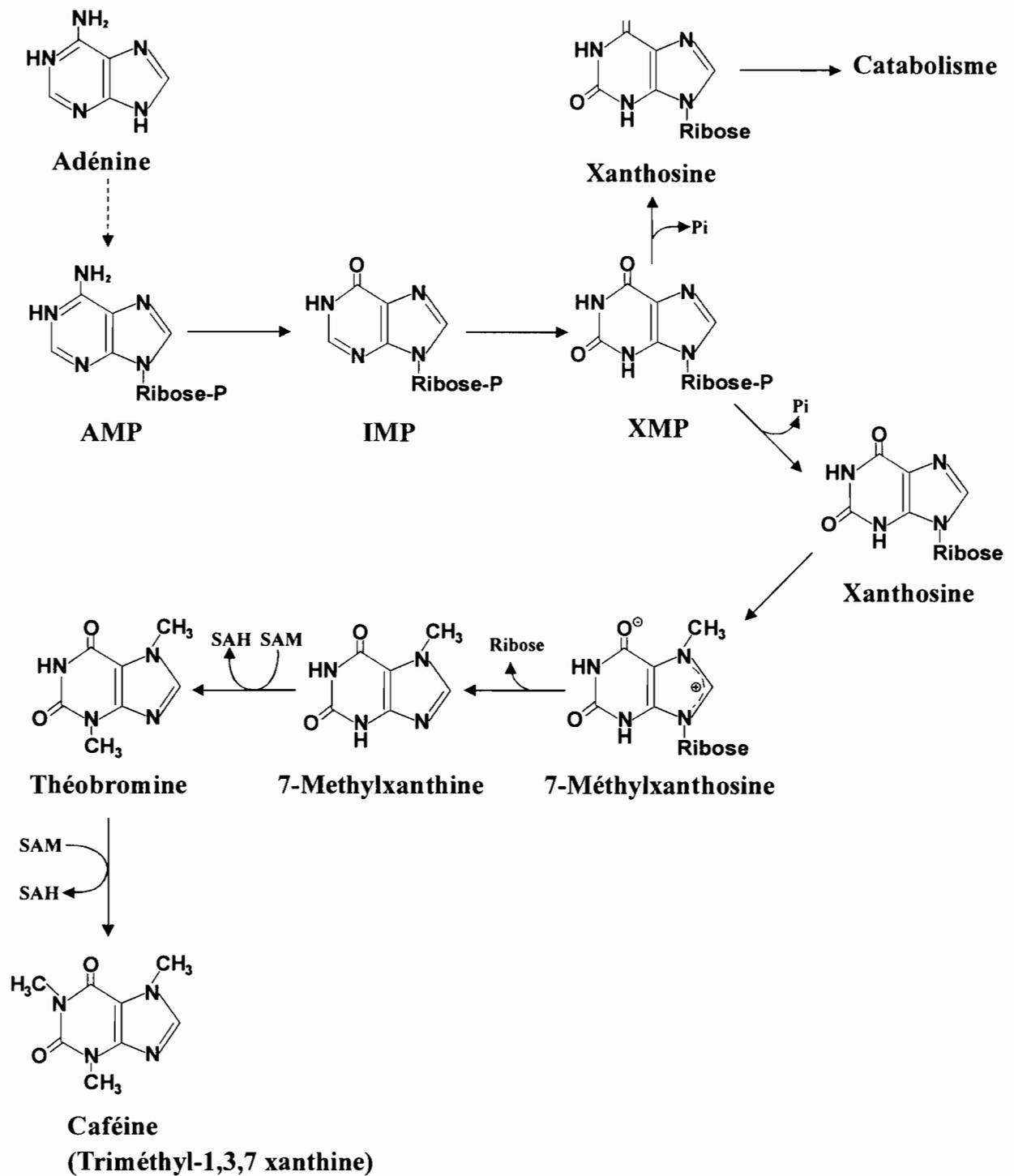
- Williams J.G.K., Kubelik A.R., Livak K.J., Rafalski J.A. et Tingey S.V.** (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, 18 : 6531-6535.
- Wilson W.A., Harrington S.E., Woodman W.L., Lee M., Sorrells M.E. et McCouch S.R.** (1999) Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated panicoids. *Genetics Society of America*, 153 : 453-473.
- Xu Y., Zhu L., Xiao J., Huang N. et McCouch S.R.** (1997) Chromosomal regions associated with segregation distortion of molecular markers in F<sub>2</sub>, backcross, double haploids, and recombinant inbred populations in rice (*Oryza sativa* L.). *Mol Gen Genet*, 253 : 535-545.
- Yapo A., Charmetant P., Leroy T., le Pierrès D. et Berthaud J.** (1990) Les hybrides congusta (*Coffea canephora* x *C. congensis*): comportement de rétrocroisements sur *C. canephora* dans les conditions de Côte d'Ivoire. *Proc. of the Int Congr. ASIC 13* : 448-456.
- Yapo A., Leroy T. et Louarn J.** (1992) Contribution à l'amélioration de *Coffea canephora* par hybridation interspécifique avec *Coffea liberica*. *Proc. of the Int Congr. ASIC 14* : 403-411.
- Zabeau M. et Vos P.** (1993) Selective restriction fragment amplification: a general method for DNA fingerprinting. European Patent Application number: 92402629.7, Publication number 0 534 858 A1.
- Zamir D. et Tadmor Y.** (1986) Unequal segregation of nuclear genes in plants. *Bot Gazette* 147: 355-358
- Zamir D., Tanksley S.D. et Jones R.A.** (1982) Haploid selection for low temperature tolerance of tomato pollen. *Genetics*, 101: 129-137.
- Zehr B.E., Bublely J.W., Chojeki J., Maroof M.A.S. et Mowers R.P.** (1992) Use of RFLP markers to search for alleles in maize population for improvement of an elite hybrid. *Theor. Appl. Genet.* 83, 903-911.
- Zuluaga J.V., Valencia G.A. et Gonzales J.** (1971) Contribucion al estudio de la naturaleza de la resistencia del cafeto a *Ceratocystis fimbriata* (Ell. Halst.) Hunt. *Cenicafé*, 22 : 43-70.

**Annexes : tableaux et figures**

---

**Annexe 2.1. : Principales caractéristiques agronomiques des caféiers (Le Pierrès, 1995 ; Louarn, 1992).**

	Espèce	Caractéristique générale
<b>Afrique de l'ouest</b>	<i>C. canephora</i>	Cultivée en basse altitude (environ 30 % de la production mondiale), autostérile, multipliée par bouture ou par graines, vigoureuse, architecture multicaule nécessitant le recépage périodique, très rustique, productive, faible granulométrie, tolérant aux maladies.
	<i>C. congensis</i>	Spontanée, peu vigoureuse, autostérile, adaptée aux zones inondables sous fortes pluviométrie, maintenue en collection par greffage, architecture de type adapté à la taille, résistance aux souches locales de rouilles orangées, floraisons nombreuses tout au long de l'année, taille des graines très variable.
	<i>C. heterocalyx</i>	Espèce sauvage, autofertile, naine, floraisons décalées par rapport aux autres espèces par une durée plus longue d'épanouissement des fleurs après la pluie inductrice, gros grains.
	<i>C. humilis</i>	Espèce sauvage, autostérile, port arbustif très peu développé, ramifications secondaires sur les plagiotropes quasi inexistantes, grands grains.
	<i>C. liberica</i>	Très peu cultivée (sensibilité aux maladies), autostérile, multipliée par graines, grande vigueur rendant la récolte difficile, architecture unicaule ramifiée ne nécessitant pas le recépage, très productive, cycle long de fructification (10 à 11 mois), granulométrie importante (les cerises peuvent atteindre jusqu'à 2 cm de diamètre).
	<i>C. stenophylla</i>	Espèce sauvage, autostérile, architecture unicaule ramifiée, petits grains, vigoureuse et rustique en condition de sécheresse, résistante aux mineuses des feuilles, fruits noirs.
<b>Afrique de l'est</b>	<i>C. arabica</i>	Cultivée en altitude, multipliée par graines, autofertile, bon rendement, sensibilité aux maladies, architecture adaptée à la taille.
	<i>C. eugenioides</i>	Espèce sauvage peu vigoureuse en altitude, autostérile, petits arbustes buissonnants, peu productive, grains très petits.
	<i>C. pseudozanguebariae</i>	Espèce sauvage, autostérile, petits arbustes buissonnants, peu productive, cycle court de fructification (2,5 à 3,5 mois), grains très petits, fruits noirs.
	<i>C. racemosa</i>	Espèce sauvage, autostérile, court cycle de fructification (moins de 11 semaines).
	<i>C. salvatrix</i>	Espèce sauvage, autostérile.
	<i>C. sessiliflora</i>	Espèce sauvage, autostérile, court cycle de fructification (moins de 11 semaines).



**Annexe 2.2.** : Chaîne de biosynthèse conduisant à la formation de la caféine. Codes des abréviations biochimiques : Adénosine mono-phosphate (AMP) ; Inosine mono-phosphate (IMP) ; Xanthosine mono-phosphate (XMP) ; S-adénosyl-méthionine, donneur de groupe méthyl (SAM) ; S-adénosyl-homocystéine (SAH) (Schulthess *et al.*, 1995 ; Schulthess et Baumann, 1995).

### **Annexe 3.1. : Solutions courantes**

- **Bleu** : 30 % glycérol, 0,25 % bleu de bromophénol, 0,1 M EDTA, pH 8
- **Bromure d'éthidium (BET)** : 2,5 mg/l dans de l'eau Millipore
- **RNase A** : dissoudre la RNase A à une concentration de 10 mg/ml dans une solution d'acétate de sodium 0,01 M pH 5,2. Chauffer cette solution pendant 15 mn à 100°C. Laisser refroidir à température ambiante. Ajuster le pH en ajoutant 0,1 volume de Tris HCl 1 M pH 7,4. Aliquoter et conserver à -20°C.
- **TAE 50X** : 2 M Tris, 0,05 M EDTA, 5,71 % d'acide acétique glacial pH 8.
- **TBE 10X** : 0,9 M Tris, 0,89 M acide borique, 0,02 M EDTA pH 8.
- **TE** : 0,01 M Tris, 0,001 M EDTA pH 8.

**Annexe 3.2.** : Origine des 7 géotypes PSE (Kenya) et des 13 géotypes DEW (République Centrafricaine) retenus pour l'étude de génétique quantitative (Chapitre 5).

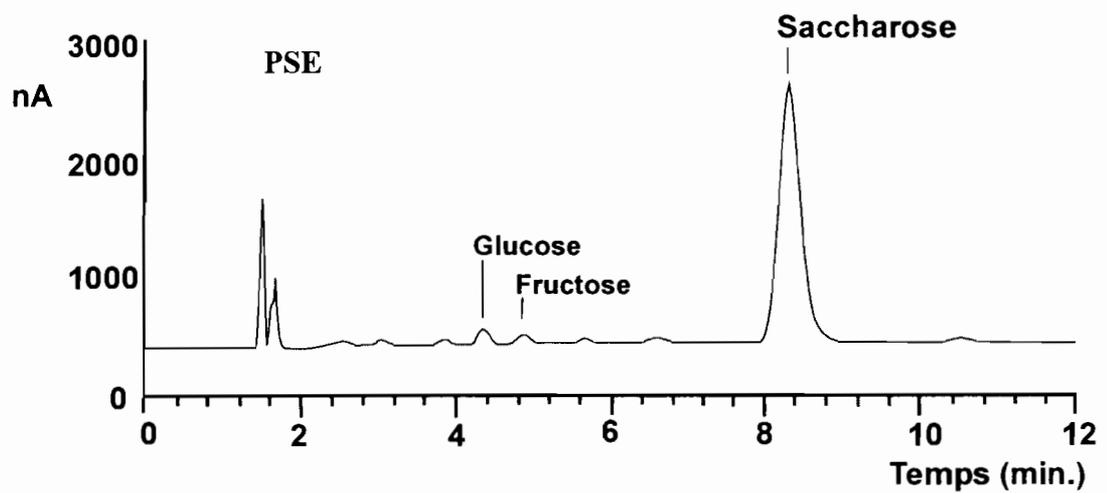
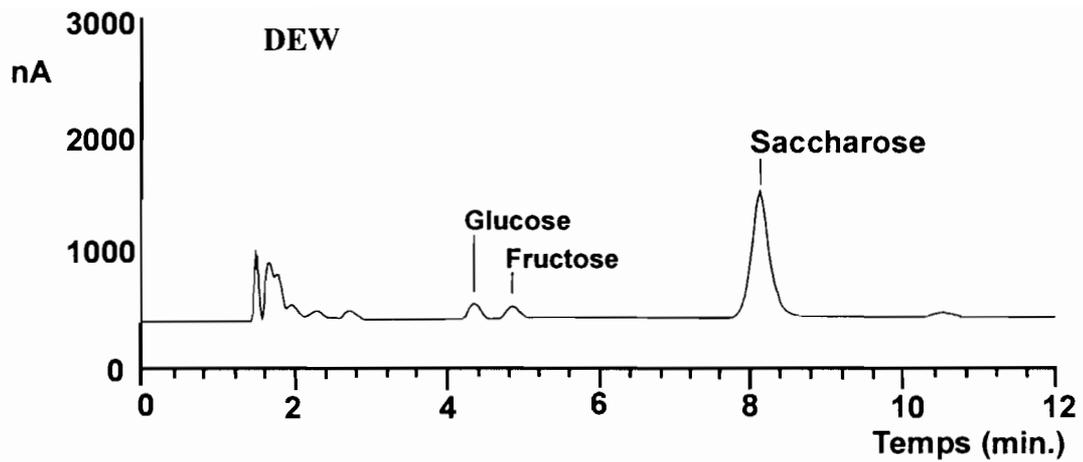
Espèce	Géotype	Localité
PSE	8013	Shimba
	8038	Shimba
	8044	Diani
	8053	Diani
	8056	Diani
	8066	Shimoni
	8107	Rabaï
DEW	5303	Sobou
	5307	Bimbo
	5341	Pama 1
	5372	Bozo
	5390	Oumba
	5462	Karagda
	5495	Ouango
	5599	Oulo
	5751	Zime
	5801	Limbo
	5832	Balifondo
	5850	Siolo-Nvuna
	5914	Bangassou

### Annexe 3.3. : Présentation des 55 variables utilisées pour la recherche des QTLs.

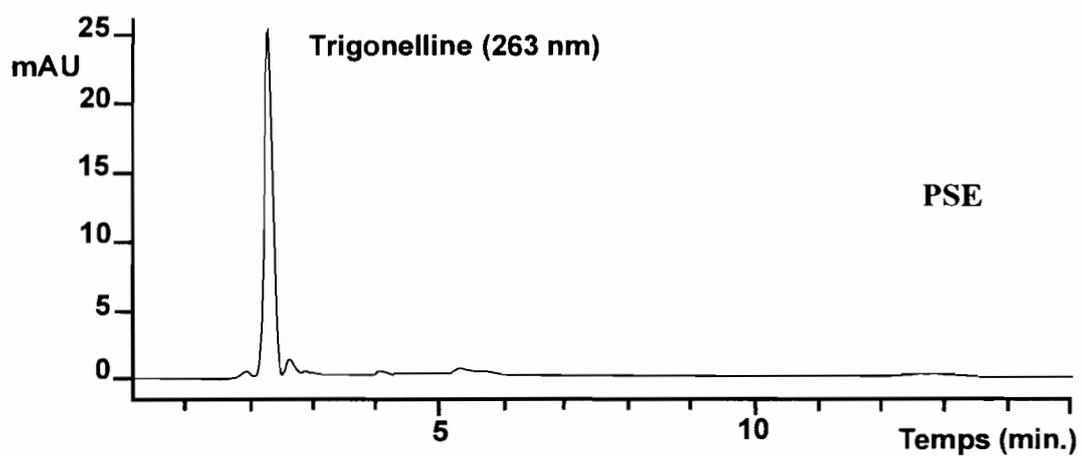
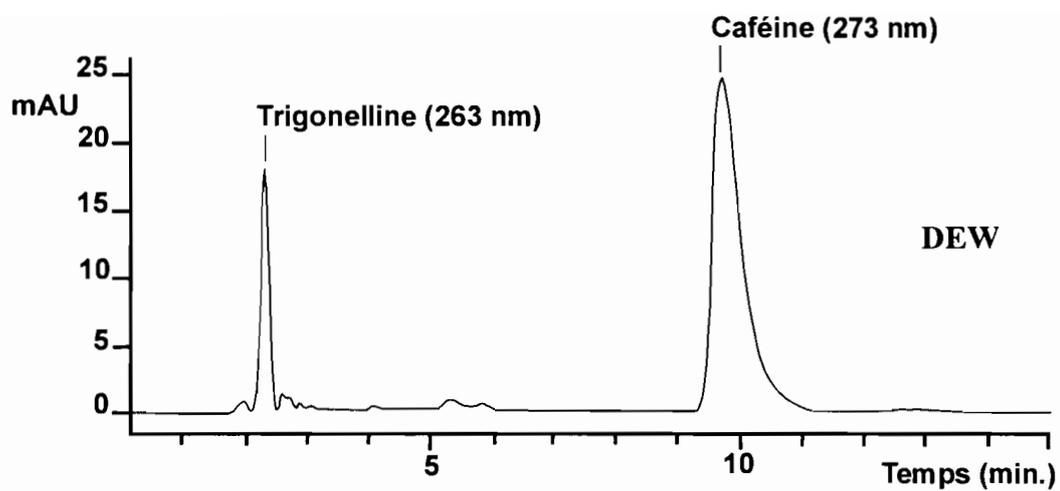
Variable	Définition
DCF	Durée du cycle de fructification
P100	Poids de cent grains
P100/ J	Rapport du poids de cent grains sur la durée du cycle de fructification
3-CQA	Teneur en acide caféyl-3 quinique
4-&5-CQA	Teneur en acide caféyl-4 quinique et en en acide caféyl-5 quinique
3-FQA	Teneur en acide férulyl-3 quinique
4-FQA	Teneur en acide férulyl-4 quinique
5-FQA	Teneur en acide férulyl-5 quinique
3,4-DQA	Teneur en acide dicaféyl-3,4 quinique
3,5-DQA	Teneur en acide dicaféyl-3,5 quinique
4,5-DQA	Teneur en acide dicaféyl-4,5 quinique
CQAt	Teneur totale en acides caféylquiniques (3-CQA + 4-&5-CQA)
FQAt	Teneur totale en acides férulylquiniques (3-FQA + 4-FQA + 5-FQA)
DQAt	Teneur totale en acides dicaféylquiniques (3,4-DQA + 3,5-DQA + 4,5-DQA)
CGA1	Teneur totale des isomères possédant l'unité acide caféique (CQAt + DQAt)
CGA	Teneur totale en acides chlorogéniques (CQAt + FQAt + DQAt)
% 3-CQA	Pourcentage en acide caféyl-3 quinique (3-CQA/ CGA)
% 3-FQA	Pourcentage en acide férulyl-3 quinique (3-FQA/ CGA)
% 4-FQA	Pourcentage en acide férulyl-4 quinique (4-FQA/ CGA)
% 5-FQA	Pourcentage en acide férulyl-5 quinique (5-FQA/ CGA)
% 3,4-DQA	Pourcentage en acide dicaféyl-3,4 quinique (3,4-DQA/ CGA)
% 3,5-DQA	Pourcentage en acide dicaféyl-3,5 quinique (3,5-DQA/ CGA)
% 4,5-DQA	Pourcentage en acide dicaféyl-4,5 quinique (4,5-DQA/ CGA)
% CQAt	Pourcentage en acides caféylquiniques (CQAt/ CGA)
% FQAt	Pourcentage en acides férulylquiniques (FQAt/ CGA)
% DQAt	Pourcentage en acides dicaféylquiniques (DQAt/ CGA)
% CGA1	Pourcentage des isomères possédant l'unité acide caféique
3-CQA/J	Rapport de la teneur en acide caféyl-3 quinique sur la durée du cycle de fructification
4-&5-CQA/J	Rapport de la teneur en acide caféyl-4 quinique et en en acide caféyl-5 quinique sur la durée du cycle de fructification
3-FQA/J	Rapport de la teneur en acide férulyl-3 quinique sur la durée du cycle de fructification
4-FQA/J	Rapport de la teneur en acide férulyl-4 quinique sur la durée du cycle de fructification
5-FQA/J	Rapport de la teneur en acide férulyl-5 quinique sur la durée du cycle de fructification
3,4-DQA/J	Rapport de la teneur en acide dicaféyl-3,4 quinique sur la durée du cycle de fructification
3,5-DQA/J	Rapport de la teneur en acide dicaféyl-3,5 quinique sur la durée du cycle de fructification
4,5-DQA/J	Rapport de la teneur en acide dicaféyl-4,5 quinique sur la durée du cycle de fructification
CQAt/J	Rapport de la teneur totale en acides caféylquiniques sur la durée du cycle de fructification

FQA/J	Rapport de la teneur totale en acides férulylquiniques sur la durée du cycle de fructification
DQA/J	Rapport de la teneur totale en acides dicaféylquiniques sur la durée du cycle de fructification
CGA1/J	Rapport de la teneur totale des isomères possédant l'unité acide caféique sur la durée du cycle de fructification
CGA/J	Rapport de la teneur totale en acides chlorogéniques sur la durée du cycle de fructification
SAC	Teneur en saccharose
SAC/J	Rapport de la teneur en saccharose sur la durée du cycle de fructification
TRI	Teneur en trigonelline
TRI/J	Rapport de la teneur en trigonelline sur la durée du cycle de fructification
HET	Teneur en hétéroside
HET/J	Rapport de la teneur en hétéroside sur la durée du cycle de fructification
CAF	Teneur en caféine
CAF/J	Rapport de la teneur en caféine sur la durée du cycle de fructification
CAF/QCA	Rapport de la teneur en caféine sur la teneur en acides caféylquiniques
CAF/CGA1	Rapport de la teneur en caféine sur la teneur des isomères possédant l'unité acide caféique
CREM	Coefficient de remplissage des fruits. C'est le rapport du nombre de graines au nombre de fruits. C'est un paramètre de fertilité synthétique qui ne tient pas compte des causes de stérilité. Chez les caféiers exprimant une fertilité quasi complète, il est proche de 2.
FECA	Taux de fruits à une écaille. C'est le rapport du nombre d'écailles au nombre de fruits. Il est tributaire de bonnes conditions de pollinisation ; dans ce cas il fournit une représentation globale de la stérilité des mégaspores et des sacs embryonnaires (stérilité gamétique), de la stérilité des ovules fécondés et des jeunes albumens (stérilité zygotique précoce).
TVID	Taux de loge vide des grains. C'est le rapport du nombre de loges vides au nombre de loges total. Il donne une représentation de l'avortement tardif des albumens (stérilité zygotique tardif).
TCAR	Taux de graine « caracoli ». C'est le rapport du nombre de graines ovoïdes issues de fruits uniloculés, au nombre total de graines.
VP	Viabilité pollinique. C'est le rapport du nombre de grain de pollen viable (coloration du cytoplasme de façon dense et homogène par le carmin acétique) sur 100 grains de pollen issus de 3 fleurs récoltées au stade « chandelle » par génotype pour une floraison.

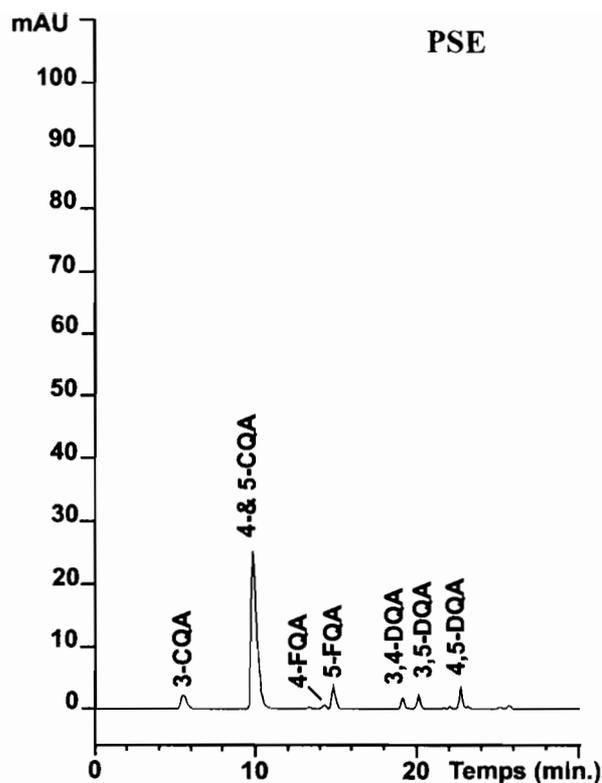
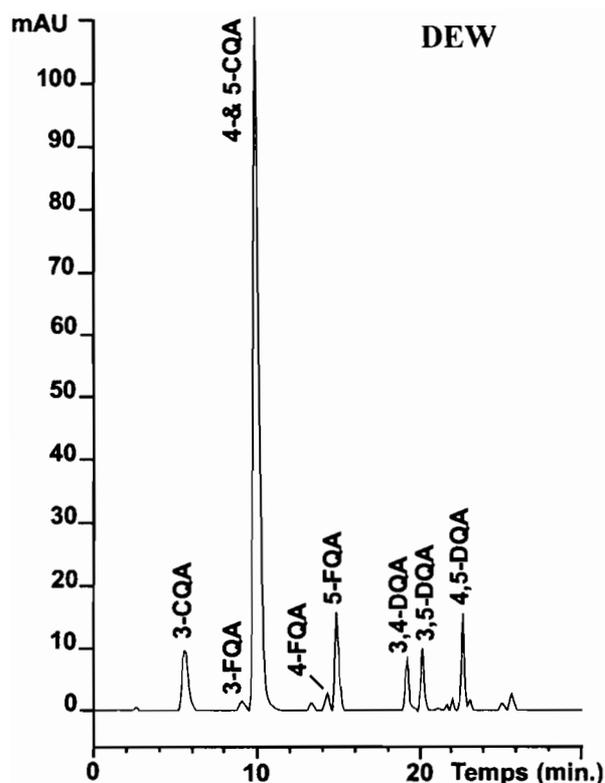
---



**Annexe 5.1.** : Profils d'élution du saccharose à partir d'extraits de grains verts à maturité des espèces *C. liberica* var. 'dewevrei' (DEW) et *C. pseudozanguebariae* (PSE) (obtenus par HPLC couplé à une détection ampérométrique, voir chapitre 3). Les chromatogramme révèlent la présence de glucose et de fructose (monosaccharides mineurs). Le symbole « nA » correspond à  $10^{-9}$  ampères.



**Annexe 5.2.** : Profils de co-élution de la trigonelline et de la caféine d'extraits de grains verts à maturité des espèces : *C. liberica* var. 'dewevrei' (DEW) et *C. pseudozanguebariae* (PSE) obtenus par HPLC (conditions chromatographiques détaillées dans le chapitre 3). Le symbole « mAU » correspond à  $10^{-3}$  unité d'absorbance.



**Annexe 5.3.** : Profils d'élution obtenus par HPLC des acides chlorogéniques d' extraits de grains verts à maturité de l'espèce *C. liberica* var. 'dewevrei' (DEW) et de l'espèce *C. pseudozanguabariae* (PSE) (conditions chromatographiques sont détaillées dans le chapitre 3). Les codes pour les isomères sont : acide caféyl-3 quinique (3-CQA), acide férulyl-3 quinique (3-FQA), acide férulyl-4 quinique (4-FQA) ; acide férulyl-5 quinique (5-FQA) ; acide dicaféyl-3,4 quinique (3,4-DQA), acide dicaféyl-3,5 quinique (3,5-DQA) et acide dicaféyl-4,5 quinique (4,5-DQA). Le symbole « mAU » correspond à  $10^{-3}$  unité d'absorbance.

**Annexe 5.4.** : Différences inter et intra-groupe (DEW, BCDEW, F1, BCPSE et PSE) révélées par les tests F et les contributions factorielles (chiffre entre parenthèse). Niveau de signification :\*\*\* très hautement significatif ( $p < 0,001$ ).

	Différences inter-groupe		Différences intra-groupes	
		F <sub>4, 30</sub>		F <sub>30, 70</sub>
CGA	114 ***	(93,9%)	41,2 ***	(5,7%)
CQA	103 ***	(93,3%)	53,9 ***	(6,3%)
DQA	54,2 ***	(87,8%)	32,5 ***	(11,2%)
FQA	6,11 ***	(41,9%)	178 ***	(57,1%)
CQA/CGA	25,8 ***	(77,7%)	126 ***	(21,7%)
DQA/CGA	17,2 ***	(69,4%)	127 ***	(29,9%)
FQA/CGA	1,95 NS	(-)	345 ***	(99,22%)
4-&5-CQA	80,3 ***	(91,7%)	63,7 ***	(8,0%)
3-CQA	158 ***	(95,5%)	38,8 ***	(4,1%)
4-&5-CQA/CQA	22,9 ***	(75,3%)	73,2 ***	(23,7%)
4,5-DQA	45,5 ***	(85,4%)	22,0 ***	(12,8%)
3,5-DQA	27,3 ***	(78,3%)	42,1 ***	(20,3%)
3,4-DQA	60,6 ***	(88,6%)	20,2 ***	(9,8%)
4,5-DQA/DQA	12,6 ***	(52,8%)	4,15 ***	(24,2%)
3,5-DQA/DQA	14,5 ***	(60,2%)	7,18 ***	(26,8%)
3,4-DQA/DQA	4,09 **	(28,3%)	7,02 ***	(60,2%)
5-FQA	5,14 **	(37,0%)	271 ***	(62,3%)
4-FQA	11,5 ***	(67,3%)	19,6 ***	(32,7%)
3-FQA	18,7 ***	(71,3%)	89,6 ***	(27,7%)
5-FQA/FQA	5,27 **	(37,2%)	70,0 ***	(60,2%)
4-FQA/FQA	5,49 **	(38,4%)	72,5 ***	(59,1%)
3-FQA/FQA	42,5 ***	(84,4%)	20,6 ***	(13,5%)

**Annexe 5.5.** : Teneur en saccharose, en trigonelline et en acides chlorogéniques chez : 13 génotypes DEW' et 7 génotypes PSE (% MS). Les génotypes en gras de DEW correspondent à ceux tirés aléatoirement pour les analyses statistiques. Les codes des composés sont ceux présents à l'annexe 3.3.

Génotype	SAC	TRI	CGA	CQAt	DQAt	FQAt	CQA3	CQA45	DQA34	DQA35	DQA45	FQA3	FQA4	FQA5
<b>5303</b>	6,24	0,51	<b>6,86</b>	5,71	0,977	0,176	0,436	5,27	0,221	0,249	0,507	0,011	0,014	0,151
<b>5307</b>	3,76	0,45	<b>6,83</b>	5,55	1,015	0,266	0,585	4,97	0,246	0,286	0,483	0,019	0,030	0,217
5341	5,49	0,80	7,47	5,38	1,233	0,865	0,494	4,88	0,287	0,325	0,621	0,066	0,124	0,676
<b>5372</b>	5,37	0,61	7,03	5,49	1,138	0,403	0,535	4,96	0,293	0,294	0,551	0,038	0,053	0,313
<b>5390</b>	5,95	0,54	6,76	5,41	1,066	0,283	0,586	4,83	0,260	0,285	0,522	0,023	0,030	0,230
5462	6,74	0,65	8,89	6,76	1,338	0,794	0,496	6,26	0,289	0,337	0,712	0,048	0,102	0,644
5495	5,29	0,50	8,72	7,04	1,213	0,461	0,474	6,57	0,231	0,310	0,672	0,027	0,039	0,396
5599	4,51	0,67	6,69	5,05	1,180	0,454	0,499	4,55	0,282	0,330	0,568	0,035	0,060	0,359
<b>5751</b>	4,91	0,64	6,14	5,01	0,892	0,235	0,656	4,35	0,256	0,281	0,355	0,027	0,030	0,178
<b>5801</b>	5,20	0,57	8,63	7,41	0,966	0,258	0,517	6,89	0,215	0,242	0,509	0,019	0,018	0,222
<b>5832</b>	6,12	0,66	8,50	6,47	1,323	0,701	0,432	6,04	0,262	0,356	0,705	0,043	0,060	0,598
5850	3,09	0,66	8,47	6,49	1,180	0,795	0,595	5,90	0,257	0,310	0,613	0,061	0,092	0,641
5914	6,00	0,51	7,95	6,30	1,372	0,281	0,500	5,80	0,302	0,385	0,685	0,019	0,037	0,225
<b>8013</b>	7,39	0,87	1,58	1,46	0,038	0,081	0,031	1,43	0,006	0,007	0,024	0,000	0,009	0,071
<b>8038</b>	5,94	1,10	1,58	1,49	0,020	0,073	0,026	1,46	0,006	0,002	0,012	0,000	0,010	0,063
<b>8044</b>	7,89	1,00	1,43	1,25	0,060	0,122	0,029	1,22	0,003	0,016	0,041	0,000	0,013	0,109
<b>8053</b>	9,25	1,16	1,44	1,34	0,036	0,066	0,018	1,32	0,006	0,007	0,024	0,000	0,016	0,050
<b>8056</b>	9,23	0,94	1,29	1,22	0,031	0,045	0,026	1,19	0,005	0,004	0,022	0,000	0,007	0,038
<b>8066</b>	5,80	1,19	1,75	1,66	0,033	0,059	0,032	1,63	0,003	0,009	0,020	0,000	0,009	0,050
<b>8107</b>	7,40	0,86	1,25	1,15	0,040	0,056	0,021	1,13	0,005	0,008	0,027	0,000	0,007	0,049

**Annexe 6.1. :** Nombre de marqueurs AFLP générés à partir de 30 combinaisons d'amorces différentes.

Combinaison d'amorces		Nombre de bandes polymorphes
<i>Eco</i> + 3	<i>Mse</i> + 3	
AAC	CAA	3
AAC	CAC	10
AAC	CAG	3
AAC	CAT	10
AAC	CTA	10
AAC	CTC	4
AAC	CTG	10
AAC	CTT	7
AAG	CAT	11
AAG	CTA	8
AAG	CTC	1
AAG	CTG	6
AAG	CTT	12
ACA	CAA	5
ACA	CAC	4
ACA	CAT	3
ACA	CTA	10
ACA	CTC	5
ACA	CTG	8
ACA	CTT	6
ACC	CAA	7
ACC	CAC	9
ACC	CAT	4
ACC	CTA	5
ACC	CTG	4
ACC	CTT	6
ACT	CAA	5
ACT	CAT	7
ACT	CTC	5
ACT	CTT	7

**Annexe 6.2. : Locus (AFLP et RFLP) montrant des distorsions de ségrégation.**

Groupe	Locus	Hmz	Htz	N	$\chi^2 (1 : 1)$	$\chi^2 (1 : 3)$	$\chi^2 (3 : 1)$
<b>B</b>	AAGCTA5	15	47	62	16,52****	0,02ns	85,35*****
	ACTCAA1	12	50	62	23,29*****	1,05ns	102,39*****
	CA24BIS	8	34	42	16,10****	0,79ns	70,13*****
	ACACAT1	17	42	59	10,59**	0,46ns	67,12*****
<b>C</b>	AAGCTT3	15	47	62	16,52****	0,02ns	85,35*****
	ACTCTC4	3	56	59	47,61*****	12,48***	153,81*****
	ACACAA1	5	52	57	38,75*****	8,01**	133,34*****
	ACTCAT7	9	52	61	30,31*****	3,42ns	118,08*****
	AACCAT8	11	50	61	24,93*****	1,58ns	105,58*****
	AACCAT9	12	50	62	23,29*****	1,05ns	102,39*****
	AAGCTG5	8	51	59	31,34*****	4,12*	118,79*****
	AACCTG9	18	43	61	10,25**	0,66ns	67,33*****
	AAGCAT4	19	43	62	9,29**	1,05ns	65,05*****
	AACCTC1	19	42	61	8,67**	1,23ns	62,56*****
	ACACAA4	19	42	61	8,67**	1,23ns	62,56*****
	AACCTG3	17	44	61	11,95***	0,27ns	72,27*****
	AAGCAT3	19	42	61	8,67**	1,23ns	62,56*****
	ACACTT3	20	41	61	7,23**	1,97ns	57,97*****
	AACCTT3	14	45	59	16,29****	0,05ns	82,72*****
AACCAT3	17	45	62	12,65***	0,19ns	74,86*****	
<b>E</b>	ACACTG8	21	37	58	4,41*	3,89*	46,55*****
	ACCCAC7	21	39	60	5,40*	3,20ns	51,20*****
	AAGCTA6	15	46	61	15,75****	0,01ns	82,67*****
	AACCAT6	14	45	59	16,29****	0,05ns	82,72*****
	AAGCTT8	18	44	62	10,90***	0,54ns	69,87*****
<b>F</b>	ACACTA2	16	42	58	11,66***	0,21ns	69,54*****
	AACCAC1	15	42	57	12,79***	0,05ns	72,05*****
	AACCAC9	17	39	56	8,64**	0,86ns	59,52*****
<b>G</b>	AACCTA5	21	40	61	5,92*	2,89ns	53,56*****
	AACCAT1	14	48	62	18,65*****	0,19ns	90,86*****
<b>H</b>	ACACTG7	40	21	61	5,92*	53,56*****	2,89ns
	AACCTG6	39	21	60	5,40*	51,20*****	3,20ns
	AAGCTA2	41	20	61	7,23**	57,97*****	1,97ns
	AACCAC5	43	17	60	11,27***	69,69*****	0,36ns
	ACCCAC4	42	19	61	8,67**	62,56*****	1,23ns
	ACTCTC1	42	20	62	7,81**	60,41*****	1,74ns
	ACCCTG4	40	16	56	10,29**	64,38*****	0,38ns
	ACTCAT1	41	21	62	6,45*	55,94*****	2,60ns
	AAGCAT2	41	21	62	6,45*	55,94*****	2,60ns
	ACCCAC1	47	14	61	17,85*****	88,14*****	0,14ns
	ACACTT5	52	4	56	41,14*****	137,52*****	9,52 **
	ACACTT6	54	4	58	43,10*****	143,47*****	10,14 **
	AACCTA3	48	10	58	24,90*****	103,20*****	1,86ns
<b>I</b>	ACCCAC3	40	20	60	6,67**	55,56*****	2,22ns
	ACCCTA4	43	18	61	10,25**	67,33*****	0,66ns
	AAGCTG6	44	17	61	11,95***	72,27*****	0,27ns
	ACACTT4	45	17	62	12,65***	74,86*****	0,19ns
	ACACTG1	45	17	62	12,65***	74,86*****	0,19ns
	ACTCAA4	46	16	62	14,52***	80,02*****	0,02ns
	AAGCTT6	40	17	57	9,28**	62,04*****	0,71ns
	AACCTG5	44	17	61	11,95***	72,27*****	0,27ns
	ACTCTT3	44	18	62	10,90***	69,87*****	0,54ns
AACCTG7	40	21	61	5,92*	53,56*****	2,89ns	
<b>J</b>	ACACTT2	17	45	62	12,65***	0,19ns	74,86*****

**Annexe 7.1.** : Principales caractéristiques des QTLs détectés pour un LOD score > 2,0 [localisation : groupe de liaison, position en cM à partir du premier marqueur du groupe, marqueurs encadrant le LOD maximum , pourcentage de variance expliqué par chaque QTL (%) et par le modèle multi-QTL (% total)].

Caractère	Groupe	Position	Intervalle	LOD	%	% total
CQA45	G	37	ACTCAT5-ACACTT1	2,67	20,7	
DQA34	C	102	ACACTT3-AACCTT3	2,10	16,7	29,2
	J	1	ACACTT2-ACTCTT6	2,67	20,7	
DQA34/J	C	102	ACACTT3-AACCTT3	2,32	24,0	
DQA45	B	84	ACCCTT3-ACACTA9	3,00	22,9	
CQAt	G	37	ACTCAT5-ACACTT1	2,65	20,6	22,4
	G	71	ACACTC2-AACCTA5	2,26	20,4	
DQAt	B	84	ACCCTT3-ACACTA9	2,42	18,9	25,4
	C	102	ACACTT3-AACCTT3	2,17	17,2	
CGA1	A	108	AR109-AACCAG2	2,27	20,7	
CGA2	A	108	AR109-AACCAG2	2,62	24,8	
% CQAt	B	76	AAGCTA1-CA30	2,85	23,2	42,0
	G	30	ACACTG6-ACACTC3	2,53	19,8	
	G	41	AAGCTT10-ACCCAA6	2,57	20,0	
% DQA34	G	37	ACTCAT5-ACACTT1	2,03	16,2	24,5
	G	82	AACCTA5-AACCAT1	2,59	24,1	
% DQA35	B	35	ACTCAA1-CA24bis	2,26	19,6	34,8
	G	87	AACCTA5-AACCAT1	2,55	25,1	
% DQA45	B	76	AAGCTA1-CA30	2,41	21,6	40,1
	B	92	AAGCTG4-AACCTT7	2,67	25,5	
	K	3	AAGCAT7P-AACCAT10	2,67	25,5	
% DQAt	B	84	ACCCTT3-ACACTA9	2,96	22,7	39,8
	G	41	AAGCTT10-ACCCAA6	2,28	18,0	
FQAt/J	C	15	AAGCTT3-ACTCTC4	2,82	49,1	
FQA5/J	C	15	AAGCTT3-ACTCTC4	2,27	40,1	
FQA4/J	C	15	AAGCTT3-ACTCTC4	3,26	61,4	
FQA3/J	C	15	AAGCTT3-ACTCTC4	3,03	43,5	
CQA3/J	A	108	AR109-AACCAG2	2,74	24,5	
CAF	A	204	AACCAC10-AACCTT4	2,66	20,3	39,5
	D	62	AACCAT5-CA71	3,12	22,3	
	E	22	G13-AACCAG1	2,58	21,7	
CAF/CGA1	A	204	AACCAC10-AACCTT4	3,59	33,6	42,4
	D	62	AACCAT5-CA71	2,95	23,4	
CAF/CQA	A	204	AACCAC10-AACCTT4	3,91	36,4	45,2
	D	62	AACCAT5-CA71	3,10	24,3	
CAF/J	D	62	AACCAT5-CA71	2,89	27,8	51,3
	G	8	Couleur-ACCCAA2	2,43	30,3	
	G	80	AACCTA5-AACCAT1	3,07	29,2	
HET	B	102	AAGCTT8P-AAGCTT5	6,51	43,3	
HET/J	B	112	AAGCTT5-AAGCTG3	4,22	39,6	
TRI	G	35	ACTCAT5-ACACTT1	3,56	28,4	30,5
	G	90	ACACTC2-AACCTA5	3,08	31,6	
TRI/J	A	152	ACTCAA3-ACCCAA3	2,73	31,6	
P100	G	85	AACCTA5-AACCAT1	4,30	34,1	
P100/J	C	32	ACACAA1-ACTCAT7	2,18	21,8	52,2
	F	21	ACCCAA1-AACCTA7	2,96	28,4	
	G	85	AACCTA5-AACCAT1	3,01	28,7	
	J	47	ACACAC4-ACACTG2	2,42	24,4	
DCF	E	15	ACCCTT1-G13	3,97	37,4	46,0
	H	1	ACACTG7-AACCAT1	2,74	26,0	
FECA	C	20	AAGCTT3-ACTCTC4	2,70	38,6	56,0
	H	55	ACACTT6-AACCTA3	3,18	33,3	
VP	D	74	CA71-ACACAC1	2,28	16,4	27,2
	L	24	AACCTG8-AAGCAT9	2,27	16,8	

**Annexes : A, B, C et D (articles)**

---

**Annexe A**

---

## COMPARISON OF FIVE PURIFICATION METHODS FOR CHLOROGENIC ACIDS IN GREEN COFFEE BEANS (*COFFEA* SP.)

KY C.L., NOIROT M. & HAMON S.

Address: Centre ORSTOM, 34032 MONTPELLIER Cedex, FRANCE

**Abstract:** Quantification of chlorogenic acid content in large populations of green coffee beans needs an accurate, fast, and unbiased purification method. Five different procedures of purification were compared. The first consisted of a successive use of different organic solvents, the second was based on a filtration through a C<sub>18</sub> cartridge, the third used two combined reagents, and the remaining two methods (4 and 5) were a simplification of the third. One of the two simplest methods of purification (method 4) was also the fastest, the most accurate and the least biased. Consequently, this method could be used routinely to quantify chlorogenic acids in green coffee beans.

**Résumé:** Le dosage quantitatif de la teneur en acides chlorogéniques chez de grandes populations de graines de café vert nécessite une méthode de purification précise, rapide et non-biaisée. Cinq procédés différents de purification ont été comparés. Le premier consiste en une utilisation successive de différents solvants organiques. Le second est basé sur une filtration au travers d'une cartouche C<sub>18</sub> et le troisième utilise deux réactifs. Les deux méthodes restantes (4 et 5) sont des simplifications de la troisième. Une des deux plus simples méthodes de purification (la méthode 4) est aussi la plus rapide, la plus précise et la moins biaisée. Cette méthode pourrait être ainsi utilisée en routine pour quantifier les acides chlorogéniques chez les graines de café vert.

**Key-words:** Chlorogenic acids, *Coffea* sp., purification procedures, HPLC

### INTRODUCTION

Quinyl esters of hydroxycinnamic acids, which are generally known as chlorogenic acids (CGA) in green coffee beans were first reported by ROBIQUET and BOUTRON in 1837 (SONDHEIMER, 1964). CGA are present in many *Coffea* species (CLIFFORD, 1985a, b; CLIFFORD *et al.*, 1989; RAKOTOMALALA *et al.*, 1993; ANTHONY *et al.*, 1993). In the two main cultivated species, *Coffea canephora* and *C. arabica*, CGA account for 7 to 10 per cent and 5 to 8 per cent of dry matter basis (dmb), respectively (CLIFFORD, 1985b; RAKOTOMALALA, 1992). CGA include three main groups which are caffeoylquinic acids (CQA), feruloylquinic acid (FQA), and dicaffeoylquinic acids (diCQA) (CLIFFORD *et al.*, 1977). Each group contains three isomers on the basis of the number and identity of the acylating residues (CLIFFORD, 1985a, b).

Quantitative analysis of CGA in large populations needs an accurate, non-biased, fast and cheap method. An accurate method allows reduction in sample size without a decrease in precision. A non-biased method should give estimations centred on average on the true value.

The experimental procedure to estimate CGA contents includes three steps: 1) extraction, 2) purification, and 3) separation and quantification of the different isomers by HPLC. The extraction, separation and quantification of the different isomers are well defined in terms of their optimal chemical reagents or solvents (COLONNA, 1979; CLIFFORD *et al.*, 1985; TRUGO, 1984; VAN DER STEGEN *et al.*, 1980). By contrast, at least four different purification procedures are described in the literature, whereas a fifth variation is the standard technique used in Germany.

The first method consists of the successive use of different organic solvents (FLEURIET & MACHEIX, 1972). The second is a purification through a C<sub>18</sub> cartridge (Sep-Pak ref.51910, Millipore Waters) (BICCHI *et al.*, 1995). The third consists of a treatment by two specific reagents (Carrez Reagents) to precipitate colloidal material (BALYAYA & CLIFFORD, 1995). Removal of the rotary evaporation step leads to the next method (TRUGO & MACRAE, 1984), whereas a fifth method differs from the others by absence of cleaning with Carrez solutions. These five methods were compared, in terms of their precision and efficiency.

### MATERIALS AND METHODS

#### PLANT MATERIAL

Green coffee beans were supplied by the Agricultural Coffee Tree Station ORSTOM-IDEFOR (Man - Ivory Coast). In the first experiment, we used 1,000 beans of the species *Coffea liberica* var *dewevrei* (BRIDSON, 1994) from ten trees. In the second experiment, 100 beans of the same species were used.

#### EXTRACTION PROCEDURE

The beans were divided into 100-bean lots to be crushed in a ball mill (Dangoumeau), which was previously frozen by liquid nitrogen in order to minimize CGA degradation (COLONNA, 1979). Crushing yielded a fine powder (15 min per lot). All powder was bulked again before being split into 1g samples. Each sample was extracted in a 250 mL Erlenmeyer flask contain-

ning 100 mL of a methanol-water mixture (70/30) and 0.5 % Na<sub>2</sub>SO<sub>3</sub>. The Erlenmeyer was shaken overnight at 4°C in darkness on a stirring table at 125 rpm (Colonna, 1979). The organic extract was filtered through cotton to eliminate powder. All extracts were bulked again in order to obtain a homogeneous extract for comparison of purification methods.

#### EXPERIMENTAL DESIGN

A first experiment involved the comparison of the four first methods. Five replicate purifications were done per method. The 20 samples were totally randomized before purification. A second experiment allowed the testing of the Carrez reagents step (comparison between methods 4 and 5). Four repetitions were performed per method.

#### PURIFICATION PROCEDURES

The first step of methods 1 to 3 was elimination of methanol by rotary evaporation (Rotavapor) to obtain an aqueous extract.

**Method 1:** (RAKOTOMALALA, 1992). To aqueous ex-tracts (about 22 ml each) were added ammonium sul-fate to a final concentration of 20 g/L (precipitation of proteins by an increase in ionic strength) and 4 % phosphoric acid (acidification of the medium to make the CGA more soluble in ethyl acetate). Extracts were then treated three times with petroleum ether (40°-60°) to remove lipids and pigments. Depigmented extracts were treated three times with CHCl<sub>3</sub> to eliminate caffeine and wax insoluble in petrol. Residual solvent was then eliminated in a rotary evaporator Bio-block 94200 (Rotavapor). CGA were extracted by four successive treatments with ethyl acetate. The four ethyl acetate phases were pooled and dried by anhydrous Na<sub>2</sub>SO<sub>4</sub>. After filtration (Whatman 1PS), ethyl acetate was evaporated (Rotavapor), and the residue was dissolved in 100 mL of methanol.

**Method 2:** (BICCHI *et al.*, 1995). Aqueous extracts (about 30 ml) were filtered through a 0.2 µm Millipore filter, and 6 mL aliquots were loaded on a C<sub>18</sub> Sep-Pak cartridge 51910 (Millipore Waters) previously conditioned with 5 mL of methanol and 3 mL of distilled H<sub>2</sub>O of HPLC grade. The cartridge was eluted with 20 mL of methanol/water mixture (70/30), and the eluate was then directly submitted to HPLC-UV analysis.

**Method 3:** (BALYAYA & CLIFFORD, 1995). Carrez reagents consist of solutions I and II. Solution I was prepared by dissolving 21.9 g of crystallized zinc acetate and 3 mL of glacial acetic acid in 100 mL of distilled water. Solution II consisted of 10.6 g of potassium hexacyanoferrate (Fe<sup>++</sup>) in 100 ml of distilled water. Colloidal material present in aqueous extracts (about 30 ml) was precipitated by adding 1 mL of so-

lution I, plus 1 mL of solution II, diluted to 50 ml with methanol-water mixture (70/30).

**Method 4:** (TRUGO & MACRAE, 1984). Methanol extracts (100 ml) were directly treated with Carrez reagents, without previous evaporation in a Rotavapor.

**Method 5:** (DIN Standard 10 767, 1992). Methanol extracts (100 ml) were directly analysed by HPLC.

#### ANALYTICAL HPLC

Chromatography was carried out on a system consisting of two Waters Associates Model 510 pumping units, an automated sample injector (Waters 717 plus autosampler), a variable-wavelength UV detector (Waters 996 Photodiode Array Detector), a C<sub>18</sub> pre-column and a 250 mm x 4 mm Merck Superspher 100 RP 18 column, 5µm particle size. The elution program used two solvents, A and B. Solvent A was 2 mM phosphoric acid, pH 2.7, containing 5% methanol. Solvent B was methanol containing 5% of 2 mM phosphoric acid, pH 3.9. These two mobile phases were filtered (0.2µm), degassed and sonicated (Ney, 300 ultrasonik) before use. Sample or standard (10 µL) was analyzed at room temperature using the following elution program: A-B mixture (75/25) to pure solvent B in 45 min. of linear gradient. Flow rate was 0.8 mL/min. UV detection was carried out at a wave-length of 325 nm, corresponding to the CGA maximum absorption.

#### ISOMER IDENTIFICATION

CGA isomers (3-CQA, 4-CQA, 5-CQA, 3-FQA, 4-FQA, 5-FQA, 3,4-diCQA, 3,5-diCQA, and 4,5-diCQA) were identified by comparing chromatograms (280 nm and 325 nm), retention times and UV spectra to published results (RAKOTOMALALA, 1992). Identification of the CQA isomers was confirmed by isomerization of 5-CQA standard (ref. D11.080-9, Aldrich-Chimie). Isomerization was achieved by heating the standard in dilute ammonia solution at pH 8 for 30 min (TRUGO *et al.*, 1984).

#### ISOMER QUANTIFICATION

Quantification was achieved by peak-area measurement and by comparison with a 5-CQA standard. A calibration curve ( $C=0.27896 \times \text{peak area}$ ) was generated using three replicate points of 5-CQA at 25, 50, 75, 100, 125, and 150 mg/L. All amounts were referred to the same dilution factors, corresponding to the final volume of 100 ml, in order to obtain comparable results.

#### STATISTICAL ANALYSIS

Results were analysed using Statistica software. The precision of the methods was compared by testing the homogeneity of their variance (LEVENE, 1960). In the first experiment, heterogeneity of variances led to compare efficiency of methods by the median test

(CONOVER, 1971). In the second experiment, homogeneity of variances permitted the use of a one-way ANOVA. These tests were applied to the absolute amounts of each isomers, total CQA, total FQA, total diCQA, and total CGA.

## RESULTS

The major problem encountered with the chromatographic resolution was the peak due to 4-CQA (TRUGO & MACRAE, 1984). The 4-CQA peak was located by isomerization of the 5-CQA standard, which yielded an equilibrium mixture of the 3-, 4- and 5-isomers. The 4-CQA and 5-CQA isomer peaks overlapped (Fig. 1), and so the common peak of these undifferentiated isomers was called the 5-& 4-CQA peak. Eight peaks were then considered for the routine statistical analysis (Fig. 2). The 5-& 4-CQA peak was the main peak whatever the method, followed by the diCQA peaks. The FQA peaks were the lowest.

### FIRST EXPERIMENT

Differences between methods were recorded for each isomer (Table 1). For CQA isomers and total CQA, the best methods were 3 and 4, with a average gain of 153% by comparison with methods 1 and 2. For FQA isomers and total FQA, method 4 was significantly better than the three others. The gain was then 84%, 111%, and 146% by comparison with methods 3, 2, and 1, respectively. For diCQA isomers and total diCQA, methods 2 and 4 were similar and better than methods 1 and 3. The gain was then 53% and 108% by comparison with methods 3 and 1, respectively.

Table 1 Median contents (% dmb), chi-square values ( $\chi^2$ ), and associated probabilities (P) for the different chlorogenic acid isomers. Results of 2-by-2 comparisons are indexed.

	Method 1	Method 2	Method 3	Method 4	$\chi^2$	P
3-CQA	0.2 <sup>a</sup>	0.2 <sup>a</sup>	0.7 <sup>b</sup>	0.8 <sup>b</sup>	19.0	0.000
5-& 4-CQA	2.7 <sup>a</sup>	2.1 <sup>a</sup>	5.6 <sup>b</sup>	6.5 <sup>b</sup>	19.0	0.000
Total CQA	2.9 <sup>a</sup>	2.3 <sup>a</sup>	6.2 <sup>b</sup>	7.3 <sup>b</sup>	19.0	0.000
3-FQA	0.03 <sup>a</sup>	0.04 <sup>a</sup>	0.04 <sup>a</sup>	0.08 <sup>b</sup>	10.1	0.017
4-FQA	0.07 <sup>a</sup>	0.09 <sup>a</sup>	0.09 <sup>a</sup>	0.15 <sup>b</sup>	10.1	0.017
5-FQA	0.22 <sup>a</sup>	0.23 <sup>a</sup>	0.27 <sup>b</sup>	0.53 <sup>b</sup>	10.1	0.017
Total FQA	0.31 <sup>a</sup>	0.36 <sup>ab</sup>	0.41 <sup>b</sup>	0.76 <sup>c</sup>	10.1	0.017
3,4diCQA	0.18 <sup>a</sup>	0.43 <sup>c</sup>	0.22 <sup>ab</sup>	0.32 <sup>b</sup>	9.5	0.022
3,5diCQA	0.19 <sup>a</sup>	0.47 <sup>c</sup>	0.26 <sup>b</sup>	0.37 <sup>b</sup>	9.5	0.022
4,5diCQA	0.39 <sup>a</sup>	0.90 <sup>c</sup>	0.56 <sup>b</sup>	0.73 <sup>b</sup>	9.5	0.022
Total diCQA	0.78 <sup>a</sup>	1.80 <sup>c</sup>	1.05 <sup>b</sup>	1.43 <sup>b</sup>	9.5	0.022
Total CGA	4.1 <sup>a</sup>	4.6 <sup>a</sup>	7.7 <sup>b</sup>	9.5 <sup>c</sup>	19.0	0.000

The lower the within-method variance, the better the absolute precision, which did not differ between the methods for CQA isomers (Table 2). By contrast, the precision of method 2 was clearly lower for FQA and diCQA isomers. The same respective conclusions were reached when isomers were regrouped in total CQA, total FQA, and total diCQA. The relative precision, measured by the coefficient of variation, was again lowest with method 2 (30-43%). Relative precision was mediocre with method 3 (8-22%), and

acceptable with method 1 (2-8%). Method 4 was still the best method (1-7%), especially for FQA (Table 2).

### SECOND EXPERIMENT

For all isomers, estimations by method 4 were higher than by method 5, even if the test was not significant for 3- and 4-FQA (Table 3). For total CQA, total FQA and total diCQA, the gain was of 15%, 6% and 37% respectively. The precision obtained did not differ between methods 4 and 5 for any isomers.

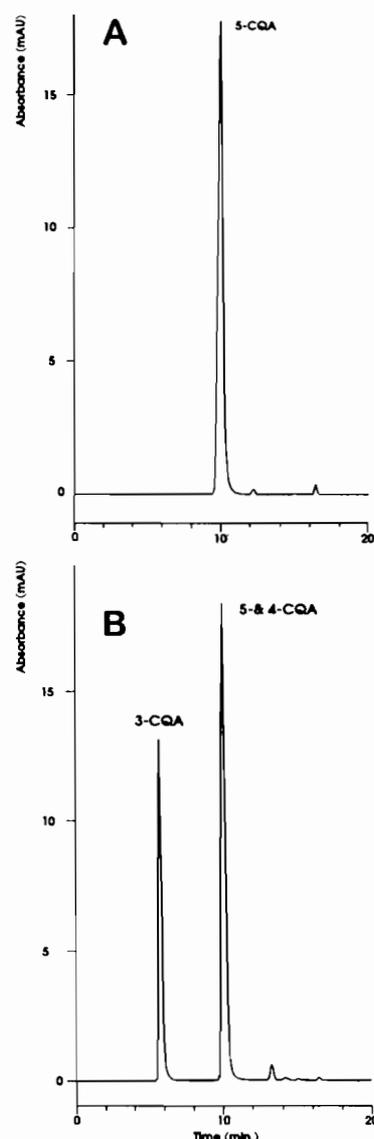


Figure 1. Chromatograms of (A) 5-CQA standard and (B) CQA isomers mixture obtained by isomerization of 5-CQA. Detection at 325 nm and chromatographic conditions as described in the text

## DISCUSSION

The method giving the higher contents was considered as being the least biased. Indeed, the true and unknown estimation would be obtained if, and only if, no loss occurs during the different steps of the analysis. This ideal case cannot be reached; however we

only expect a minimal loss (bias). Consequently, method 4 was the least biased for all chlorogenic acids, although methods 2 and 3 were similar to method 4 in the case of diCQA and CQA, respectively. Nevertheless, the quality of our choice supposes that the results we obtained with other methods were similar to literature values. Our results were comparable with published data on *Coffea liberica* var *dewevrei* reported by ANTHONY *et al.* (1993) using method 3 and by RAKOTOMALALA (1992) using method 1 (Table 4).

Table 2 Within-method variation Standard deviations and coefficients of variation (in brackets) are given for each method and variable F is the observed value and P the associated probability from comparison of variance (Levene test)

	Method 1	Method 2	Method 3	Method 4	F	P
3-CQA	0.016 (8%)	0.074 (31%)	0.070 (10%)	0.041 (5%)	1.83	>0.05
5-& 4-CQA	0.143 (5%)	0.610 (37%)	0.441 (8%)	0.156 (3%)	2.63	>0.05
Total CQA	0.156 (6%)	0.688 (31%)	0.506 (8%)	0.195 (3%)	2.51	>0.05
3-FQA	0.001 (6%)	0.015 (43%)	0.008 (18%)	0.001 (2%)	4.55	0.018
4-FQA	0.010 (5%)	0.039 (41%)	0.012 (13%)	0.001 (1%)	4.79	0.015
5-FQA	0.014 (6%)	0.105 (41%)	0.040 (14%)	0.006 (1%)	6.31	0.005
Total FQA	0.018 (6%)	0.159 (41%)	0.061 (14%)	0.005 (1%)	6.05	0.006
3,4diCQA	0.005 (3%)	0.115 (30%)	0.050 (22%)	0.022 (7%)	3.68	0.036
3,5diCQA	0.008 (4%)	0.152 (35%)	0.049 (18%)	0.025 (7%)	3.81	0.032
4,5diCQA	0.009 (2%)	0.250 (31%)	0.086 (15%)	0.034 (5%)	4.34	0.021
Total diCQA	0.013 (3%)	0.519 (32%)	0.182 (17%)	0.078 (6%)	4.07	0.026
Total CGA	0.182 (5%)	1.325 (31%)	0.733 (9%)	0.286 (3%)	2.66	>0.05

Table 3 Comparison of methods 4 and 5 Mean contents (% dmb), F values, and associated probabilities (P) are done for each chlorogenic acid isomers

	Method 4	Method 5	F	P
3-CQA	0.59	0.50	15	0.008
5-& 4-CQA	4.8	4.2	15	0.008
Total CQA	5.4	4.7	15	0.008
3-FQA	0.054	0.049	5.46	0.058
4-FQA	0.054	0.049	5.46	0.058
5-FQA	0.37	0.35	8.87	0.025
Total FQA	0.48	0.45	7.90	0.030
3,4diCQA	0.26	0.19	99	0.000
3,5diCQA	0.28	0.23	46	0.000
4,5diCQA	0.56	0.43	55	0.000
Total diCQA	1.1	0.8	63	0.000
Total CGA	6.95	6.05	20	0.004

Simplification of methods led to an improvement of results, except for method 5. Although the latter is a standard method to estimate CGA content, the absence of Carrez reagents decreased recovery in the case of coffee beans. As Carrez reagents precipitate polysaccharides, soluble proteins, and other colloidal materials, the latter observation suggests that some of these products could react with CGA during the HPLC

phases in relation with the acidification of medium (TRUGO, 1984; CLIFFORD, 1985a).

In conclusion, method 4 should be selected for routine analysis of CGA in green beans of *Coffea* species when repeatability, precision, rapidity and cost are at a premium.

Table 4 Comparison of our data with those of Anthony *et al.* (1993) using method 3 of purification, and those of Rakotomalala (1992) using method 1 Data are expressed in % dmb

	Our results		Literature results	
	Method 1	Method 3	Method 1	Method 3
Total CQA	2.9	6.2	2.6	6.1-6.65
Total diCQA	0.78	1.05	-	0.64-0.98
Total FQA	0.31	0.41	0.28	0.38-0.61

## ACKNOWLEDGEMENTS

We are grateful to Mr P. TROUSLOT for technical assistance, to Mrs S. DOULBEAU, Mr P. BARRE and Dr. J.-L. VERDEIL for help with HPLC, to Dr. B. GUYOT (CIRAD-CP) for technical advice and to Dr. J. TREGEAR for the english manuscript correction.

## LITERATURE CITED

- ANTHONY, F., CLIFFORD, M.N. & NOIROT, M. 1993. Biochemical diversity in the genus *Coffea* L.: chlorogenic acids, caffeine and mozambioside contents. *Gen. Res. and Crop Evol.* 40: 61-70.
- BALYAYA, K.J. & CLIFFORD, M.N. 1995. Individual chlorogenic acids and caffeine contents in commercial grades of wet and dry processed indian green robusta coffee. *J. Food Sci. Technol.* 32: 104-108.
- BICCHI, C.P., BINELLO, A.E., PELLEGRINO, G.M. & VANNI, A.C. 1995. Characterization of green and roasted coffees through the chlorogenic acid fraction by HPLC-UV and principal component analysis. *J. Agric. Food Chem.* 43: 1549-1555.
- BRIDSON, D.M. 1994. A new forma in *Coffea liberica* (Rubiaceae). *Kew Bull.* 36: 314.
- CLIFFORD, M.N. & STANIFORD, P.S. 1977. A critical comparison of six spectrophotometric methods for measuring chlorogenic acids in green coffee beans. *Proc Int. Congr. of ASIC* 8: 109-113.
- CLIFFORD, M.N. 1985a. Chlorogenic acids. In *Chemistry*; CLARKE R.J., MACRAE R. (Eds.), Elsevier Appl. Sci., London, Vol. 1, Chapter 5.
- CLIFFORD, M.N. 1985b. Chemical and physical aspects of green coffee and coffee products. In *Coffee Botany, Biochemistry and Production of Beans and Beverage*; Clifford, M.N., Willson, K.C., Eds.; The Avi Publishing Company: Westport, Connecticut, Chapter 13.

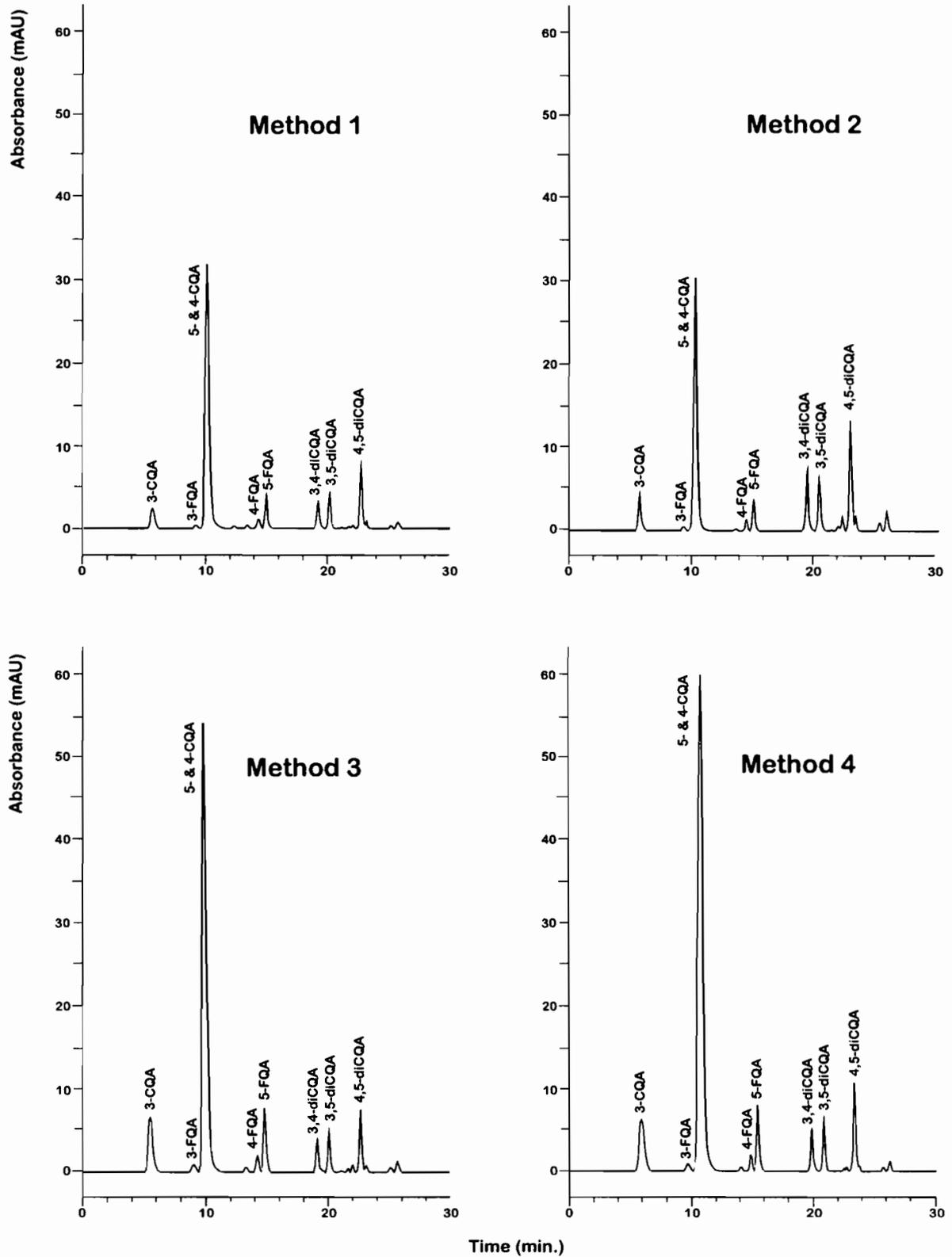


Figure 2 HPLC chromatograms of the four methods of purification of chlorogenic acids from green *Coffea liberica dewevrei* beans. Chromatographic conditions Solvent A 5 % methanol in 2 mM phosphoric acid, Solvent B 5% of 2 mM phosphoric acid in pure methanol, gradient A-B mixture (75-25) to pure B solvent in 45 min, column 4 mm x 250 mm Merck Superspher C<sub>18</sub>, 10 µl injection, flow-rate, 0.8 mL min<sup>-1</sup>, detection at 325 nm

- CLIFFORD, M.N., OHIKPEHAI, O. & DE MENEZES, H. 1985. The influence of extraction method and analytical method on the chlorogenic acid content of green coffee beans. *Proc. Int. Congr. ASIC* 11: 252-262.
- CLIFFORD, M.N., WILLIAMS, T. & BRIDSON, T. 1989. Chlorogenic acids and caffeine as possible taxonomic criteria in *Coffea* and *Psilanthus*. *Phytochem.* 28: 829-838.
- COLONNA, J.P. 1979. *L'acide chlorogénique et les depsides de divers caféiers africains et malgaches: leur participation au métabolisme et leur signification biologique*. Thesis of the Univ. Paul Sabatier, Toulouse, France.
- CONOVER, W.J. 1971. *Practical Nonparametric Statistics*; CONOVER, W.J., Eds.; John Wiley & sons, New York.
- FLEURIET, A. & MACHEIX, J.J. 1972. Séparation et dosage par chromatographie en phase gazeuse de l'acide chlorogénique et des catéchines des fruits. *J. Chromatogr.* 198: 373-376.
- LEVENE, H. 1960. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, OLKIN, I. et al. (Eds.), Stanford Univ. Press, Stanford, California, pp 278-292.
- RAKOTOMALALA, J.J.R. 1992. *Diversité biochimique des caféiers, Particularités des caféiers sauvages de la région malgache*. Thesis of the Univ. Sci. & Techn. of Languedoc, Montpellier, France
- RAKOTOMALALA, J.J.R, CROS, E., CHARRIER, A. ANTHONY, A. & NOIROT, M. 1993. Marqueurs biochimiques de la diversité des caféiers. *Proc. Int. Congr. ASIC*, 15: 79-84.
- SONDHEIMER, E. 1964. Chlorogenic acids and related depsides. *Bot. Rev.* 30: 667-712.
- TRUGO, L.C. 1984. *HPLC in coffee analysis*. Ph. D Univ. of Reading, Reading, U.K.
- TRUGO, L.C. & MACRAE, R. 1984. Chlorogenic acid composition of instant coffees. *Analyst* 109: 263-266.
- VAN DER STEGEN, G.H.D. & VAN DUJIN, J. 1980. Analysis of chlorogenic acids in coffee. *Proc. Int. Congr. ASIC* 9: 107-112.



## RELATIONS BETWEEN AND INHERITANCE OF CHLOROGENIC ACID CONTENTS IN AN INTERSPECIFIC CROSS BETWEEN *COFFEA PSEUDOZANGUEBARIAE* AND *C. LIBERICA* VAR *DEWEVREI*

KY C.L.<sup>\*</sup>, LOUARN J.<sup>\*\*</sup>, GUYOT B.<sup>\*\*\*</sup>, CHARRIER A.<sup>\*\*\*\*</sup>, HAMON S.<sup>\*</sup> & NOIROT M.<sup>\*</sup>

Addresses: \* Centre ORSTOM, BP 5045, 34032 MONTPELLIER Cedex 1, FRANCE.

\*\* Station ORSTOM, BP 434, MAN, COTE-D'IVOIRE.

\*\*\* CIRAD-CP, BP 5035, 34032 MONTPELLIER Cedex 1, FRANCE.

\*\*\*\* ENSA, place Viala, F-34060 MONTPELLIER, FRANCE

**Abstract:** Chlorogenic acids (CGA) are phenolic compounds commonly found in green coffee beans. The main CGA classes are caffeoylquinic acids (CQA), dicaffeoylquinic acids (diCQA), and feruloylquinic acids (FQA). Each contains three isomers differing in the number and identity of the acylating residues. An interspecific cross between *Coffea pseudozanguebariae* (low CGA content) and *C. liberica* var *dewevrei* (high CGA content) was investigated for CGA contents in F1 and back-cross hybrids. Relations within and between CGA classes were studied and confirmed the known biosynthesis pathway. A single major gene was noted for the 3-FQA isomer (absence was dominant). Additivity was found for most other isomers either with or without transformation of variables. Conversely, most ratios were not additive, due to a curvilinear relation between some isomers. The consequences for breeding both in terms of cup taste improvement and disease resistance are discussed.

**Résumé:** Les acides chlorogéniques (ACG) sont des composés phénoliques communs chez les graines vertes de café. Les principales classes de ACG sont les acides caféylquiniques (ACQ), les acides dicaféylquiniques (AdiCQ) et les acides férulylquiniques (AFQ). Chacune contient trois isomères différant par la position des résidus. Un croisement interspécifique entre *Coffea pseudozanguebariae* (faible teneur en ACG) et *C. liberica* var *dewevrei* (forte teneur en ACG) a été étudié pour les teneurs en ACG chez les hybrides F1 et rétrocroisés. Les relations intra- et interclasses d'ACG confirment les voies connues de la biosynthèse. Un gène majeur a été noté pour l'isomère 3-AFQ (l'absence est dominante). L'additivité a été observée pour la plupart des autres isomères, soit avec ou sans transformation préalable de la variable. A l'inverse, la plupart des rapports ne sont pas additifs et ceci est dû à la présence de relations curvilinéaire entre certains isomères. Les conséquences pour l'amélioration sont discutées à la fois pour l'amélioration du goût à la tasse et la résistance aux maladies.

**Key words** Chlorogenic acids, Quantitative inheritance, Interspecific cross, *Coffea pseudozanguebariae*, *C. liberica dewevrei*

### INTRODUCTION

The esters of quinic acid are trivially named chlorogenic acids (CGA) and constitute the major phenolic secondary metabolites found in the plant kingdom. They control germination and cell growth, participate in defense mechanisms against phytopathogens or act as lignin precursors. CGA are particularly abundant in the coffee plant (STREULI, 1970; AMORIN *et al.*, 1974; MAIER, 1987), especially in green beans and their derivatives - roasted coffee beans, soluble coffee powders and coffee brews - (CLIFFORD, 1985a; BALYAYA & CLIFFORD, 1995). In coffee trees, CGA include 6 classes: caffeoylquinic acids (CQA), dicaffeoylquinic acids (diCQA), feruloylquinic acids (FQA), *p*-coumaroylquinic acids (*p*CoQA), caffeoylferuloylquinic acids (CFQA), and feruloylcaffeoylquinic acids (FCQA). CGA-like components where the caffeic unit is coupled with an amino acid such as tyrosine are also present as traces (CLIFFORD *et al.*, 1989a; CORREIA *et al.*, 1995). The first three classes represent about 98 % of the CGA content (CLIFFORD & STANFORTH, 1977; MORI-SHITA *et al.*, 1989). Each of these classes is commonly divided into three isomers on the basis of the number and position of the acylating residues (CLIFFORD, 1985b).

The CGA content of coffee beans modifies cup taste through direct and indirect effects. The direct effect is due to the presence of CGA after roasting [36% in *C. canephora* var *Robusta* (LENTNER & DEATHERAGE, 1959; FELDMAN *et al.* 1969)]. As an example, diCQA is known to increase astringency (OHIKPEHAI *et al.*, 1982; CLIFFORD & OHIKPEHAI, 1982). The indirect effects are due to molecular changes during roasting which have positive or negative influences. Degradation of CGA into phenolic derivatives (LELOUP *et al.*, 1995) and inhibition of pyrazine formation by Maillard's reactions (GUYOT *et al.*, 1997) have negative effects. Conversely, the CGA reaction with trigonelline, sucrose, and amino acids which enriches headspace volatiles profiles recorded by gas chromatography is a positive effect (DE MARIA *et al.*, 1994). To summarize, the quality of the beverage increases when the CGA content decreases. This largely explains taste differences between Robus-

ta and Arabica coffees (CLIFFORD, 1985b; GUYOT *et al.*, 1988; GUYOT *et al.*, 1996).

Chlorogenic acid content increases resistance to *Ceratocystis fimbriata* at high concentrations but stimulates pathogen growth at low concentrations (ECHANDI & FERNANDEZ, 1962; ZULUAGA *et al.*, 1971). Coffee species producing more CGA will be well protected. It seems that the effects of CGA content on taste and disease resistance are in conflict.

CGA content in green coffee beans varies greatly between species (CLIFFORD & JARVIS, 1988; CLIFFORD *et al.*, 1989b; ANTHONY *et al.*, 1993; RAKOTOMALALA *et al.*, 1993). CGA content varies from 0.14 % dry matter basis (dmb) in *Coffea rhamnifolia* BRIDSON to 9.90 % dmb in *C. sessiliflora* BRIDSON. CQA ranges from 61 % in *C. sp Mouloundou* to 100 % CGA content in *C. farafanganensis* LEROY. DiCQA reaches 36 % CGA content, whereas FQA decreases to 12 % CGA content in *C. canephora*. If we consider the African continent, a CGA gradient exists between East-African and West-African species (ANTHONY *et al.*, 1993). The large range of CQA, diCQA, and FQA contents in the genus *Coffea* and their implications for organoleptic qualities and disease resistance emphasize the importance of CGA in breeding programs.

*C. pseudozanguebariae* BRIDSON (PSE), which has about 1.3 % dmb CGA (ANTHONY *et al.*, 1993) was hybridized with *C. liberica* var *dewevrei* (DEW), an anciently cultivated species from West-Africa with approximately 6.9 % dmb CGA (ANTHONY *et al.*, 1993). F1 and back-cross hybrids were obtained (LOUARN, 1982, 1992). The objectives of this paper were (1) to investigate the quantitative inheritance of CGA, and (2) to study the relations between the isomers.

## MATERIALS AND METHODS

### PLANT MATERIAL

Plant material was maintained at the Agricultural Station ORSTOM-IDEFOR (Man, Côte-d'Ivoire).

Parental species were investigated from the 1996 harvest. For each species, beans were separately harvested from seven trees, which were native to the Republic of Central Africa or Kenya for DEW and PSE, respectively.

The F1 generation included 7 hybrids from a controlled cross between genotype 8044 (PSE) and 5851 (DEW) parents. Biochemical analysis was performed on the 1995 harvest.

The second generation included 14 hybrids which consisted of 7 F1 x DEW (BCDEW) and 7 F1 x PSE (BCPSE). Twelve hybrids BCDEW were harvested in both 1995 and 1997 for the environmental effects study.

### SAMPLE PREPARATION

All coffee cherries were harvested at complete maturity and depulped using the wet processing method. Fifty green beans per tree were immediately frozen using liquid nitrogen to avoid CGA degradation by heat during the crushing. Two-minute crushing in a ball mill (Dangoumill) yielded a fine powder, which was divided into 6 samples, 3 to estimate water content and 3 for the extraction and analysis of CGA.

### EXTRACTION AND PURIFICATION

Extraction and purification of CGA were done using a fast and accurate method. Coefficients of variation are 3-5% for CQA, 1-2% for FQA and 5-7% for diCQA (KY *et al.*, 1997). Each sample was extracted in a 250 ml Erlenmeyer flask. Each flask contained about 250 mg of powder and 100 ml of methanol-water (70/30 v/v) mixture and 0.5% sodium bisulfite. The Erlenmeyer flasks were shaken over-night at 4°C in darkness on a stirring table at 125 rpm.

The organic extracts were directly treated with Carrez reagents (0.5 ml each). Solutions were then filtered through a 0.2µm filter and directly analyzed by HPLC.

### ANALYTICAL HPLC

Chromatography was carried out on a system consisting of two Waters Associates Model 510 pumping units, an automated sample injector (Waters 717 plus autosampler), a variable-wavelength UV detector (Waters 996 Photodiode Array Detector), a C<sub>18</sub> pre-column and a 250 mm x 4 mm Merck Superspher 100 RP 18 column, 5µm particle size. The elution program used two solvents, A and B. Solvent A was 2 mM phosphoric acid, pH 2.7, containing 5% methanol. Solvent B was methanol containing 5% 2 mM phosphoric acid, pH 3.9. These two mobile phases were filtered (0.2 µm), degassed and sonicated (Ney, 300 ultrasonik) before use. Sample or standard (10 µL) was analyzed at room temperature using the following elution program: A-B mixture (75/25 v/v) to pure solvent B in 45 min. of linear gradient. Flow rate was 0.8 mL/min. UV detection was at 325.2 nm, corresponding to the CGA maximum absorption.

The processing order of extracts was fully randomized for HPLC analysis. Every ten extracts, a 5-CQA standard (ref. D11.080-9, Aldrich-Chimie) at 100 mg/l was used to verify the stability of the HPLC measurements. Every 40 extracts, the C<sub>18</sub> pre-column and the 100 RP 18 column were washed overnight with a solution of pure methanol at a flow rate of 0.2 ml/min.

### CGA IDENTIFICATION AND QUANTIFICATION

Chlorogenic acid isomers were identified by means of their chromatograms (at 280 nm and 325 nm),

retention times, and UV spectrum (KY *et al.*, 1997) which were compared to RAKATOMALALA's results (1992). Identification of CQA isomers was confirmed by isomerization of 5-CQA (ref. D11.080-9, Aldrich-Chimie) following the method proposed by TRUGO and MACRAE (1984).

Quantification was achieved by peak-area measurement (PA) and by comparison with the standard 5-CQA. A 5-CQA calibration curve ( $C=2.768 \cdot 10^{-5} \times PA$ ) was generated using three replicate points at 25, 50, 75, 100, 125, and 150 mg/l. CGA isomer contents were expressed in per cent of dry matter basis (% dmb).

#### STATISTICAL ANALYSIS

All results were analyzed using Statistica software.

The relative importance of between-group (PSE, BCPSE, F1, BCDEW, and DEW) and within-group diversities was analyzed using the nested model of ANOVA with two factors: group (fixed effects) and plant nested in group (random effects).

Differences between harvest years in the 12 BCDEW hybrids were tested using a crossed model of ANOVA with two factors: hybrids (random effects) and harvest year (fixed effects).

Additivity was tested using linear fitting between expected values (PSE = 0, BCPSE = 0.25, F1 = 0.5, BCDEW = 0.75, DEW = 1) and observed data.

Relations between isomers were supposed to be allometric ( $y = a x^b$ ). Fitting was done after double transformation of variables [ $Y = \log(y)$  and  $X = \log(x)$ ] to obtain a linear model. When estimation of the b parameter was close to 1, a classic linear regression without transformation was applied to data. Figures represented relations for initial data, but  $R^2$  estimations are those obtained from transformed data. Within-group relations were studied using covariance analysis (within-group regression and parallelism tests).

#### RESULTS

Eight peaks were observed in DEW instead of the nine expected. Isomerization of pure 5-CQA led to 2 peaks for 3 isomers showing that 5- and 4-CQA peaks were in fact confounded (KY *et al.*, 1997).

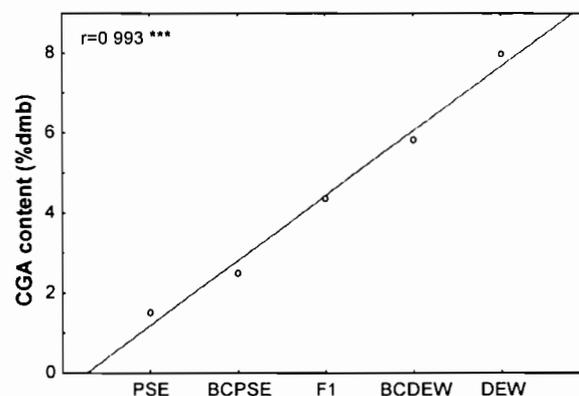
Within-group variation was detected for all studied traits (Table 1). Its relative importance, as regards the general variation, varied among traits from 4% (3-CQA) to 99% (FQA/CGA). Some trends can be highlighted: within-group variation was lower for CQA isomer contents than for FQA and diCQA isomer contents, and was lower for isomer contents than for ratios. Otherwise, within-group variation was related to species average (CGA, CQA, 4-&5-CQA) in some cases and was similar in all groups (diCQA, 4,5-diCQA) or group-dependent (3,5-diCQA) in other cases.

Some environmental effects (between years 1995 and 1997) were also studied. Between-tree differences were recorded again for all isomer contents (Table 2), but in particular, most traits showed both year and interaction effects. Some traits, mainly ratios, showed only interaction effects, whereas the 4,5-diCQA/diCQA ratio presented a year effect without interaction. The importance of environmental effects (years + interactions) varied between traits from 8.4% (3-FQA/FQA) to 57% (CQA/CGA). In general, minor isomers presented greater environmental effects.

#### QUANTITATIVE INHERITANCE OF TOTAL CHLOROGENIC ACID CONTENT (CGA)

Chlorogenic acid contents differed between parental species (Table 1): CGA content of green beans of DEW was 8.0 % dmb, i.e. about 5 times that of PSE. In F1 hybrids, CGA content was similar to that of the mid-parental species, and differed significantly from those of parental species. CGA content in BCPSE was intermediate between F1 and PSE averages, whereas CGA content in BCDEW was between F1 and DEW values. This suggests an additivity hypothesis for this trait, as can be verified in figure 1.

Fig 1 Regression analysis showing additivity of CGA content



#### Quantitative inheritance of caffeoylquinic acid content (CQA)

CQA represented 79% and 93% of CGA in DEW and PSE, respectively, and these values were significantly different (Table 1). Nevertheless, the quantitative inheritance of the ratio was not simple. Indeed, F1 and BCDEW showed a ratio close to that of DEW, but especially the BCPSE ratio (72%) was significantly lower than the DEW ratio. Conversely, quantitative inheritance of CQA content was simpler than for the ratio. CQA content, which was 4.5-fold higher in DEW than in PSE, showed an intermediate value between parental species in F1 hybrids suggesting additivity. This was confirmed by values observed in the two back-crosses (Table 3).

The contents of 4- and 5-CQA isomers varied from 1.37% dmb in PSE to 5.80% dmb in DEW and the

additivity hypothesis fitted the data well (Table 3). 4-&5-CQA content represented 91% of CQA in DEW and 98% in PSE, the two values being significantly different (Table 1). Quantitative inheritance of the ratio seems additive (Table 3), but discrepancies between observed and expected values were larger than for CGA and CQA. This is probably due to the relative importance of the within-group variation associated with the small number of trees used to estimate each average.

The 3-CQA content varied from 0.026 % dmb in PSE to 0.532% dmb in DEW. Its quantitative inheritance was not additive. In fact, only the square root of its content was additive (Table 3).

An inter-group and curvilinear relations was emphasized between 3-CQA and 4-&5-CQA contents (Fig. 2a). The allometric model ( $y = a x^b$ ) statistically accounted for 89% of variance and the b parameter was close to 2 ( $b = 1.803$ ) meaning that relation is quadratic. The curvilinearity explains why the 4-&5-CQA/CQA ratio varied between groups. Conversely, no relation was observed within groups ( $F_{1, 29} = 1.93$ ;  $P = 0.17$ )

Table 1 Chlorogenic acid contents Contents are expressed in % dmb. Table includes main group characteristics (2<sup>nd</sup> to 5<sup>th</sup> columns) with averages (upper number) and min-max (lower number), and F tests (6<sup>th</sup> & 7<sup>th</sup> columns) with test results (upper number) and factorial contributions (lower number in brackets) (\*\*\*) Very highly significant,  $P < 0.001$ ) Results of the Newmann & Keuls test are indexed with letters

Table 1 Chlorogenic acid contents Contents are expressed in % dmb. Table includes main group characteristics (2<sup>nd</sup> to 5<sup>th</sup> columns) with averages (upper number) and min-max (lower number), and F tests (6<sup>th</sup> & 7<sup>th</sup> columns) with test results (upper number) and factorial contributions (lower number in brackets) (\*\*\*) Very highly significant,  $P < 0.001$ ) Results of the Newmann & Keuls test are indexed with letters

	Averages and ranges					Between-group	Within-group
	DEW	BCDEW	F1	BCPSE	PSE	differences	differences
CGA	7.97 <sup>a</sup> 6.2-8.8	5.81 <sup>b</sup> 5.0-6.6	4.36 <sup>c</sup> 3.6-5.2	2.50 <sup>d</sup> 1.9-3.1	1.51 <sup>e</sup> 1.3-1.8	F <sub>4, 30</sub> 114 *** (93.9%)	F <sub>30, 70</sub> 41.2 *** (5.7%)
CQA	6.33 <sup>a</sup> 5.1-7.5	4.41 <sup>b</sup> 3.6-4.9	3.34 <sup>c</sup> 2.7-3.8	1.81 <sup>d</sup> 1.3-2.3	1.40 <sup>d</sup> 1.2-1.7	103 *** (93.3%)	53.9 *** (6.3%)
diCQA	1.18 <sup>a</sup> 0.9-1.4	0.97 <sup>b</sup> 0.8-1.2	0.67 <sup>c</sup> 0.5-0.8	0.45 <sup>d</sup> 0.2-0.9	0.04 <sup>e</sup> 0.02-0.06	54.2 *** (87.8%)	32.5 *** (11.2%)
FQA	0.465 <sup>a</sup> 0.24-0.82	0.431 <sup>a</sup> 0.13-0.90	0.344 <sup>a</sup> 0.20-0.61	0.236 <sup>ab</sup> 0.09-0.43	0.075 <sup>b</sup> 0.04-0.13	6.11 *** (41.9%)	178 *** (57.1%)
CQA/CGA	0.794 <sup>b</sup> 0.76-0.86	0.760 <sup>ab</sup> 0.71-0.81	0.767 <sup>ab</sup> 0.72-0.83	0.725 <sup>a</sup> 0.63-0.81	0.925 <sup>c</sup> 0.87-0.95	25.8 *** (77.7%)	126 *** (21.7%)
diCQA/CGA	0.149 <sup>a</sup> 0.11-0.18	0.167 <sup>a</sup> 0.14-0.19	0.155 <sup>a</sup> 0.11-0.19	0.176 <sup>a</sup> 0.10-0.32	0.025 <sup>b</sup> 0.01-0.04	17.2 *** (69.4%)	127 *** (29.9%)
FQA/CGA	0.058 <sup>a</sup> 0.03-0.08	0.072 <sup>a</sup> 0.02-0.14	0.078 <sup>a</sup> 0.06-0.12	0.099 <sup>a</sup> 0.05-0.15	0.049 <sup>a</sup> 0.03-0.09	1.95 NS	345 *** (99.22%)
4-&5-CQA	5.80 <sup>a</sup> 4.4-6.9	4.15 <sup>b</sup> 3.3-4.6	3.21 <sup>c</sup> 2.6-3.7	1.74 <sup>d</sup> 1.3-2.2	1.37 <sup>d</sup> 1.2-1.6	80.3 *** (91.7%)	63.7 *** (8.0%)
3-CQA	0.532 <sup>a</sup> 0.43-0.66	0.257 <sup>b</sup> 0.24-0.31	0.134 <sup>c</sup> 0.12-0.15	0.073 <sup>d</sup> 0.04-0.14	0.026 <sup>e</sup> 0.020-0.032	158 *** (95.5%)	38.8 *** (4.1%)
4-&5-CQA/CQA	0.914 <sup>a</sup> 0.87-0.93	0.941 <sup>b</sup> 0.91-0.95	0.959 <sup>c</sup> 0.95-0.97	0.960 <sup>c</sup> 0.94-0.98	0.981 <sup>d</sup> 0.97-0.99	22.9 *** (75.3%)	73.2 *** (23.7%)
4,5-diCQA	0.596 <sup>a</sup> 0.36-0.71	0.512 <sup>a</sup> 0.41-0.69	0.325 <sup>b</sup> 0.28-0.42	0.214 <sup>c</sup> 0.09-0.40	0.025 <sup>d</sup> 0.012-0.045	45.5 *** (85.4%)	22.0 *** (12.8%)
3,5-diCQA	0.321 <sup>a</sup> 0.24-0.39	0.254 <sup>b</sup> 0.24-0.29	0.227 <sup>b</sup> 0.14-0.30	0.166 <sup>b</sup> 0.08-0.40	0.007 <sup>c</sup> 0.002-0.015	27.3 *** (78.3%)	42.1 *** (20.3%)
3,4-diCQA	0.262 <sup>a</sup> 0.22-0.30	0.204 <sup>b</sup> 0.15-0.28	0.122 <sup>c</sup> 0.08-0.17	0.071 <sup>d</sup> 0.02-0.14	0.005 <sup>e</sup> 0.004-0.007	60.6 *** (88.6%)	20.2 *** (9.8%)
4,5-diCQA/diCQA	0.501 <sup>b</sup> 0.40-0.55	0.527 <sup>b</sup> 0.43-0.59	0.488 <sup>b</sup> 0.58-0.74	0.481 <sup>b</sup> 0.41-0.54	0.657 <sup>a</sup> 0.58-0.74	12.6 *** (52.8%)	4.15 *** (24.2%)
3,5-diCQA/diCQA	0.273 <sup>b</sup> 0.25-0.31	0.264 <sup>b</sup> 0.25-0.28	0.332 <sup>a</sup> 0.27-0.38	0.371 <sup>a</sup> 0.27-0.48	0.184 <sup>c</sup> 0.08-0.29	14.5 *** (60.2%)	7.18 *** (26.8%)
3,4-diCQA/diCQA	0.225 <sup>a</sup> 0.19-0.28	0.209 <sup>ab</sup> 0.18-0.29	0.180 <sup>ab</sup> 0.15-0.21	0.148 <sup>b</sup> 0.12-0.20	0.159 <sup>b</sup> 0.06-0.29	4.09 ** (28.3%)	7.02 *** (60.2%)
5-FQA	0.382 <sup>a</sup> 0.18-0.66	0.296 <sup>a</sup> 0.05-0.63	0.253 <sup>a</sup> 0.17-0.48	0.200 <sup>ab</sup> 0.05-0.35	0.064 <sup>b</sup> 0.04-0.12	5.14 ** (37.0%)	271 *** (62.3%)
4-FQA	0.049 <sup>a</sup> 0.02-0.10	0.115 <sup>b</sup> 0.05-0.23	0.091 <sup>b</sup> 0.04-0.13	0.036 <sup>a</sup> 0.01-0.05	0.011 <sup>a</sup> 0.007-0.016	11.5 *** (67.3%)	19.6 *** (32.7%)
3-FQA	0.036 <sup>a</sup> 0.02-0.06	0.020 <sup>b</sup> 0.0-0.041	0.000 <sup>c</sup> 0.0-0.0	0.000 <sup>c</sup> 0.0-0.0	0.000 <sup>c</sup> 0.0-0.0	18.7 *** (71.3%)	89.6 *** (27.7%)
5-FQA/FQA	0.818 <sup>a</sup> 0.76-0.86	0.662 <sup>b</sup> 0.41-0.81	0.734 <sup>ab</sup> 0.62-0.82	0.810 <sup>a</sup> 0.77-0.91	0.849 <sup>a</sup> 0.76-0.90	5.27 ** (37.2%)	70.0 *** (60.2%)
4-FQA/FQA	0.106 <sup>a</sup> 0.07-0.13	0.298 <sup>b</sup> 0.18-0.59	0.266 <sup>b</sup> 0.18-0.38	0.190 <sup>ab</sup> 0.09-0.23	0.151 <sup>a</sup> 0.10-0.24	5.49 ** (38.4%)	72.5 *** (59.1%)
3-FQA/FQA	0.076 <sup>a</sup> 0.06-0.13	0.040 <sup>b</sup> 0.00-0.07	0.000 <sup>c</sup> 0.0-0.0	0.000 <sup>c</sup> 0.0-0.0	0.000 <sup>c</sup> 0.0-0.0	42.5 *** (84.4%)	20.6 *** (13.5%)

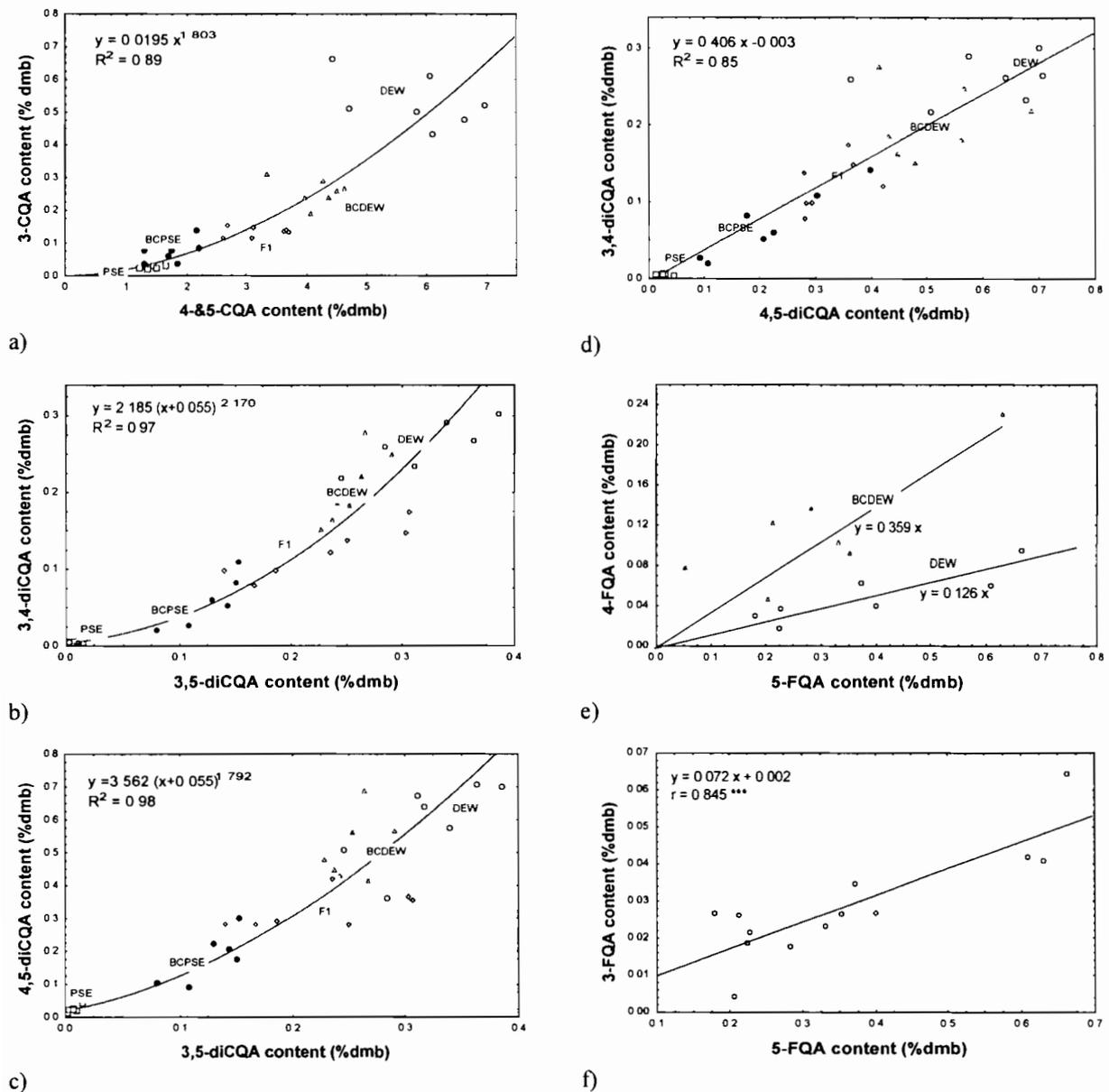


Fig. 2. Relations between isomer contents within CGA classes

### Quantitative inheritance of dicaffeoylquinic acid content (diCQA)

DiCQA represented 15% and 2.5% of CGA in green beans of DEW and PSE, respectively, these values being significantly different (Table 1). Quantitative inheritance of the ratio was not additive: averages of BCDEW, F1, and BCPSE hybrids did not differ from the DEW average. Note the presence of a BCPSE hybrid for which diCQA represented up to 32% of CGA. No explanation can be given for this outlier: for all other chlorogenic acids, it gave values expected for a BCPSE hybrid. Quantitative inheritance of diCQA content was simpler than that of the ratio. diCQA content was 29-fold higher in DEW, and

F1 hybrids showed an intermediate value between parental species. Additivity was confirmed by values observed in the two back-crosses (Table 3).

The 4,5-diCQA content varied from 0.025 in PSE to 0.596 in DEW (Table 1). F1 hybrids showed an average intermediate between parental species and additivity was confirmed in back-crosses (Table 3). These contents represented 50% and 66% of diCQA content in DEW and PSE, respectively, and these ratios were significantly different (Table 1). Hybrid groups (BCDEW, BCPSE, and F1) did not differ from DEW for the 4,5-diCQA proportion.

The 3,5-diCQA content was 46-fold higher in DEW than in PSE (Table 1). Additivity can be assu-

med, but discrepancies between expected and observed values were not negligible. By contrast, additivity became evident when data were transformed using  $y = x^2$  (Table 3). This isomer represented 27% and 18% of diCQA content in DEW and PSE, respectively, and these ratios were significantly different (Table 1). Genetic behavior of the 3,5-diCQA ratio was complex: BCDEW did not differ from DEW, whereas F1 and BCPSE showed the highest values (35%).

Table 2 Quantitative effects of hybrid, harvest year, and interaction on chlorogenic acid contents and ratios. Factorial contributions are in brackets. NS Not significant ( $P > 0.05$ ), \* Significant ( $0.01 < P \leq 0.05$ ), \*\* Highly significant ( $0.001 < P \leq 0.01$ ), \*\*\* Very highly significant ( $P < 0.001$ )

	Hybrid	Year	Interaction
	df=11, 48	df=1, 11	df=11, 48
CGA	48.4 *** (77.7 %)	0.25 NS	7.80 *** (22.3 %)
CQA	71.4 *** (77.8 %)	1.49 NS	11.0 *** (22.2 %)
diCQA	23.5 *** (65.3 %)	24.7 *** (26.9 %)	2.35 * (7.8 %)
FQA	121 *** (71.3 %)	20.8 *** (18.6 %)	9.48 *** (22.0 %)
CQA/CGA	82.2 *** (42.7 %)	66.4 *** (49.3 %)	8.63 *** (8.0 %)
diCQA/CGA	56.3 *** (73.2 %)	54.0 *** (24.0 %)	2.06 * (2.8 %)
FQA/CGA	166 *** (68.2 %)	15.3 *** (17.8 %)	18.0 *** (14.1 %)
5-& 4-CQA	77.8 *** (77.7 %)	1.14 NS	12.0 *** (22.3 %)
3-CQA	21.5 *** (61.3 %)	8.46 * (17.1 %)	4.60 *** (21.6 %)
5-& 4-CQA/CQA	55.6 *** (84.3 %)	4.63 NS	6.07 *** (15.7 %)
4,5diCQA	21.6 *** (74.4 %)	13.9 ** (17.0 %)	2.18 * (8.6 %)
3,5diCQA	22.0 *** (62.0 %)	21.0 *** (27.4 %)	2.78 ** (10.5 %)
3,4diCQA	13.9 *** (44.8 %)	29.0 *** (43.5 %)	2.69 ** (11.7 %)
4,5diCQA/diCQA	4.26 *** (76.9 %)	4.90 * (23.1 %)	1.50 NS -
3,5diCQA/diCQA	5.54 *** (59.5 %)	0.026 NS	2.55 * (40.5 %)
3,4diCQA/diCQA	8.50 *** (52.1 %)	6.27 * (18.9 %)	3.09 ** (29.1 %)
5-FQA	121 *** (76.3 %)	20.6 *** (15.5 %)	7.45 *** (8.2 %)
4-FQA	65.1 *** (62.4 %)	4.49 * (8.9 %)	15.8 *** (28.7 %)
3-FQA	261 *** (73.4 %)	12.2 ** (13.1 %)	24.9 *** (13.5 %)
5-FQA/FQA	173 *** (67.3 %)	0.074 NS	42.7 *** (32.7 %)
4-FQA/FQA	232 *** (69.7 %)	0.001 NS	51.2 *** (30.3 %)
3-FQA/FQA	205 *** (91.6 %)	4.58 NS	10.4 *** (8.4 %)

The 3,4-diCQA content was 52-fold higher in DEW than in PSE (Table 1). Table 3 shows clearly that additivity can be fully accepted for this isomer content. Proportions of 3,4-diCQA in diCQA were similar to, but slightly lower than proportions of 3,5-diCQA. As for other ratios, quantitative inheritance was not simple: the additivity hypothesis can be accep-

ted for F1 and BCDEW hybrids, and rejected for BCPSE hybrids.

Table 3 Biochemical traits showing additivity either without or with data transformation. Non additive traits are not presented. Correlation coefficient (r) and its significance from regression analysis as in figure 1. \*\*: Highly significant ( $0.001 < P \leq 0.01$ ), \*\*\* Very highly significant ( $P < 0.001$ )

Biochemical trait	Transformation	r	Significance
CQA	No transformation	0.981	**
4-&5-CQA/CQA	No transformation	0.965	**
4-&5-CQA	No transformation	0.984	**
3-CQA	$y = \sqrt{x}$	0.984	**
DiCQA	No transformation	0.993	***
4,5-diCQA	No transformation	0.993	***
3,5-diCQA	$y = x^2$	0.998	***
3,4-diCQA	No transformation	0.998	***
FQA	$y = x^2$	0.994	***
5-FQA	$y = x^2$	0.984	**

An inter-group and linear relation was noted between 3,4-diCQA and 3,5-diCQA contents (Fig. 2b). The allometric model [ $y = a(x+c)^b$ ] statistically accounted for 97% of variance and the b parameter was close to 2 ( $b = 2.17$ ) meaning that the relation is quadratic. Parameter c was necessary to linearize the relation using logarithmic transformation and means that 3,4-diCQA could theoretically be present even if 3,5-diCQA is absent. A within-group relation was also detected ( $F_{1,28} = 26.3$ ;  $P = 0.00002$ ) with different slopes (parallelism test:  $F_{4,24} = 9.33$ ;  $P = 0.0001$ ). In fact, slopes of within-group relations seemed close to the tangent to the curve, i.e. increased linearly with the 3,5-diCQA content.

An inter-group and curvilinear relation was also recorded between 4,5-diCQA and 3,5-diCQA contents (Fig. 2c). The allometric model [ $y = a(x+c)^b$ ] statistically explained 98% of variance and the b parameter was again close to 2 ( $b = 1.79$ ). Note that the c parameter had the same value as in the 3,4-diCQA-3,5-diCQA relation. A within-group relation was again noted ( $F_{1,28} = 14.1$ ;  $P = 0.0008$ ), with different slopes ( $F_{4,24} = 3.48$ ;  $P = 0.022$ ).

Another inter-group, but linear, relation was observed between 3,4-diCQA and 4,5-diCQA contents (Fig. 2d). Linearity can be easily explained when we consider the two other relations as quadratic ( $b=2$ ). Again, a within-group relation was noted ( $F_{1,28} = 12.0$ ;  $P = 0.0017$ ), but with a common slope ( $F_{4,24} = 1.13$ ;  $P = 0.37$ ).

Quadratic relations explain why the three diCQA ratios differed between species and are consistent with our results on 3,5-diCQA additivity after transformation.

#### Quantitative inheritance of feruloylquinic acid content (FQA)

FQA represented 7% of CGA on average, and no significant difference was observed between groups for the ratio (Table 1). By contrast, FQA content differed greatly between groups, especially between

DEW and PSE. As for 3,5-diCQA, additivity can be recorded when we take into account the transformed values using “ $y = x^2$ ” (Table 3).

The 5-FQA isomer accounted for the same proportion of FQA (83%) in the two parental species (Table 1). As expected, F1 hybrids and BCPSE hybrids showed the same ratio as parental species, but BCDEW hybrids gave a significantly lower ratio. By contrast with the ratio, 5-FQA content varied greatly between species, with PSE presenting a 6-fold lower ratio (Table 1). As for CQA, additivity was accepted for 5-FQA (Table 3).

The proportion of 4-FQA in FQA did not differ significantly between parental species (Table 1). Nevertheless, the ratio increased linearly from PSE (15%) to BCDEW (30%). More surprising, the 4-FQA content showed a genetic behavior close to the ratio: content increased linearly from PSE to BCDEW, but DEW had a twofold lower content than BCDEW.

3-FQA isomer was a minor isomer. It was absent in PSE, BCPSE and F1, and represented only 7.6% of FQA, i.e. 0.4% of CGA in DEW. Nevertheless, this isomer is genetically interesting: indeed, results could be interpreted as a dominance of the absence of 3-FQA. In addition, when we assume 50% of heterozygotes in BCDEW, we expect a BCDEW average intermediate between F1 and DEW, as observed (Table 1).

A relation between 5-FQA and 4-FQA was detected in four groups. It was linear, but weak (only 37% of the 4-FQA content variation was statistically explained by 5-FQA content variation). In fact, two situations can be defined: 1) absence of a relation within PSE, BCPSE, and F1; and 2) presence of a relation for BCDEW and DEW, i.e. in groups with highest 5-FQA content. Nevertheless, the relation was different within DEW and BCDEW: the slope was 3-fold higher in BCDEW than in DEW (Fig 2e).

A linear relation was also recorded between 3-FQA and 5-FQA for trees with 3-FQA, i.e. in DEW and in BCDEW. The two within-group relations did not differ in slope ( $F_{1,9} = 0.11$ ;  $P=0.75$ ) or intercept ( $F_{1,10} = 3.00$ ;  $P=0.11$ ). A common relation was then estimated (Fig. 2f), for which the intercept can be considered as equal to 0 ( $t = 0.484$ ;  $df=11$ ;  $P=0.637$ ).

#### RELATIONS BETWEEN CGA CLASSES

An almost linear between-group relation existed between diCQA and CQA content (Fig. 3a), but PSE was unrelated. The allometric model [ $y = a(x)^b$ ] statistically explained 81% of variance and the  $b$  parameter was again close to 1 ( $b = 0.95$ ) meaning that linear model could be appropriate. A within-group relation was noted ( $F_{1,22} = 4.63$ ;  $P = 0.043$ ), with a common slope ( $F_{3,19} = 0.29$ ;  $P = 0.830$ ).

Another allometric, but weaker, relation was seen between groups for FQA and CQA contents (Fig. 3b). As before, CQA content seems to be over-represented in PSE. Two fittings were done: 1) with BCPSE, F1, BCDEW, and DEW groups using  $y = a(x)^b$ , and 2) with all groups using  $y = a(x+c)^b$ . The two fittings gave similar values of  $b$  (0.437 and 0.501, respectively), close to 0.5. The relation seems to be of the square root type. This is consistent with the fact that FQA must be transformed using  $X^2$  to become additive, while CQA is additive without transformation. Note that CQA additivity would be better if we consider that CQA is over-represented in PSE (Fig 1).

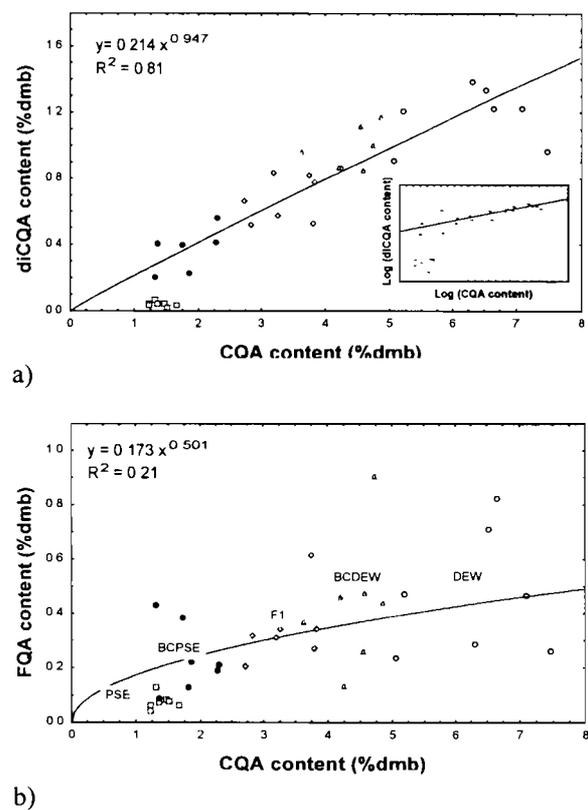


Fig 3 Relations between CGA class contents

#### DISCUSSION

An overview of chlorogenic acid biosynthetic pathways is necessary to interpret our results. In coffee trees, all biosynthetic pathways of CGA formation occur in leaves (COLONNA, 1986). CGA is derived from phenylalanine (VILLEGAAS & KOJIMA, 1986). The caffeic acid unit, which exists as caffeoyl-D-glucose, reacts with the quinate unit to give the 5-CQA isomer. The enzyme catalyzing the reaction is the quinate hydroxycinnamoyl transferase (CQT). The isomeric monosubstituted quinic acids, e.g. 3-CQA and 4-CQA, are produced by enzymatic transesterification of 5-CQA.

The 3,5-diCQA was formed by caffeate substitution of 5-CQA in the 3-position by a chlorogenic acid caffeoyl transferase. The other two disubstituted quinate, e.g. 3,4-diCQA and 4,5-diCQA, were also formed by enzymatic trans-esterification of 3,5-diCQA (FRIEDMAN, 1997). So all diCQA isomers derived from 5-CQA.

By contrast, FQA isomer formation was independent of 5-CQA synthesis. A quinate transferase other than CQT catalyzes the reaction of feruloyl-CoA with quinate. The 5-FQA is first formed, whereas the 3- and 4-FQA were derived from trans-esterification (ULBRICH & ZENK, 1979). In CQA and FQA, the migration of the cinnamoyl group is an intramolecular process which occurs in the sequence  $5 \leftrightarrow 4$  and  $5 \leftrightarrow 3$  (HANSON, 1965). Useful descriptions can be found in MORIGUSHI *et al.* (1988) and TANAKA and KOJIMA (1991).

CGA are then exported to beans where their accumulation is maximum when fruits are green. During fruit maturation, the CGA level decreases (PETNGA, 1986). Thus, from their biosynthesis in leaves to their accumulation in beans at complete maturity, a lot of processing occurs and leads to quantitative variation in their contents.

#### RELATIONS BETWEEN ISOMERS WITHIN CGA CLASSES

Relations between isomers within CGA classes seems to result from isomerization processes. In the CQA and diCQA classes, results agreed well with known pathways. The relations between 4-FQA and 5-FQA in the BCDEW and DEW groups is also due to the isomerization process. Its absence in PSE, BCPSE and F1 groups could be explained by a lack of accuracy due to the low content of 4-FQA in such hybrids. This is also true for the relation between 3-FQA and 5-FQA (3-FQA was not detected in PSE, BCPSE, and F1 groups).

Changes in relations are more surprising: some relations were linear whereas others were curvilinear. Differences appeared even for isomers belonging to the same CGA class. As an example, in the case of the diCQA class, 4,5-diCQA and 3,4-diCQA were curvilinearly related to 3,5-diCQA, while 3,4-diCQA was linearly related to 4,5-diCQA. A linear relation is expected from enzymatic reactions at equilibrium: the product/ reactant ratio was indeed constant. Deviation from linearity could be due to the involvement of 3,5-diCQA in other specific pathways. Indeed, 3,5-diCQA should react with caffeine to form a complex (HORMAN & VIANI, 1971; KAPPELER *et al.*, 1987; BAUMANN & ROHRIG, 1989). The curvilinear relation between 3-CQA and 4-&5-CQA could be similarly explained. Indeed, 5-CQA, which contributes mainly to the 4-&5-CQA peak (CLIFFORD, 1985b), is known to complex with caffeine (BAUMANN *et al.*, 1993). By contrast, the linear relation between 4,5-diCQA and

3,4-diCQA suggests that 4,5-diCQA and 3,4-diCQA would not complex with caffeine or would complex at a much lower rate.

The slope difference observed for the relation between 4-FQA and 5-FQA in BCDEW and DEW is strange. Because the intersect of straight lines was nil, the slope can be interpreted as a 4-FQA/5-FQA ratio, which would be relatively constant within groups (as expected), but would differ greatly between BCDEW and DEW. Nevertheless, parental species similarity in 4-FQA/5-FQA ratios complicates the interpretation problems of such results, for which further studies appear necessary.

#### RELATIONS BETWEEN CGA CLASSES

The addition of 3 isomer contents in order to obtain CQA content, or diCQA content, or FQA content has a biological meaning when biosynthesis pathways are taken into account. For example, CQA content in fact reflects the quantity of 5-CQA synthesized before isomerization.

A relation between CQA and diCQA contents is again expected from our knowledge of biosynthesis: diCQA derives directly from CQA. Conversely the relation between FQA content and CQA content is indirect: FQA does not derive from CQA, but 5-CQA and 5-FQA synthese use a common substrate, the quinate unit. Indirect effects also explain why the FQA-CQA relation was weaker its diCQA-CQA counterpart. Similarly, absence of a relation between FQA and diCQA contents accords with the larger distance between their biosynthesis pathways.

The two between-class relations were not linear. Concavity would mean that diCQA and FQA are implicated in biosynthesis pathways other than CQA. FQA is known to be involved in lignin biosynthesis pathways (HAHLBROCK & GRISEBACH, 1979) and diCQA plays a role in caffeic acid detoxification (COLONNA, 1979).

#### GENETIC CONTROL OF THE 3-FQA ISOMER CONTENT

The absence of 3-FQA in PSE, BCPSE and F1 could be an artefact due to a lack of accuracy for individuals with very low contents. Nevertheless, 3-FQA was also absent from BCDEW hybrids for which 5-FQA content was high enough to suggest a detectable 3-FQA content.

The 3-FQA isomer content seems to be controlled by one major gene with a dominant allele leading to the absence of 3-FQA. Parental species would be FF and ff for PSE and DEW, respectively, whereas F1 hybrids would be Ff. All BCPSE would have the allele F leading to absence of 3-FQA. By contrast, BCDEW hybrids presented would be Ff or ff in a 1:1 proportion, explaining why the average 3-FQA con-

tent of BCDEW (0.020 %dmb) was intermediate between F1 (0.00 %dmb) and DEW (0.36 %dmb).

Physiologically, a compound may fail to accumulate because of the: 1) absence of the enzyme responsible for its biosynthesis, 2) absence of transport from leaves (biosynthesis tissues) to beans (storage tissues), and 3) presence of an enzyme responsible for its transformation before transport i.e. in leaves or later in fruits. Nevertheless, the dominance of absence favors the third hypothesis. Further studies on leaves and fruits during maturation would provide some explanations.

#### ADDITIVITY

Genetic inheritance of CGA isomer contents showed three types of genetic behavior. The first case concerned traits showing additivity without data transformation. This was the case for CGA, CQA, diCQA, 5-CQA, 4,5-diCQA, and 3,4-diCQA.

The second case concerned traits for which the additive model was present after data transformation. Results are consistent with previous interpretations on relations between isomers or between classes. For example, 5-CQA being additive and 3-CQA being non linearly related to 5-CQA, the inverse transformation to obtain additivity for 3-CQA is expected to be efficient. We can verify that additivity after data transformation concerns isomers (3,4-diCQA, 4,5-diCQA, 3-CQA) and classes (FQA) which are non linearly related to true additive traits. An additive model after data transformation has already been encountered in the same interspecific cross for caffeine (BARRE *et al.*, 1998). In this example, additivity of the square root of caffeine content has been observed.

The third case concerned the contents for which additivity was not or partly observed. For 3-FQA, which was discussed above, additivity was verified for F1, BCDEW and DEW. The situation was more complex for 4-FQA ( $PSE = BCPSE = DEW < F1 = BCDEW$ ) and cannot be easily interpreted.

Lastly, note the absence of additivity for most ratios except 4-&5-CQA/CQA. Here too, curvilinear relations explain well the absence of additivity for most ratios.

#### ENVIRONMENTAL EFFECTS AND INTERACTIONS

Genetic effects represented 78% of variation in the best cases showing the relative importance of environmental effects and interactions. Environmental effects and interactions would have probably been more important in crosses of less genetically differentiated species. These effects include inter-annual climatic variation during bean maturation, but also maturity variation at the harvest, and bean-drying conditions after harvest. In particular, the greater importance of inter-actions for diCQA could be explained by their

degradation during bean drying, which would change between years. Observed interactions constitute an example of « genotype x environment » interactions. This implies that CGA are sensitive to all types of environmental changes, and consequently, interactions should exist when genotypes are growing in different regions.

#### CONCLUSIONS AND PROSPECTS

The consequences for breeding are various: 1) additivity implies the possibility of choosing parents to obtain an expected average in progenies; 2) genetic diversity is the main source of variation allowing clonal selection; 3) the "hybrid-year" interaction was not important (22% of effects) meaning that observations should be recorded over several years to obtain a good estimate of the phenotypic average; 4) the presence of "genotype x environment" interactions should lead to a multilocal selection for coffee tree varieties for taste quality improvement. However, choosing hybrids with low CGA content (i.e. having good cup taste) would conflict with the resistance role played by high CGA content.

Biosynthetically, the data on isomers and between-class relations confirmed the presently accepted pathways both in 1) interdependence of each CGA class and 2) formation of each isomer within each class. Genetically, the results showed that the 3-FQA isomer content was a trait governed by a single major gene, dominant for 3-FQA absence. Other isomers showed different additive patterns. This study was the first using interspecific progenies to explain both genetic and physiological aspects of the CGA trait for a particular aim: comprehension and improvement of coffee cup taste or disease resistance.

As CGA is a complex trait implicated in aroma, physiology and genetics, further studies are needed. Study of relations between CGA and other interesting traits such as morphological traits (leaf area, bean weight...), phenological traits (fructification cycle period...) and biochemical traits (sucrose, trigonelline, free amino acid contents...) would lead to better planning of diploid coffee tree breeding programs. Molecular tools would be helpful in screening for markers near interesting quantitative trait loci (QTL) to develop marker-assisted selection of plantlets.

#### REFERENCES

- AMORIN, H.V., TEIXEIRA, A.A., GUERCIO, M.A., CRUZ, V.F. & MALAVOLTA, E. 1974. Chemistry of Brazilian green coffee and the quality of the beverage II phenolic compounds. *Turrialba* 24: 217-221.
- ANTHONY, F., CLIFFORD, M.N. & NOIROT, M. 1993. Biochemical diversity in the genus *Coffea* L.: chlorogenic acids, caffeine and mozambioside contents. *Gen. Res. Crop Evol.* 40: 61-70.

- BALYAYA, K.J. & CLIFFORD, M.N. 1995. Individual chlorogenic acids and caffeine contents in commercial grades of wet and dry processed Indian green *Robusta* coffee. *J. Food Sci. Technol.* 32: 104-108.
- BARRE, P., AKKAFU, S., LOUARN, J., CHARRIER, A., HAMON, S. & NOIROT, M. 1998. Inheritance of caffeine and heteroside contents in an interspecific cross between a cultivated coffee species *Coffea liberica* var *dewevrei* and a wild species caffeine free *C. pseudozanguebariae*. *Theor. Appl. Genet.* 96: 306-311.
- BAUMANN, T.W. & RÖHRIG, L. 1989. Formation and intracellular accumulation of caffeine and chlorogenic acid in suspension cultures of *Coffea arabica*. *Phytochemistry* 28: 2667-2669.
- BAUMANN, T.W., MOSLI, S.S., SCHULTHESS, B.H. & AEARTS, R.J. 1993. Interdependence of caffeine and chlorogenic acid (5-CQA) metabolism in coffee. *Proc. Int. Congr. ASIC* 15: 134-140.
- CLIFFORD, M.N. 1985a. Chemical and physical aspects of green coffee and coffee products. In: *Coffee, botany, biochemistry and production of beans and beverage*. Clifford & Wilson (eds), Avi publishing company, Westport, Connecticut, pp 305-374.
- CLIFFORD, M.N. (1985b) Chlorogenic acids. In: *Coffee I Chemistry*. Clarke & Macrae (eds), Elsevier Applied Science Publisher, pp 153-202.
- CLIFFORD, M.N. & JARVIS, T. 1988. The chlorogenic acids content of green robusta coffee beans as possible index of geographic origin. *Food Chem.* 29: 291-298.
- CLIFFORD, M.N. & OHIKPEHAI, O. 1982. Coffee as-tringency. *Royal Society of Chemistry Special Techniques group Meeting*. Metal Box Ltd., Wantage.
- CLIFFORD, M.N. & STANFORTH, P.S. 1977. A critical comparison of six spectrophotometric methods for measuring chlorogenic acids in green coffee beans. *Proc. Int. Congr. ASIC* 8: 109-113.
- CLIFFORD, M.N., KELLARD, B. & AH-SING, E. 1989a. Caffeoyl-tyrosine from green robusta coffee beans. *Phytochemistry* 28: 1989-1990.
- CLIFFORD, M.N., WILLIAMS, T. & BRIDSON, T. 1989b. Chlorogenic acids and caffeine as possible taxonomic criteria in *Coffea* and *Psilanthus*. *Phytochemistry* 28: 829-838.
- COLONNA, J.P. 1979. *L'acide chlorogénique et les depsides de divers caféiers africains et malgaches: leur participation au métabolisme et leur signification biologique*. Ph.D. Thesis, Université Paul Sabatier, Toulouse.
- COLONNA, J.P. 1986. Biosynthèse et renouvellement de l'acide chlorogénique et des depsides voisins dans le genre *Coffea*. *Café Cacao Thé* 30: 83-92 & 247-258.
- CORREIA, M.N.G., LEITAO, M.C.A. & CLIFFORD, M.N. 1995. Caffeoyl-tyrosine and Angola II as characteristic markers for Angolan robusta coffees. *Food Chem.* 53: 309-313.
- DE MARIA, C.A.B., TRUGO, L.C., MOREIRA, R.F.A. & WERNECK, C.C. 1994. Composition of green coffee fractions and their contribution to the volatile profile formed during roasting. *Food Chem.* 50: 141-145.
- ECHANDI, E. & FERNANDEZ, C.E. 1962. Relacion entre el contenido de acido chlorogenico y la resistencia a la Haga macana o cancer de los cafetos causados por *Ceratocystis fimbriata*. *Turrialba* 12: 87-90.
- FELDMAN, J.R., RYDER, W.S. & KUNG, J.T. 1969. Importance of nonvolatile compounds to the flavour of coffee. *J. Agric. Food Chem.* 17: 733-739.
- FRIEDMAN, M. 1997. Chemistry, biochemistry, and dietary role of potato polyphenols. A review. *J. Agric. Food Chem.* 45: 1523-1540.
- GUYOT, B., CHRISTOPHE, T. & GUEULE, D. 1997. Effet inhibiteur de l'acide caféyl 5-quinique sur la formation des pyrazines dans les réactions de Mailard. *Proc. Int. Congr. ASIC* 17: 169-175.
- GUYOT, B., PETNGA, E. & VINCENT, J.C. 1988. Analyse qualitative d'un café *Coffea canephora* var robusta en fonction de la maturité. Partie I. Evolution des caractéristiques physiques, chimiques et organoleptiques. *Café Cacao Thé* 32: 229-242.
- GUYOT, B., GUEULE, D., MANEZ, J.C., PERRIOT, J.J., GIRON, J. & VILLAIN, L. 1996. Influence de l'altitude et de l'ombrage sur la qualité des cafés arabicas. *Plantation Recherche Développement* 3: 272-280.
- HAHLBROCK, K. & GRISEBACH, H. 1979. Enzymatic controls in the biosynthesis of lignin and flavonoids. *Ann. Rev. Plant Physiol.* 30: 105-130.
- HANSON, K.R. 1965. Chlorogenic acid biosynthesis. Chemical synthesis and properties of the mono-*O*-cinnamoylquinic acids. *Biochemistry* 4: 2719-2730.
- HORMAN, I. & VIANI, R. 1971. The caffeine-chlorogenate complex of coffee NMR study. *Proc. Int. Congr. ASIC* 5: 102-111.
- KAPPELER, A.W., BAUMANN, T.W. & GREUTERT, P.D. 1987. Complexation of purine alkaloids in the coffee plant. *Proc. Int. Congr. ASIC* 12: 247-253.
- KY, C.L., NOIROT, M. & HAMON, S. 1997. Comparison of five purification methods for chlorogenic acids in green coffee beans (*Coffea* sp.). *J. Agric. Food Chem.* 45: 786-790.
- LELOUP, V., LOUVRIER, A. & LIARDON, R. 1995. Degradation mechanisms of chlorogenic acids during roasting. *Proc. Int. Congr. ASIC* 16: 192-198.
- LENTNER, C. & DEATHERAGE, F.E. 1959. Organic acids in coffee in relation to the degree of roast. *Food res.* 24: 483-492.

- LOUARN, J. 1982. Bilan des hybridations interspécifiques entre caféiers africains diploïdes en collection en Côte-d'Ivoire. *Proc. Int. Congr. ASIC* 10: 375-383.
- LOUARN, J. 1992. *La fertilité des hybridations interspécifiques et les relations génomiques entre caféiers diploïdes d'origine africaine (genre Coffea L. sous genre coffea)*. PhD Thesis, ORSTOM, Paris, 200 pp.
- MAIER, H.G. 1987. The acids of coffee. *Proc. Int. Congr. ASIC* 12: 229-237.
- MORIGUSHI, T., VILLEGAS, R.J.A., KONDO, T. & KOJIMA, M. 1988. Isolation of 1-O-*trans-p*-coumaroyl- $\beta$ -D-glucopyranose from sweet potato roots and examination of its role in chlorogenic acid biosynthesis. *Plant Cell Physiol.* 29: 1221-1226.
- MORISHITA, H., IWAHASHI, H. & KIDO, R. 1989. Chlorogenic acids composition of green and roasted coffee beans. *Bull. Fac. Edu. Wakayama Univ. Nat. Sci.* 38: 33-39.
- OHIKPEHAI, O., BRUMEN, G. & CLIFFORD, M.N. 1982. The chlorogenic acids content of some peculiar green coffee beans and the implications for beverage quality. *Proc. Int. Congr. ASIC* 10: 177-185.
- PETNGA, E. 1986. *Application de l'analyse multidimensionnelle à la classification qualitative du café Robusta en fonction de la maturité*. Ph.D. Thesis, Université des Sciences et Techniques du Languedoc.
- RAKOTOMALALA, J.J.R. 1992. *Diversité biochimique des caféiers, particularités des caféiers sauvages de la région malgache*. Ph.D. Thesis, Université des Sciences et Techniques du Languedoc.
- RAKOTOMALALA, J.J.R., CROS, E., CHARRIER, A., ANTHONY, A. & NOIROT, M. 1993. Marqueur biochimique de la diversité des caféiers. *Proc. Int. Congr. ASIC* 15: 47-55.
- STREULI, H. 1970. Alkaloidhaltige genussmittel, gewürze, kochsalz. In: *Kaffee*. Schormüller (ed), Springer-Verlag, Berlin, 55p.
- TANAKA, M. & KOJIMA, M. 1991. Purification and characterization of *p*-coumaroyl-D-glucose hydroxylase of sweet potato (*Ipomoea batatas*) roots. *Arch Biochem. Biophys.* 284: 151-157.
- TRUGO, L.C. & MACRAE, R. 1984. A study of the effect of roasting on the chlorogenic acid composition of coffee using HPLC. *Food Chem.* 15: 219-227.
- ULBRICH, B. & ZENK, M.H. 1979. Partial purification and properties of hydroxycinnamoyl-CoA: quinate hydroxycinnamoyl transferase from higher plants. *Phytochemistry* 18: 929-933.
- VILLEGAAS, R.J.A. & KOJIMA, M. 1986. Purification and characterization of hydroxycinnamoyl D-glucose. *J. Biol. Chem.* 261: 8729-8733.
- ZULUAGA, J.V., VALENCIA, G.A. & GONZALES, J. 1971. Contribucion al estudio de la naturaleza de la resistencia del cafeto a *Ceratocystis fimbriata* (Ell. Halst.) Hunt. *Cenicafé* 22: 43-70.

**Annexe C**

---

**INHERITANCE OF COFFEE BEAN SUCROSE CONTENT  
IN THE INTERSPECIFIC CROSS:  
*COFFEA PSEUDOZANGUEBARIAE* X *COFFEA LIBERICA DEWEVREI***

KY C.L.<sup>1</sup>, DOULBEAU S.<sup>1</sup>, GUYOT B.<sup>2</sup>, AKAFFOU S.<sup>3</sup>, CHARRIER A.<sup>4</sup>, HAMON S.<sup>1</sup>, LOUARN J.<sup>3</sup> & NOIROT M.<sup>1</sup>

<sup>1</sup> I.R.D., GeneTrop, BP 5045, 34032 MONTPELLIER Cedex 1, France;

<sup>2</sup> CIRAD-CP, Laboratoire Chimie-Technologie, BP 5035, F-34032 Montpellier Cédex 1, France ;

<sup>3</sup> I.R.D., Man, B.P. 434, Côte-d'Ivoire ;

<sup>4</sup> ENSA, UFR Génétique et Amélioration des Plantes, Place Viala, F-34060 Montpellier, France.

**Abstract:** The sucrose content of coffee beans is an important component of the coffee flavour: the higher the sucrose content is in green beans, the more intense coffee cup flavour. Beans of the anciently cultivated cultivar *Coffea liberica dewevrei* have low levels of sucrose compared with beans of the wild species *C. pseudozanguebariae*. In the present study, the inheritance of the sucrose accumulation in the interspecific cross *C. pseudozanguebariae* x *C. liberica dewevrei* was examined. The sucrose content was measured in mature beans of both parental species, and in F1 and backcross hybrids using high-pressure liquid chromatography coupled to pulsed amperometric detection. The sucrose accumulation, in all but one hybrid, showed genetic additivity. A segregation distortion was identified in the offspring of the backcross to *C. pseudozanguebariae*. There was no year effect and only a low "genotype x year" interaction. Consequences for breeding, in relation to the coffee cup taste improvement, are discussed.

**Résumé:** La teneur en saccharose des graines de caféier est une composante importante de l'arôme du café. Plus la teneur est élevée, plus l'arôme est intense. Les graines de l'espèce anciennement cultivée *Coffea liberica dewevrei* ont une faible teneur en sucrose par rapport aux graines de l'espèce sauvage *C. pseudozanguebariae*. Dans cette étude, nous avons examiné l'héritage de l'accumulation du saccharose dans un croisement interspécifique *C. pseudozanguebariae* x *C. liberica dewevrei*. La teneur en saccharose a été mesurée sur les graines matures des 2 espèces parentales, des hybrides F1 et des hybrides "back-cross" en utilisant une HPLC couplée à une détection ampérométrique pulsée. La teneur en saccharose montre une additivité sauf chez un type d'hybride. Une distorsion de ségrégation a été identifiée dans les descendants du rétrocroisement sur *C. pseudozanguebariae*. Il n'y a pas d'effet "année" et seulement une faible interaction "génotype x année". Les conséquences pour l'amélioration génétique du goût à la tasse sont discutées.

**Key-words:** *Coffea pseudozanguebariae*, *Coffea liberica*, interspecific hybrids, HPLC, sucrose, inheritance

## INTRODUCTION

Only two species, *Coffea arabica* (2n=44) and *C. canephora* (2n=22) – or Robusta – have a worldwide economic importance. The former represents 68% of the production because, in the cup, it has a low bitterness and a good flavour. Preferred by consumers, its price is twice that of Robusta. These two species are also cultivated in different ecological environments. *C. arabica* is cultivated in tropical highlands (> 600 m), mainly in Latin America, East Africa and India, whereas *C. canephora* is cultivated in lowlands of West Africa, Indonesia, Vietnam and Brazil. The choice of Robusta for cultivation is mainly due to unfavourable conditions for *C. arabica*. Consequently, if lowland farmers want to improve their incomes they also need improvement of the *C. canephora* cup taste quality.

Coffee flavour is developed during the roasting process from aroma precursors present in green beans. The aroma formation is very complex and includes Maillard and Stecker's reactions and flavour precursor degradations (DE MARIA *et al.*, 1994). Among them, sucrose which is degraded in furans, is one of the most important (CLIFFORD, 1985; DART & NURSTEN, 1985; DE MARIA *et al.*, 1996). In addition, the sucrose also reacts with amino acids producing pyrazine and carbonyl compounds also involved in flavour (THALER & ARNETH, 1967; FELDMAN *et al.*, 1969). The "sucrose/amino acids" ratio in green beans widely determines the profile of volatile compounds (HASSAN, 1970; TRESSL *et al.*, 1982). Low-molecular-weight sugars (glucose, fructose, raffinose, stachyose, arabinose, ribose, mannose, galactose and rhamnose) may also react with amino acids or be degraded but they were recorded only as traces (0.5% of total) in green beans (KROPLIEN, 1971; TRESSL *et al.*, 1982).

In *C. arabica*, the green bean sucrose content is higher than in *C. canephora*: 8.2-8.3 % vs 3.3-4.1 % dmb (TRESSL *et al.*, 1982). Because the "sucrose/amino acids" ratio is involved in organoleptic differences, sucrose increase in *C. canephora* green beans constitutes a priority in breeding programmes. This improvement could not be achieved using

*C. arabica* because all attempts to include this species led to sterility problems (BERTHAUD, 1977; LE PIERRES, 1995). Among wild *Coffea* species, *C. pseudozanguebariae*, native from East Africa, has a sucrose content similar to that of *C. arabica* (unpubl. data : C.L. Ky) and could be a gene donor for Robusta.

The genetic inheritance of the sucrose content in coffee beans has only been studied in intraspecific crosses involving low differences between genitors (MONTAGNON *et al.*, 1998). At an interspecific level, more genetic information is needed for this particular situation: (1) some major genes vs. many minor genes, (2) additivity vs. dominance, and (3) relative importance of genotypic and environmental effects.

Because only the F1 hybrids *C. pseudozanguebariae* x *C. canephora* were available, the cross *C. pseudozanguebariae* x *C. liberica dewevrei*, for which both F1 and back-cross hybrids were available, were used. In this cross, parental species also present large differences in sucrose content. In addition, the low genetic distance between *C. liberica dewevrei* and *C. canephora* should facilitate the inference of results to the cultivated species.

The aims of the present study were: (1) to evaluate the sucrose content of green beans in parental species and their hybrids using anion exchange chromatography coupled to pulsed amperometric detection, (2) to estimate the interannual variation and (3) to determine the genetic additivity of the sucrose content.

## MATERIALS AND METHODS

### PLANT MATERIALS

Plant materials were maintained at the IRD Agricultural Station (Man, Côte-d'Ivoire). *Coffea liberica dewevrei* (DEW) and *C. pseudozanguebariae* (PSE) trees were grown on two separate plots of the station. Both species are allogamous and heterozygous, but their fructification time vary from 3 (PSE) to 10 months (DEW). Interspecific F1 hybrids were obtained from a cross between genitors PSE8044 (now no longer available) and DEW5851 (available and included in the sample). The reciprocal cross (DEW x PSE) never gave progeny. The crossing procedure was made as described by LOUARN (1992). Thirty F1 hybrids were grown on the same plot (D 16) of the station. Backcross hybrids - 62 BCDEW (F1 x DEW) and 7 BCPSE-1 (F1 x PSE) - were obtained by open pollination of 13 of the 30 F1 hybrids (used as female parent). Their origins were checked using molecular markers. In contrast, BCPSE-2 hybrids were originated from a hand-pollinated cross. BCDEW and BCPSE-1 hybrids were grown in the same plot (the two types of hybrids were randomly mixed) whereas BCPSE-2 were grown in another plot. In all plots the plant density was 2220 plants/ha. The same fertiliza-

tion was applied for each tree: 150 g of urea and 100 g of NPK (10:18:18), three times per year.

For the additivity test, seven trees were harvested in each group: PSE, BCPSE, F1, BCDEW and DEW. All these plants, as well as the parental species, were harvested in 1996, except the BCPSE-2 hybrids in 1997. To test a potential effect of the harvesting year, two BCPSE-1 and five BCDEW were harvested in both 1995 and 1997.

### SAMPLE PREPARATION

All beans were harvested at full maturity and de-pulped using the wet processing method. Fifty green beans per tree were immediately frozen using liquid nitrogen to avoid sucrose degradation during the crushing. This was done in a ball mill (Dangoumill, Prolabo, Fontenay-sous-bois, France) for about 2 min, until a fine powder was obtained. The powder was divided into eight samples. Three were used to estimate the water content and five to quantify the sucrose contents.

### EXTRACTION PROCEDURE

Each sample was extracted in a 50-ml capped tube (Sarstedt, Nümbrecht, Germany). Each tube contained about 50 mg of green coffee bean powder to which 25 ml of distilled water were added. Extraction of sucrose was achieved by heating the tubes at 60°C for 15 minutes, tubes being hand-shaken every 5 minutes. Colloidal material present in aqueous extracts was precipitated as described by TRUGO and MACRAE (1985) for chlorogenic acids, by adding 0.5 ml of Carrez solution I, plus 0.5 ml of Carrez solution II, and diluted to 50 ml with distilled water. Carrez solution I was prepared by dissolving 21.9 g of crystallized zinc acetate and 3 ml of glacial acetic acid in 100 ml of distilled water. Carrez solution II consisted of 10.6 g of potassium hexacyanoferrate ( $\text{Fe}^{++}$ ) in 100 ml of distilled water. The solution was then filtered (0.2  $\mu\text{m}$  pore diameter) and directly analysed by high-pressure liquid chromatography (HPLC).

### APPARATUS AND CHROMATOGRAPHIC CONDITIONS USED

Sucrose was analysed using anion exchange chromatography coupled to pulsed amperometric detection. This method is currently used in our laboratory. The chromatographic set-up comprised the following items from Dionex Chromatography Co. (Jouy-en-Josas, France): A GP 40 Gradient Pump was used to propel the eluent, consisting of distilled filtered water (0.2  $\mu\text{m}$  pore diameter) (solvent A) and 150 mM NaOH (solvent B) (50% sodium hydroxide from Mallinckrodt Baker, Deventer, The Netherlands according to the following isocratic gradient: 25% solvent A and 75% solvent B. The eluent was delivered at a flow rate of 1 ml/ min and degassed with helium. An

automated sampler was used with a fixed loop of 50  $\mu$ l. A Carbo-Pac PA 1 (4 x 250 mm) column was used in conjunction with a Carbo-Pac PA 1 Guard precolumn (10 x 32 mm). The pulsed amperometric detector II, equipped with a gold working electrode, was used at the following settings: measurement potential,  $E_1=50$  mV; measurement time,  $t_1=300$  ms;  $E_2=60$  mV;  $t_2=300$  ms;  $E_3=-800$  mV;  $t_3=480$  ms. Solid sucrose standard from Sigma (Ref. S-9378, HPLC grade, Saint-Quentin-Fallavier, France) was dissolved in water and filtered (0.2  $\mu$ m).

#### STATISTICAL ANALYSIS

Genotypic and environmental effects were tested using a two-way ANOVA with hybrids (random effect) and years (fixed effect) as factors.

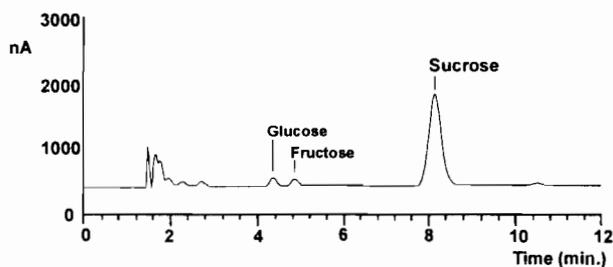
Differences between groups (PSE, BCPSE, F1, BCDEW, and DEW) and between trees within groups were tested using the nested model of ANOVA with two factors: group (fixed effect) and tree nested in group (random effect).

Additivity was tested using a linear regression between the sucrose content average and the expected part of the PSE genome in PSE (100%), BCPSE (75%), F1 (50%), BCDEW (25%) and DEW (0%) (see Fig. 2).

All results were analysed using Statistica software (5.1 version, 1997 for Microsoft Windows).

#### RESULTS

A chromatogram, corresponding to a DEW genotype, is shown in Fig. 1. Only the sucrose peak, detected after a retention time (RT) of 8.3 minutes, was high. As expected, peaks corresponding to glucose (RT of 4.5 min) and fructose (RT of 5.0 min) were present as traces. The within-tree coefficient of variation between extractions was 8.6 %.



**Figure 1** : HPLC chromatogram of sucrose from mature green beans of *C. liberica* 'dewevrei'. HPLC is coupled to a pulsed amperometric detection. The chromatogram shows traces of glucose and fructose. The symbol "nA" correspond to  $10^{-9}$  Amperes.

Sucrose content variation, tested with seven backcrosses hybrids over two years, is reported in Table 1. Between trees, differences were clearly observed ( $F_{6, 56} = 121$ ;  $p < 0.001$ ). In contrast, no difference between years was found ( $F_{1, 6} = 0.50$ ;  $p = 0.51$ ). Presence of an

interaction ( $F_{6, 56} = 5.34$ ;  $p < 0.001$ ) means that the year effect changed between trees.

**Table 1** Variation in bean sucrose content (% dmb) between backcrosses hybrids and years (1995 and 1997)

Trees	Groups	95	97	Means
9/8	BCPSE-1	5.34	5.22	5.28
14/3	BCPSE-1	5.33	4.36	4.84
1/9	BCDEW	3.96	3.67	3.82
12/12	BCDEW	8.25	8.57	8.41
19/13	BCDEW	5.55	6.54	6.04
20/17	BCDEW	5.90	6.54	6.22
9/3	BCDEW	3.32	4.04	3.68
Means		5.38	5.56	

Sucrose content in parental species, F1 hybrids and backcross hybrids are reported in Table 2. Sucrose content in DEW (5.37% dmb) was 30% lower than in PSE. Between trees variation, within the same species, was also highly significant (Table 2).

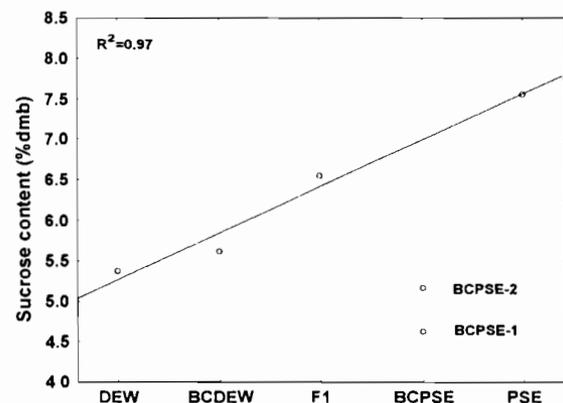
The sucrose content of F1 hybrids observed was close to the theoretical mid-parental value (6.46 % dmb) suggesting additivity (Table 2). The between-F1 hybrid variance ( $s^2 = 1.141$ ) did not differ from the within-parental species one ( $s^2 = 1.335$ ).

**Table 2** Bean sucrose contents (% dmb) in parental species, F1 hybrids and backcross hybrids (BCDEW and BCPSE-1)

	DEW	BCDEW	F1	BCPSE-1	PSE
Average	5.37	5.62	6.55	4.63	7.57
Range	3.76-6.24	3.67-7.47	5.09-8.18	3.83-5.22	5.80-9.25
$F_{6, 2x}$	43.5 ***	64.0 ***	49.6 ***	8.1 ***	10.9 ***

\*\*\* Significant at  $P < 0.001$

Sucrose contents in the backcross hybrids BCDEW and BCPSE-1 overlap (Table 2). As observed for the two species and their F1, the between-tree variation (within backcross) was very highly significant. Nevertheless, the within-BCPSE-1 variance was four times lower ( $s^2 = 0.252$ ) than in the F1, whereas the within-BCDEW variance was twice ( $s^2 = 2.130$ ) the F1. The lower BCPSE-1 variance is explained by its low maximal value (5.2 % dmb).



**Figure 2** . Inheritance of bean sucrose content in the interspecific cross *Coffea pseudozanguebariae* x *Coffea liberica* "dewevrei" (regression analysis)

Sucrose content in BCDEW was close to the expected value in the additivity model  $[(F1+DEW)/2 = 5.15\% \text{ dmb}]$ . In contrast, sucrose content in BCPSE-1 was significantly different from the expected value  $[(F1+PSE)/2 = 7.06\% \text{ dmb}]$ . A second sample of seven BCPSE hybrids (BCPSE-2) showed the same discrepancy with values ranging from 4.54 to 5.65% dmb with an average of 5.17%. The difference between BCPSE-1 and BCPSE-2 remained significant.

The result of the additivity test is summarized in Fig. 2. The linear regression involved the DEW, BCDEW, F1 and PSE group averages. The relationship is linear and significant representing 97 % of the variance.

## DISCUSSION

### INHERITANCE OF SUCROSE ACCUMULATION.

Our results showed: (1) additivity of bean sucrose content in F1 and in BCDEW and (2) within-F1 variation similar to the intraspecific variation. Additivity of sucrose content was also observed in *C. canephora* intraspecific crosses (MONTAGNON *et al.*, 1998). Additivity means that the sucrose content average of a progeny should be close to the mid-parental value. In our study this is true except for the BCPSE group for two independent samples. This exception could be explained by segregation distortion but further experiments are required to explain such discrepancy.

The two BCPSE samples were considered as different with a risk I close to 5%. There are two interpretations: (1) the difference did not exist and our observation corresponds to the risk I, and (2) the difference exists and could be due to the "year x genotype" interaction.

The large within-F1 variation suggested a polygenic inheritance with independent genes and a high level of parental heterozygosity. This last assumption is consistent with their allogamy. Similarly, the larger variance of BCDEW reflected the polygeny, the high interspecific heterozygosity of the F1 and the high level of recombination.

### ABSENCE OF BETWEEN-YEAR EFFECT AND PRESENCE OF HYBRID X YEAR INTERACTION.

The final sucrose content in green beans results from a long biochemical process where several external parameters are involved: (1) sucrose content will vary when climatic conditions change during fructification because it is a major photosynthesis output (HUBER, 1989); (2) sucrose content will vary when harvest takes place at different bean maturity levels because it reaches its maximum in mature beans (PETNGA, 1986); and (3) sucrose content can also be reduced during post-harvest processing, depending on drying and storage conditions (elevated temperatures

enhanced the development of microorganisms) (GUYOT *et al.* 1995; POKORNY, 1975). These different external factors could explain between-year variation of sucrose content, which is expected when fructification periods are synchronous for all trees and pre and post harvest processing are similar within year but can differ between years.

In contrast, we did not observe between-year difference, but, instead, an "hybrid x year" interaction, which means that external factors influencing the sucrose content changed between hybrids. In fact, fructification times of parental species are 3 and 10 months for PSE and DEW, respectively, and a large variation of this trait exists in hybrids of the second generation. Consequently, climatic conditions from flowering to seed storage were different for each hybrid, explaining the interaction.

### IMPORTANCE OF THE GENETIC EFFECTS IN THE SUCROSE CONTENT VARIATION.

Differences between species, observed using ANOVA, could be interpreted as between plots differences (species are growing in different plots). Nevertheless, the linear regression we used to emphasize additivity, cannot be obtained under an environmental hypothesis (plots of PSE, DEW, F1 and BC are randomly located on the station) highlighting the importance of genetic effects in the between species differences.

Within groups (species, F1, and backcrosses), between-tree differences constituted the main source of sucrose content variation. Similar results were recorded in apricot, *Prunus armeniaca* L., where constant patterns of sucrose content were observed over three years for six parents and their progenies (BASSI *et al.*, 1996).

In *Prunus*, as in *Coffea*, the importance of the between-tree variation in a backcross is due to the large difference in sucrose content of parental species. We expect a decreasing of this part of variation in backcrossing generations.

### CONSEQUENCES FOR BREEDING PROGRAMMES

As seen before, *C. arabica* bean sucrose content is higher than that of *C. canephora* (8.2-8.3% vs. 3.3-4.1% dmb) and constitutes one major parameter of the coffee cup quality. Within the coffee gene pool, *C. pseudozanguebariae* has a sucrose value close to *C. arabica* and consequently could be used for *C. canephora* breeding for this character.

Our results clearly indicate that, using an interspecific cross with PSE, we can expect an improvement of the sucrose content in *C. canephora* green beans. The intraspecific variability of both parental species (*C. canephora* and *C. pseudozanguebariae*) and the genetic additivity allow a choice of genotypes as parents for breeding. As an example, the PSE geno-

type having 9.25% dmb of sucrose is of special interest (Table 2). The large within-BCDEW variance, coupled with additivity, strongly suggests that the choice of parents for the next generation BC2 should be efficient.

#### PROSPECTS

The sucrose impact in aroma is very complex and further physiological and genetic studies are needed. On one hand, studies must be made on sucrose-metabolizing enzymes activities occurring during bean growth and maturation: sucrose synthase, sucrose phosphate synthase and acid invertase have already been isolated (HUB-BARD *et al.*, 1989; BRUNEAU *et al.*, 1991; MORIGUCHI *et al.*, 1992; SUN *et al.*, 1992). On the other hand, we also need to focus on sucrose content relationships with: (1) traits such as morphology (leaf area, bean weight, etc.); (2) phenology (flowering-ripening period, etc.); and (3) biochemistry of other compounds (fats, trigonelline, free amino acids, etc.). Such data, added to what we already know on chlorogenic acids, caffeine and heteroside inheritance on the same interspecific cross (BARRE *et al.*, 1998; KY *et al.*, 1999), would certainly enhance diploid coffee breeding programmes. In addition, molecular tools will be used to screen quantitative trait loci (QTL) associated with sucrose accumulation, and consequently to develop marker-assisted selection at the plantlet stage.

#### REFERENCES

- BARRE, P., AKAFFOU, S., LOUARN, J., CHARRIER, A., HAMON, S. & NOIROT, M. 1998. Inheritance of caffeine and heteroside contents in an interspecific cross between a cultivated coffee species *Coffea liberica* var *dewevrei* and a wild species caffeine free *C. pseudozanguabariae*. *Theor. Appl. Genet.* 96: 306-311.
- BASSI, D., BARTOLOZZI, F., & MUZZI, E. 1996. Patterns and heritability of carboxylic acids and soluble sugars in fruits of apricot (*Prunus armeniaca* L.). *Plant Breeding* 115: 67-70
- BERTHAUD, J. 1977. Caractéristiques comparées des hybrides interspécifiques tétraploïdes et hexaploïdes *Coffea arabica* L. x *C. canephora* Pierre. *Proc. Int. Congr ASIC* 8: 393-397
- BRUNEAU, J.M., WORELL, A.C., CAMBOU, B., LANDO, D. & VOELKER, T.A. 1991. Sucrose phosphate synthase, a key enzyme for sucrose biosynthesis in plants. *Plant Physiology* 96: 473-478
- CLIFFORD, M.N. 1985. Chemical and physical aspects of green coffee and coffee products. In: *Coffee, botany, biochemistry and production of beans and beverage* CLIFFORD & WILSON (eds) Publishing Company Inc., Westport, Connecticut, pp. 305-374.
- DART, S.K. & NURSTEN, H.E. 1985. Volatile components. In: *Coffee: Chemistry (Vol. 1)* CLARKE & MACRAE (eds) Elsevier Applied Science Publishers Ltd, London, UK, pp. 223-265
- DE MARIA, C.A.B., TRUGO, L.C., MOREIRA, R.F.A. & WERNECK, C.C. 1994. Composition of green coffee fractions and their contribution to the volatile profile formed during roasting. *Food Chem.* 50: 141-145
- DE MARIA, C.A.B., TRUGO, L.C., AQUINO NETO, F.R., MOREIRA, R.F.A. & ALVIANO, C.S. 1996. Composition of green coffee water-soluble fractions and identification of volatiles formed during roasting. *Food Chem* 55: 203-207
- FELDMAN, J.R., RYDER, W.S. & KUNG, J.T. 1969. Importance of nonvolatile compounds to the flavor of coffee. *J Agric. Food Chem.* 14: 733-739
- GUYOT, B., GUEULE, D., ASSEMAT, S., TCHANA, E. & POMATHIOS, L. 1995. Influence du mode de préparation du café vert Robusta sur sa composition chimique et ses qualités organoleptiques. *Proc. Int. Congr. ASIC* 16: 267-277.
- HASSAN, M.U. 1970. *The precursor of coffee aroma*. Ph D Thesis, University of Strathclyde, Glasgow.
- HUBART, N.L., HUBER, S.C. & PHARR, D.M. 1989. Sucrose phosphate synthase and acid invertase as determinants of sucrose concentration in developing muskmelon (*Cucumis melo* L.) fruits. *Plant physiol.* 91: 1527-1534
- HUBER, S.C. 1989. Biochemical mechanism for regulation of sucrose accumulation in leaves during photosynthesis. *Plant physiol.* 91: 656-662
- KROPLIEN, U. 1971. Monosaccharides in coffees and coffee substitutes. *Proc. Int Congr. ASIC* 5: 217-223
- KY, C.-L., LOUARN, J., GUYOT, B., CHARRIER, A., HAMON, S. & NOIROT, M. 1999. Relations between and inheritance of chlorogenic acids contents in an interspecific cross between *C. pseudozanguabariae* and *Coffea liberica* var 'dewevrei'. *Theor. Appl. Genet.* 98: 628-637
- LE PIERRES, D. 1995. *Etude des hybrides interspécifiques tétraploïdes de première génération entre C. arabica L. et les caféiers diploïdes*. Ph D Thesis, University of Orsay, France.
- LOUARN, J. 1992. *La fertilité des hybrides interspécifiques et les relations génomiques entre les caféiers diploïdes d'origines africaines (Genre Coffea L. sous-genre Coffea)*. Ph D Thesis, University of Orsay, France.
- MONTAGNON, C., GUYOT, B., CILAS, C. & LEROY, T. 1998. Genetic parameters of several biochemical compounds from *Coffea canephora* green coffee. *Plant Breeding* 117: 576-578

- MORIGUCHI, T., ABE, K., SANADA, T. & YAMAKI, S. 1992. Levels and role of sucrose synthase, sucrose phosphate synthase, and acid invertase in sucrose accumulation in fruit of asian pear. *J. Amer. Soc. Hort.* 117: 274-278
- PETNGA, E. 1986. *Application de l'analyse multidimensionnelle à la classification qualitative du café Robusta en fonction de la maturité*. Ph D Thesis, University de Montpellier, France.
- POKORNY, J., CON, N.H., SMIDRKALOVA, E. & JANICEK, G. 1975. Non-enzymic browning XII. Maillard reactions in green coffee beans on storage. *Z. Lebensm. Unters. Forsch.* 158: 87-92
- SUN J., LOBODA, T., SUNG S.J.S. & BLACK, C.C. 1992. Sucrose synthase in wild tomato, *Lycopersicon chmielewskii*, and tomato fruit sink strenght. *Pl. physiol.* 98: 1163-1169
- THALER, H. & ARNETH, W. 1967. Hochpolymere Kohlenhydrate des rohen end gerösteten Kaffees. *Coll. Int. Chimie des Cafés* 3: 127-136
- TRESSL, R, HOLZER, M. & KAMPERSCHOER, H. 1982. Bildung von Aromastoffen in Röstkaffee in Abhängigkeit vom Gehalt an Aminosäuren und Reduzierenden Zuckern. *Proc. Int. Congr. ASIC* 10: 279-292
- TRUGO, L.C. & MACRAE, D R 1984. Chlorogenic acid composition of instant coffees. *Analyst* 109: 263-266

**Annexe D**

---

## INTERSPECIFIC GENETIC LINKAGE MAP, SEGREGATION DISTORTION AND GENETIC CONVERSION IN COFFEE (*COFFEA* SP.).

KY, C.L.<sup>\*</sup>, BARRE, P.<sup>\*</sup>, LORIEUX, M.<sup>\*</sup>, TROUSLOT, P.<sup>\*</sup>, AKAFFOU, S.<sup>\*\*</sup>, LOUARN, J.<sup>\*\*</sup>, CHARRIER, A.<sup>\*\*\*</sup>, HAMON, S.<sup>\*</sup> & NOIROT M.<sup>\*</sup>

Addresses : <sup>\*</sup>GeneTrop, Centre IRD, B.P. 5045, 34032 MONTPELLIER cedex 1, FRANCE

<sup>\*\*</sup> Station IRD de Génétique du café, B.P. 434, MAN, COTE-D'IVOIRE.

<sup>\*\*\*</sup> ENSA, UFR Génétique et Amélioration des Plantes, Place Viala, F-34060 MONTPELLIER, FRANCE

**Abstract:** An interspecific partial genetic linkage map of *Coffea* sp. based on 62 backcross hybrids is presented. F1 hybrids were generated by a cross between the wild *C. pseudozanguebariae* and the anciently cultivated *C. liberica* var. *dewevrei* (DEW). Progeny derived from a back cross between F1 hybrid and DEW. The map construction consisted on a two-step strategy using 5.5 and 3.1 LOD score revealed by simulation file. The map consisted of 181 loci: 167 amplified fragment length polymorphism (AFLP) and 13 random fragment length polymorphism (RFLP). The markers were assembled into 14 linkage groups, each with 4 to 31 markers covering 1144 cM. Segregation distortion was observed for 30 % of all loci, with a particular (3 : 1) and (1 : 3) ratios, equally favouring each of the two parents. The origin of such ratio suggests the implication of genetic conversion events. This map represents also an initial step towards the detection of quantitative trait loci.

**Résumé:** Une carte génétique interspécifique partielle de caféier, obtenue à partir de 62 hybrides rétrocroisés, est présentée. Les hybrides F1 ont été obtenus par croisement entre l'espèce sauvage *C. pseudozanguebariae* et l'espèce anciennement cultivée *C. liberica* var. *dewevrei* (DEW). La descendance provient d'un rétrocroisement entre l'hybride F1 et DEW. La construction de la carte est basée sur une stratégie en 2 étapes utilisant les valeurs de LOD score 5.5 et 3.1, estimées sur un fichier simulé. La carte comprend 181 loci: 167 *amplified fragment length polymorphism* (AFLP) et 13 *random fragment length polymorphism* (RFLP). Les marqueurs sont regroupés dans 14 groupes de liaison ayant de 4 à 31 marqueurs. La longueur de la carte est de 1144 cM. Des distorsions de ségrégation s'observent pour 30 % des loci. Deux types de rapports existent alors en fréquence égale: (3:1) et (1:3). De tels rapports suggèrent l'implication d'événements de conversion génétique. Cette carte génétique constitue aussi une étape préliminaire à la recherche de *quantitative trait loci* (QTL).

**Key-Terms:** *Coffea* sp., AFLPs, RFLPs, Linkage map, Segregation distortion, Genetic conversion

## INTRODUCTION

PCR-based methods reduce cost and time for screening large number of genetic markers, such as AFLP markers (amplified fragment length polymorphism) (VOS *et al.*, 1995). The AFLP technology has two main advantages: (1) more reliability and consistence than RAPDs (random amplified polymorphism digested), (2) a large number of markers on a single high-resolution sequencing gel. Consequently, a high density map could be obtained from a low number of primers and minute amounts of DNA. AFLP markers are typically dominant and most fragments correspond to unique positions on the chromosome and can be exploited in genetic and physical mapping (THOMAS *et al.*, 1995; MEKSEM *et al.*, 1995). RFLP (restriction fragment length polymorphism) markers do not have such advantages: the number of markers obtained from a gel is lower and require higher DNA amount. By contrast, they are a co-dominant and allows between-map comparison.

Coffee trees belong to the genus *Coffea* sub-genus *Coffea*, family *Rubiaceae* and are mostly present in tropical and subtropical regions of the world. Most species (up to 80 taxa) are diploid ( $2n=22$ ), except *C. arabica* which is tetraploid (LEROY, 1980; BRIDSON, 1987). Two species are currently cultivated over the world: *C. arabica* and *C. canephora*. Wild species constitute the base of genetic resources for breeding. Interspecific crosses have been realised to test breeding possibilities by introgression (LOUARN, 1992). Those including East African species and a West African species constitute an extreme situation, which is illustrated by the cross between *C. pseudozanguebariae*, native from Kenya and Tanzania, and *C. liberica* var. *dewevrei*, native from Centrafrique. We retain the *C. pseudo-zanguebariae* x *C. liberica* 'dewevrei' cross for genetic mapping for two main reasons: (1) the availability of F1 hybrids as well as backcrossed hybrids and (2) the low genetic distance between *C. liberica* 'dewevrei' and *C. canephora* (CARVALHO & MONACO, 1967; CHARRIER, 1978; LOUARN, 1992), which facilitate the results inference for the cultivated species.

Markers distorted segregation was reported in several interspecific linkage analysis (GRANT, 1975). This

systematic deviation from equal representation of alleles among functional gametes involves all chromosomes and could be caused by selection processes at gamete or zygote stage (GADISH & ZAMIR, 1986; ZAMIR & TADMOR, 1986). This study was also designed to obtain a quantitative estimation for the proportion of unequally segregating loci in the interspecific cross and to give a possible explanation of it in the *Coffea* genera. The main purposes of this paper are: (1) to present the first interspecific AFLP marker-based genetic linkage map of a cross in *Coffea*, (2) to identify sets of markers showing important segregation distortion.

## MATERIALS AND METHODS

### PLANT MATERIAL

Plant material is maintained at the Agricultural Station IRD (Man, Côte-d'Ivoire). The cross between *C. pseudozanguebariae* Bridson (PSE) (geno-type 8044 used as female parent) and *C. liberica* var. *dewevrei* (DEW) (genotype 5851 used as male parent) was made by hand pollination under bag according to LOUARN (1992). Twenty five F1 hybrids were obtained and grew on the same plot of the station. Sixty two backcross hybrids on DEW (BCDEW) were obtained by open pollination of thirteen F1 hybrids (used as female parent). Backcross hybrids were identified using RFLP and AFLP molecular markers (180 PSE and 30 DEW specific markers). Twelve trees of both species were used to identify species specific AFLP markers.

Before being sent to France, fresh coffee leaves were packed in wet newspaper in an isotherm box. They were lyophilised and then stored in a cold room (4°C, 20 % relative humidity).

### DNA EXTRACTION

Coffee leaves of PSE and hybrids were recalcitrant to DNA extraction. This was due to the presence of polysaccharides and other secondary metabolites. A direct extraction with CTAB (DOYLE & DOYLE, 1987) or MATAB gave DNA with gelatinous substance (polysaccharide), limiting PCR amplification and enzymatic digestion. A two step large-scale DNA extraction (nuclei isolation followed by MATAB lysis) constituted the protocol for DNA extraction in coffee plant:

#### Step 1 - Nuclei extraction:

- Crushing of lyophilised leaves during 4 min in a ball mill (Dangoumill) to obtain a fine powder;
- Mixing of powder (1 g) with 100 mL of nuclei extraction buffer in an Erlenmeyer flask. The slightly modified lysis buffer of Dolezel (1989) was used: 15 mM Tris, 2 mM EDTA, 80 mM KCl, 20 mM NaCl, 0.5% Triton X-100, 0.1 %  $\beta$ -mercaptoethanol<sup>1</sup>, pH 9;

- Shaking of Erlenmeyer flask for 2 h on a stirring table at 125 rpm;
- Transfer of the suspension in two 50 ml capped tubes (Sarstedt);
- Centrifugation for 20 min. at 3000 g;
- Pooling of pellets and re-suspensions in 20 mL of lysis buffer, which was composed of 4 % MATAB, 0.1 M Tris, 1.25 M NaCl, 0.02 M EDTA, pH 8.0;
- Storage of samples -20°C.

#### Step 2 - Nuclei lysis and DNA extraction:

- Incubation of capped tubes at 65°C for 4 h;
- Centrifugation at 3000 g for 10 min;
- Transfer of supernatant into a new 50 mL tube;
- Addition of twenty millilitres of chloroform/isoamyl alcohol (24/1: v/v);
- After a brief agitation, centrifugation of the emulsion for 10 min at 3000 g. Washing with chloroform/isoamyl alcohol was done twice;
- RNA suppression by incubation for 30 min at 37°C with 200  $\mu$ L of RNAase solution (10 mg/mL of RNAase A from bovine pancreas, Boehringer);
- DNA precipitation with 20 mL of isopropyl alcohol;
- Centrifugation at 3000 g for 10 min;
- Re-suspension of pellet in 0.7 mL of TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0);
- Transfer into 1.5 mL tubes;
- DNA precipitation with 0.6 mL of isopropyl alcohol and 70  $\mu$ L of 3M sodium acetate (pH 5.2);
- Centrifugation at 12000 g for 10 min;
- Washing of the pellet in 70 % ethanol;
- Centrifugation at 12000 g for 10 min;
- Drying of pellet;
- Re-suspension in TE.

DNA content was evaluated on 1% agarose gel comparing the sample to the  $\lambda$  DNA/EcoR I + Hind III marker (Promega). DNA yield depended considerably on genotypes and varied from 125  $\mu$ g to 1000  $\mu$ g per 1 g of lyophilised leaves.

### RFLP MARKERS

DNA probes came from two genomic libraries. Probes for which the name began by G or AR derived from a *Bam*HI library of Arabusta genotype (hybrid issue from a cross between *C. arabica* and *C. canephora*). The probe designed CA were obtained from a *C. arabica* *Pst*I library. Preparation and characteristics of the libraries were previously described (PAILLARD, *et al.* 1996). DNA digestion, separation by electro-phoresis, blotting and hybridisation were performed following the manufacturer's instructions.

### AFLP MARKERS

AFLP protocol by ZABEAU and VOS (1993) and VOS *et al.* (1995) was carried out with minor modifications. Total genomic DNA (50 ng/  $\mu$ l) was restricted with two enzyme combinations, 10 U *Eco*RI and 5 U *Mse*I (from Gibco BRL), in 5X T4 ligase buffer for 2 h. at

37°C. Twenty-five microlitres of a mixture containing 100 ng/µl of *EcoRI* adapter, 1 µg/µl of *MseI* adapter, 1 U T4 DNA ligase (from Gibco BRL) and AFLP water grade were added to the restricted-DNA. The ligation reaction was incubated at 37°C for 3 h.

The pre-amplification profile by PCR was performed with 5 µl of template DNA using a primer pair based on the sequences of the *EcoRI* and *MseI* adapters. This lasts had one additional selective nucleotide at the 3' end (*EcoRI* + 1 primer and *MseI* + 1 primer) (VOS, *et al.*, 1995). This first round of amplification was as follows: 20 cycles (30 s at 94°C, 1 min at 56°C and 1 min at 72°C). Pre-amplification products were then diluted 30-fold with AFLP water grade. The second and final round of amplification was performed using primer based on the same sequence but with two additional selective nucleotides at the 3' end (*EcoRI* + 3 primers and *MseI* + 3 primers). The *EcoRI* + 3 primers were labelled by phosphorylating the 5' end with [ $\gamma$  33 P] ATP (Amersham). The amplification profile was as follows: 13 cycles (30s at 94°C, 30s at 65°C, 1 min at 72°C) and 33 cycles (30s at 94°C, 30s at 56°C, 1 min at 72°C). The PCR was performed on a MJ Research PTC 200 thermocycler.

The PCR products were mixed with an equal volume of 98% formamide, 10 mM EDTA, 0.05% bromophenol blue and xylene cyanol as tracking dyes. They were denaturated at 90°C (3 min) and maintained at 70°C. Aliquots (5 µl) were loaded onto a denaturing 8 % polyacrylamide gel which consisted of: 40 g of urea, 12 mL of acrylamide, 8 mL of TBE 10 X (108.0g Tris, 55.0g boric acid, 40.0 mL EDTA 0.5 M pH 8 for 1 litre) and 28 mL of distilled water. Electrophoresis buffer was 1 X TBE. Gels were run at constant power (50 W for 3h40), fixed, dried as for sequencing gels and then exposed to X-ray film (Kodak Biomax film) for 48-72 h.

#### MORPHOLOGICAL MARKER

Only one morphological and monogenic marker was used for mapping : the fruit color at maturity (purple in PSE and red in DEW, the colour purple being dominant).

#### DATA ANALYSIS

AFLP and RFLP data were obtained by visual scoring of autoradiograms, independently by two readers. Only unambiguous bands specific to PSE were scored (present = 1, absent = 0). Marker segregations were checked for deviation from the expected Mendelian segregation (1:1) by chi-squared analysis.

#### MAP CONSTRUCTION STRATEGIES

Linkage analysis was assessed with MAPMAKER / EXP version 3.0 b computer software (LANDER, *et al.* 1987, LINCOLN, *et al.* 1992). Map units (cM) were derived using KOSAMBI (1944) mapping function. The marker order was confirmed with the ripple command.

A two-step procedure was carried out to manage the two statistical error risks (I and II). In the first step, MAPMAKER was separately applied to distorted markers and non-distorted markers. The LOD score was defined to obtain highly likely linkage groups. In the second step, MAPMAKER was applied to regroup previously defined linkage groups that can be linked. The LOD score was then defined to avoid pseudo-linkage.

**Table 1** Number of AFLP amplification products generated with 30 primer combinations *Eco* +3 3' end selective nucleotides of the primers complementary to the *Eco* and *Mse* adapter, respectively

Primer combinations		Number of discriminant bands
<i>Eco</i> + 3	<i>Mse</i> + 3	
AAC	CAA	3
AAC	CAC	10
AAC	CAG	3
AAC	CAT	10
AAC	CTA	10
AAC	CTC	4
AAC	CTG	10
AAC	CTT	7
AAG	CAT	11
AAG	CTA	8
AAG	CTC	1
AAG	CTG	6
AAG	CTT	12
ACA	CAA	5
ACA	CAC	4
ACA	CAT	3
ACA	CTA	10
ACA	CTC	5
ACA	CTG	8
ACA	CTT	6
ACC	CAA	7
ACC	CAC	9
ACC	CAT	4
ACC	CTA	5
ACC	CTG	4
ACC	CTT	6
ACT	CAA	5
ACT	CAT	7
ACT	CTC	5
ACT	CTT	7

A simulation was carried out to define the two LOD score values. The simulated table of 62 individuals and 80 markers was structured as follows:

- four independent linkage groups A, B, C, D of ten non-distorted markers. The first marker of each linkage group was obtained using a binomial distribution (N=2, p=0.5). The second marker rose from the first marker by random swapping (0 became 1 and vice-versa) for 10% of the 62 individuals. The same process was used to obtain the third marker from the second one, and so on, up to the tenth marker. Markers were the approximately equally spaced;
- an independent linkage group E of ten distorted markers (segregation 3:1). Its obtaining differed only by the initial use of a binomial distribution with p=0.75;
- a second distorted linkage group F, related to the group A of non-distorted markers. This group F also included of ten markers, but its first marker rose from the first marker of the group A in which 50%

of 0 (selected at random) became 1. Thus A and F should be linked after mapping;

- ten distorted and independent markers, obtained using a binomial distribution ( $N=2$ ,  $p = 0.75$ );
- ten distorted markers simulating band superposition. Each marker "M3" resulted from two markers ("M1" and "M2") randomly selected among the groups A, B, C and D. The algorithm was: if "M1"=1 or "M2"=1 then "M3"=1, else "M3"=0.

The two LOD scores, we defined from the simulation, were 5.5 for the first step and 3.1 for the second one.

## RESULTS

A total of 192 PSE-specific AFLP markers (present in PSE and F1 hybrids and absent in DEW) were scored on 60 gels. Table 1 gives the 30 primer combinations used. The number of selected bands varied from 1 to 12, with an average of 6 bands per combination. Sixteen RFLP markers were also scored, corresponding data could be found in BARRE (1997).

In a way to applied our mapping strategies, segregation distortions were tested for each marker, leading to define two groups of markers: distorted and non-distorted.

### GENETIC LINKAGE MAP

The first step of the mapping strategy (using a LOD score of 5.5) defined 19 linkage segments of non distorted markers and six linkage segments of distorted markers (black segments on Fig. 1). Twelve non distorted markers and 16 distorted markers were found independent by MAPMAKER and were not mapped. The second step (LOD score value of 3.1) led to a final genetic map with 14 linkage groups (four to 31 loci per linkage group) (Fig. 1).

**Table 2** Linkage groups main characteristics (size, number of markers and distances) of the genetic map resulted from the 62 BCDEW hybrids progeny

Linkage groups	Length (cM)	umber of markers	Means distances (cM)	Min-max (cM) distances
A	224.4	31	7.5	0.0 - 25.3
B	150.6	21	7.5	0.0 - 22.3
C	112.7	16	7.5	0.0 - 24.4
D	101.6	15	7.2	0.0 - 17.8
E	100.4	12	9.1	1.9 - 15.0
F	97.3	13	8.1	0.0 - 17.6
G	96.1	19	5.3	0.0 - 20.1
H	70.7	13	5.9	0.0 - 28.7
I	51.8	11	5.2	0.0 - 22.0
J	48.6	11	4.9	0.0 - 24.4
K	26.9	5	6.7	0.0 - 20.3
L	25.2	6	5.0	0.0 - 12.6
M	19.7	4	6.7	0.0 - 16.8
N	18.2	4	6.1	1.6 - 13.4

Table 2 summarised the main statistics of each linkage group. Linkage groups could be classified in three categories: i) large-sized linkage groups (150 to 225 cM) with 31 to 21 loci (groups A and B), ii) medium-sized linkage groups (48 to 113 cM) with 11 to 19 loci

(groups C to J), and iii) small-sized linkage groups (18 to 27 cM) with 4 to 6 markers (groups K to N). The genetic linkage map covered 1144 cM.

The distances between two consecutive markers varied from 0 to 29 cM, with a mean of 7 cM. This distance distribution, shown on figure 2, reveals a strong skewness. Among the 167 intervals, sixty were smaller than 2 cM (36 %) and twelve were larger than 20 cM (7 %). The small intervals were gathered and eight clusters could be easily observed in large and medium-sized linkage groups (except groups D and E on Fig. 1). The density on these clusters varied from 6 to 15 markers for an interval segment of 20 cM.

### SEGREGATION DISTORTION

Segregation distortion were observed for 30% of loci (53 AFLP and 1 RFLP markers). The distribution of segregation ratio is symmetric with three modes centred on three different ratios: (1 : 3), (1 : 1) and (3 : 1) (Fig. 3). If we assume that there are mainly three segregation types resulting in three binomial distribution, with p parameter equal to 0.25, 0.5 and 0.75, respectively. Tested with chi-square, the distorted loci are classified as: i) 26 loci (14 % of the 181 markers) segregated according to a (1 : 3) ratio in favour of PSE, ii) 21 loci (12 %) segregated as a (3 : 1) ratio in favour of DEW and iii) 7 loci presenting stronger distortion of segregation (Table 3). Proportions of (1 : 3) and (3 : 1) loci were similar.

**Table 3.** Average numbers of segregation types of the 181 markers in the BCDEW progeny

Segregation ratio			
1 1*	1 3*	3 1*	thers
70 %	14 %	12 %	4 %

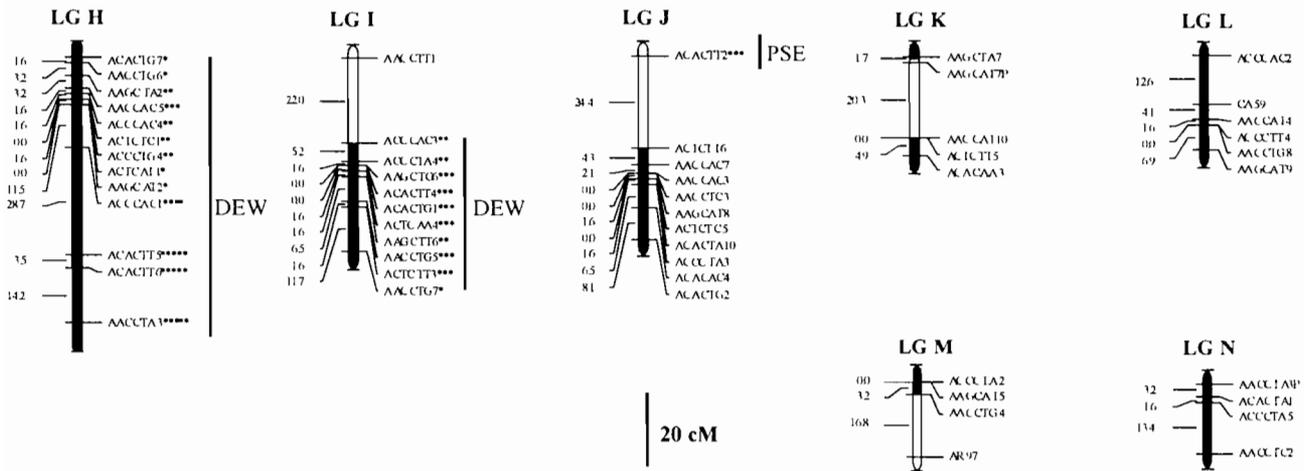
\*Significant at  $\alpha = 0.05$  level

The distribution of the 54 distorted markers on the genetic map is illustrated on figure 1. They were grouped at the end of long and intermediate linkage groups size and were not found on the four smaller groups. The segments which are distorted in favour of DEW constituted the whole of the H group and a part of I group. The distorted fragments in favour of PSE were located on 6 linkage groups: the totality of C, the halves of the B, E and F groups and, for one and two markers, on the J and G groups, respectively.

## DISCUSSION

The PSE x DEW interspecific coffee map consisted of 181 markers distributed on 14 linkage groups covering 1144 cM. AFLP clusters were observed, suggesting location of centromeric regions. This map also revealed a relative important part of distorted loci (30 %), which were symmetrically distributed following a (3 : 1) and (1 : 3) ratios in favour of PSE (14 %) and DEW (12 %) respectively.





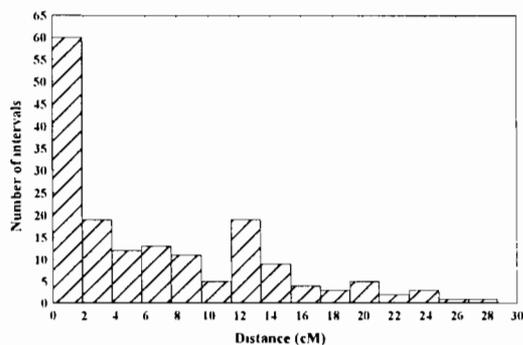
**Figure 1** Genetic linkage map constructed from 62 backcross hybrids derived from the interspecific cross [(PSE x DEW) x DEW], showing the location of 168 AFLP, 12 RFLP AFLP markers are described in terms of the three selective nucleotides used as follows: for example, "AACCTG8", the *Eco*RI-primer is given to the left (AAC), the *Mse*I primer to the right (CTG). RFLP markers are prefixed with "G" or "AR" for loci detected by Arabusta cDNA clone and "CA" for loci detected by random genomic clones of *C. arabica*. Recombination distances are given in cM (Kosambi units) on the left side of each linkage group (LG) and markers names are given to the right of each LG. Fourteen LG covering 1144 cM are indicated by letters A through N, ordered by their length. Loci marked \*, \*\*, \*\*\*, \*\*\*\*, and \*\*\*\*\* deviated significantly from (1 : 1) ratio at  $P < 0.01$ ,  $P < 0.001$ ,  $P < 0.0001$ ,  $P < 0.00001$  and  $P < 0.000001$  respectively. The lines specified by PSE or DEW correspond to distortion in favor of PSE and DEW respectively.

**GENETIC LINKAGE MAP**

PSE and DEW species have a strictly allogamous reproductive behaviour. Genitors are highly heterozygous and alleles segregate in the F1 generation. The segregation of 2 to 4 alleles per loci and the impossibility in parents to distinguish between coupling and repulsion phase makes genetic mapping more complicated than for autogamous species. Double pseudo-test cross is then recommended (GRATTAPAGLIA & SEDEROFF, 1994). Our approach constituted an alternative: we selected specific PSE markers so DEW gametes can be considered as genetically homozygous (double absence) for these markers. By contrast, the F1 hybrids generated different gametes segregating in BCDEW hybrids. Loci from parental or recombinant can be distinguished.

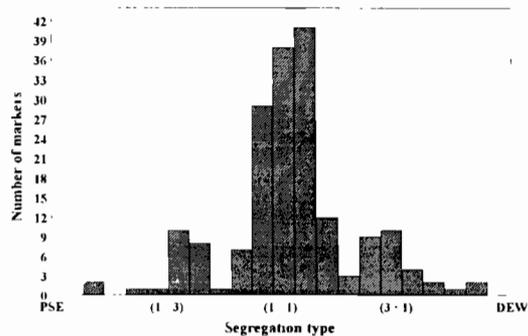
this number should be high to fill all the large intervals.

Eight AFLP clusters were observed. Such clusters are known to characterise centromeric regions (Q1, *et al.*, 1998; VUYLSTECKE, *et al.*, 1999) and this particularity was used to locate this region on the chromosome 3 of *Arabidopsis thaliana* (ALONSA-BLANCO, *et al.*, 1998). In addition, our markers underline: i) the very conserved parts of the PSE genome (no within-PSE polymorphism) and ii) the interspecific differentiation of the two genomes. Consequently, the centromeric markers should be markers of the speciation. Thus concern repeated sequences and could be implied in chromosome pairing. Their presence on several centromers should also allow to define specific markers for chromosome *in situ* hybridisation studies.



**Figure 2** Distribution of the distance between two successive markers

Fourteen linkage group were identified for 11 chromosomes, meaning that genome is not fully covered. Seven per cent of 2 successive marker intervals were higher than 20 cM, confirming that additional markers are needed to cover all the genome. The strong skewness of the distance distribution shows that



**Figure 3** Distribution of the 181 markers mapped in the BCDEW progeny in relation to the relative contribution of PSE. The figure clearly shows a predominant fraction of markers presenting a Mendelian segregation of (1 : 1), and two categories of segregation distortion following a (3 : 1) ratio in favor to PSE and a (1 : 3) ratio in favor to DEW.

The size difference between intraspecific (1402 cM) and interspecific (1144 cM) maps would result from

the lower genome homology between parents in an interspecific cross, reducing recombination and map size. For example, in potatoes, the map size of an interspecific cross was 65 % smaller than the intraspecific one (BONIERBALE, *et al.*, 1988; GEBHARDT, *et al.*, 1991). In coffee map size is 18 % smaller than the intraspecific one on *C. canephora*: 1144 cM (PAILLARD *et al.*, 1996). The reduce difference between map size in coffee trees could reflect an higher genome homology in the genus *Coffea* than in potatoes.

Among the 12 RFLP markers mapped, only one (CA 30) is common with the intraspecific map developed by PAILLARD *et al.* (1996), the other RFLP markers being homozygous in *C. canephora*.

#### SEGREGATION DISTORTION

With AFLP markers, co-migration and superimposition of non-allelic amplification products is possible (CLOUTIER, *et al.*, 1997) leading to pseudo-distortion of (3 : 1) ratio. Using our two-step mapping method, the mapping of pseudo-distorted markers is not likely. Nevertheless, the symmetric situation between the (1 : 3) and (3 : 1) ratios (Fig. 3) constitutes the main argument in favour of true distorted markers.

Skewed segregation ratios have been frequently reported in interspecific cross for all types of markers (morphological, isozyme, RFLP or AFLP). The percentage of loci showing segregation distortions was highly variable: 69 % in *Cryptomeria japonica* (NIKAIDO, *et al.*, 1999), 37% in *Citrus* (LURO, *et al.*, 1994), 36% in *Oryza* (XU, *et al.*, 1997; VIRK, *et al.*, 1998), 33% in *Prunus* (FOOLAD, *et al.* 1995), 23% in *Helianthus* (QUILLET, *et al.*, 1995), 8.4 % in *Lens* sp. (EUJAYL, *et al.*, 1998) and only 1.4% in *Hevea* spp. (LESPINASSE, *et al.*, 2000). Compared to these data, our distorted frequency appears to be intermediate (30%).

In the genus *Lycopersicum*, according to the wild species crossed with the cultivated *L. esculentum*, interspecific cross leads to distortion rates varying from 51 % to 80 %, (PATERSON, *et al.*, 1988; PATERSON, *et al.*, 1991; de VINCENTE and TANKSLEY, 1993). Our cross including two species among the most differentiated of the genus (LASHERMES, *et al.*, 1997), it should be interesting to observe the variation of this parameter when the genetic distance between parental species is lower.

Explanations given for such observed distortion are mostly gametic, zygotic or/ and post-zygotic selection but their implying to explain a symmetric and polymodal distribution, especially with (3 : 1) and (1 : 3) ratios is not plausible. By contrast, gene conversion implying DNA heteroduplex formation is known to give such ratios (LEBLON & ROSSIGNOL, 1973; NICOLAS & ROSSIGNOL, 1983; NAG, *et al.*, 1989). Post-meiotic segregation is another process leading to distortion segregation, but with (5 : 3) and (3 : 5)

ratios. A high rate of conversion and low rate of post-meiotic segregation reflect a high level of heteroduplex formation and efficient repair of the resulting mismatch (NAG, *et al.* 1989). This explanation fit well with our results. High rate of heteroduplex formation and mismatch are expected during meiosis pairing in interspecific hybrids. This phenomenon could be due to difference in DNA amount among parents as it is for DEW genome size which is 25% higher than PSE one (BARRE, *et al.*, 1996). The intergeneric cross *Citrus reticulata* x *Poncirus trifoliata* (tangerine tree x grapefruit tree) studied by LURO *et al.* (1994) showed 37% of distorted loci (for both RFLP and RAPD). Although not emphasised by authors, data clearly show to types of distorted markers: 9 markers in favour of *Poncirus* with  $p = 0.30$  and 29 markers in favour of *Citrus* with  $p=0.75$ . Again the  $p$  values fit well with (1 : 3) and (3 : 1) ratio, respectively. Obviously, gene conversion seems to characterise some interspecific cross and does not depends on the type of marker (AFLP, RFLP and RAPD). These results led us to make a software to test such ratios. MapDisto (the software) is available on the web: <http://www.mpl.ird.fr/~lorieux>.

#### PERSPECTIVES

The present work demonstrates the feasibility of an AFLP marker-based genetic linkage map of the diploid coffee genome. This map is a starting point for further molecular study in coffee genetics.

Firstly, this map could be used to identify quantitative trait loci (QTL) for highly discriminant and specific traits. Biochemical traits implied in the coffee cup quality will be interesting because of the difference between species. This is the case of the caffeine with 0% in PSE and 1% in DEW (BARRE, *et al.*, 1998), the chlorogenic acids with 1.2 % in PSE and 5.5 % in DEW (KY, *et al.*, 1999), and the sucrose (8 % in PSE and 3 % in DEW (KY, *et al.*, 1999). Identification and location of interspecific QTL will lead to use bordering molecular markers as tools for assisting the introgression of this interesting characteristic from wild coffee to cultivated coffee species (Molecular Assisted Selection).

Secondly, this map could be used to select AFLP markers for sequencing and to obtain co-dominant PCR markers. These markers will be used to compare our result with other interspecific maps currently in works, such as these three under studied crosses: *C. pseudozanguebariae* x *C. canephora*, *C. heterocalyx* x *C. canephora* and *C. eugenioides* x *C. canephora*. And finally, sequenced AFLP of either distorted loci or clustered loci will permits to approach meiotic pairing mismatch using *in situ* hybridisation.

#### REFERENCES

ALONSO-BLANCO, C., PEETERS, A.J.M., KOORNNEEF, M., LISTER, C., DEAN, C., VAN DEN BOSCH, N.,

- POT, J. & KUIPER M.T.R. 1998. Development of an AFLP based map of Ler, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a Ler/ Cvi recombinant inbred lines population. *Plant J.* 14: 259-271.
- BARRE, P. 1997. *Les transferts de gènes entre espèces de caféiers diploïdes : étude des hybrides F1 et BC1 entre C. pseudozanguabariae Bridson et C. liberica Bull ex Hiern (C. dewevrei De Wild et Th. Dur)*. PhD thesis, Ecole Nationale Supérieure Agronomique de Montpellier, France.
- BARRE, P., NOIROT, M., LOUARN, J., DUPERRAY, C. & HAMON, S. 1996. Reliable flow cytometric estimation of nuclear DNA content in coffee trees. *Cytometry* 24: 32-38.
- BARRE, P., AKAFFOU, S., LOUARN, J., CHARRIER, A., HAMON, S. & NOIROT, M. 1998. Inheritance of caffeine and heteroside contents in an interspecific cross between a cultivated coffee species *Coffea liberica* var 'dewevrei' and a wild species caffeine free *C. pseudozanguabariae*. *Theor. appl. Genet.* 96: 306-311.
- BONIERBALE, M.W., PLAISTED, R.L. & TANKSLEY, S.D. 1988. RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* 120: 1095-1103.
- BRIDSON, D.M. 1987. Nomenclatural notes on *Psilanthus*, including *Coffea* sect. *Paracoffea* (Rubiaceae tribe *Coffeae*). *Kew Bull.* 36 : 817-860.
- CARVALHO, A. & MONACO, L.C. 1967. Genetic relationships of selected *Coffea* species. *Ciencia e cultura* 19: 151-165.
- CHARRIER, A. 1978. La structure génétique des caféiers spontanés de la région malgache et leurs relations avec les caféiers d'origine africaine (*Eucoffea*). *Mém ORSTOM* 87: 1-223.
- CLOUTIER, S., CAPPADOCIA, M. & LANDRY, B.S. 1997. Analysis of RFLP mapping inaccuracy in *Brassica napus* L. *Theor. appl. Genet.* 95: 83-91.
- DE VICENTE, M.C. & TANKSLEY, S.D. 1993. QTL analysis of transgressive segregation in an interspecific tomato cross. *Genetics* 134: 585-596.
- DOLEZEL, J., BINAROVA, P., LUCRETTI, S. 1989. Analysis of nuclear DNA content in plant cells by flow cytometry. *Biol. Plant* 32: 113-120.
- DOYLE, J.J. & DOYLE J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19: 11-15.
- EUIJAYL, I., BAUM, M., POWELL, W., ERSKINE, W. & PEHU, E. 1998. A genetic linkage map of lentil (*Lens sp.*) based on RAPD and AFLP markers using recombinant inbred lines. *Theor. appl. Genet.* 97: 83-89.
- FOOLAD, M.R., ARULSEKAR, S., BECERRA, V. & BLISS, F.A. 1995. A genetic map of *Prunus* based on an interspecific cross between peach and almond. *Theor. appl. Genet.* 91: 262-269.
- GADISH, I. & ZAMIR, D. 1986. Differential zygotic abortion in an interspecific *Lycopersicum* cross. *Genome* 29: 156-159.
- GEBHARDT, C., RITTER, E., BARONE, A., DEBENER, T., WALKEMEIER, B., SCHACHTSCHABEL, U., KAUFMANN, H., THOMPSON, R.D., BONIERBALE, M.W., GANAL, M.W. & TANKSLEY, S.D. 1991. RFLP maps of potato and their alignment with the homologous tomato genome. *Theor. appl. Genet.* 83: 49-57.
- GRANT, V. 1975. *Genetics of flowering plants*. Columbia University Press, New York.
- GRATTAPAGLIA, D. & SEDEROFF, R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross mapping strategy and RAPD markers. *Genetics* 137: 1121-1137.
- KOSAMBI, D.D. 1944. The estimation of map distances recombination values. *Ann. Eugen.* 12: 172-175.
- KY, C.-L., LOUARN, J., GUYOT, B., CHARRIER, A., HAMON, S. & NOIROT, M. 1999. Relations between and inheritance of chlorogenic acids contents in an interspecific cross between *C. pseudozanguabariae* and *Coffea liberica* var 'dewevrei'. *Theor. appl. Genet.* 98: 628-637.
- KY, C.-L., DOULBEAU, S., GUYOT, B., AKKAFU, S., CHARRIER, A., HAMON, S., LOUARN, J., NOIROT, M. 1999. Inheritance of coffee bean sucrose content in the interspecific cross: *Coffea pseudozanguabariae* x *Coffea liberica* 'dewevrei'. *Plant breed.* (in press)
- LASHERMES, P., COMBES, M.C., TROUSLOT, P. & CHARRIER, A. 1997. Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. *Theor. appl. Genet.* 94: 947-955.
- LANDER, E.S., GREEN, P., ABRAHAMSON, J., BARLOW, A., DALY, M.J., LINCOLN, S.E. & NEWBURG, L. 1987. MAPMAKER: an interactive computer package for primary genetic linkage maps of experimental and natural populations. *Genomics* 1: 174-181.
- LEBLON, G. & ROSSIGNOL, J.L. 1973. Mechanism of gene conversion in *Ascobolus immersus*. The interaction of heteroalleles in the conversion process. *Mol. Gen. Genet.* 122: 165-182.
- LEROY, J.F. 1980. Evolution et taxogénèse chez les caféiers (*Coffea* L., *Psilanthus* Hook. f. et *Nostolachma* Durand). Hypothèse sur leur origine. *C. R. Acad. Sci.* 291: 593-596.
- LESPINASSE, D., RODIER-GOUD, M., GRIVET, L., LÉCONTE, A., LEGNATE, H. & SEGUIN, M. 2000. A saturated genetic linkage map of rubber tree (*Hevea* spp.) based on RFLP, AFLP, microsatellite, and isozyme markers. *Theor. appl. Genet.* 100: 127-138.
- LINCOLN, S., DALY, M. & LANDER, E. 1992. *Constructing genetic maps with MAPMAKER / EXP version 3.0*. Whitehead Institute Technical Report, 3<sup>rd</sup> edn
- LOUARN, J. 1992. *La fertilité des hybrides interspécifiques et les relations génomiques entre les caféiers diploïdes d'origines africaines (Genre Coffea L. sous-genre Coffea)*, Ph D Thesis, University of Orsay, France.

- LURO, F., LORIEUX, M., LAIGRET, F., BOVÉ, J.M. & OLLITRAULT, P. 1994. Cartographie du génome des agrumes à l'aide des marqueurs moléculaires et distorsions de ségrégation. *In*: Techniques et utilisations des marqueurs moléculaires, Montpellier, France, Ed. INRA, pp: 69-82.
- MEKSEM, K., LEISTER, D., PELEMAN, J., ZABEAU, M., SLAMINI, & GEBHARDT, C. 1995. A high-resolution map of the vicinity of the RI locus on chromosome V of potato based on RFLP and AFLP markers. *Mol. Gen. Genet.* 249: 74-81.
- NAG, D.K., WHITE, M.A. & PETES, T.D. 1989. Palindromic sequences in heteroduplex DNA inhibit mismatch repair in yeast. *Nature* 340: 318-320.
- NICOLAS, A. & ROSSIGNOL, J.L. 1983. Gene conversion: point-mutation heterozygosities lower heteroduplex formation. *EMBO J.* 2: 2265-2270.
- NIKAIDO, A., YOSHIMARU, H., TSUMURA, Y., SUYAMA, Y., MURAI, M. & NAGASAKA, K. 1999. Segregation distortion for AFLP markers in *Cryptomeria japonica*. *Gen. Genet. Syst.* 74: 55-59.
- PAILLARD, M., LASHERMES, P. & PÉTIARD, V. 1996. Construction of a linkage map in coffee. *Theor. appl. Genet.* 93: 41-47.
- PATERSON, A.H., LANDER, E.S., HEWITT, J.D., PETERSON, S., LINCOLN, S.E. & TANKSLEY, S.D. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335: 721-726.
- PATERSON, A.H., DAMON, S., HEWITT, J.D., ZAMIR, D., RABINOWITZ, H.D., LINCOLN, S.E., LANDER, E.S. & TANKSLEY, S.D. 1991. Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. *Genetics* 127: 181-197.
- QI, X., STAM, P. & LINDHOUT, P. 1998. Use of locus-specific AFLP markers to construct a high density molecular map in barley. *Theor. appl. Genet.* 96: 376-384.
- QUILLET, M.C., MADJIDAN, N., GRIVEAU, Y., SERIEYS, H., TERSAC, M., LORIEUX, M. & BERVILLÉ, A. 1995. Mapping genetic factors controlling pollen viability in an interspecific cross in *Helianthus* sect. *Helianthus*. *Theor. appl. Genet.* 91: 1195-1202.
- THOMAS, C.M., VOS, P., ZABEAU, M., JONES, D.A., NORCOTT, K.A., CHADWICK, B.P. & JONES, J.D.G. 1995. Identification of amplified restriction fragment polymorphism (AFLP) markers tightly linked to the tomato *Cf-9* gene for resistance to *Cladosporium fulvum*. *Plant J.* 8: 785-794.
- VIRK, P.S.; FORD-LLOYD, B.V. & NEWBURY, H.J. 1998. Mapping AFLP markers associated with sub-specific differentiation of *Oryza sativa* (rice) and an investigation of segregation distortion. *Heredity* 81: 613-620.
- VOS, P., HOGERS, R., BLEEKER, M., REIJANS, M., VAN DE LEE, T., HORNES, M., FRIJTERS, A., POT, J., PELEMAN, J., KUIPER, M. & ZABEAU, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23: 4407-4414.
- VUYLSTEKE, M., MANK, R., ANTONISE, R., BASTIAANS, E., SENIOR, M.L., STUBER, C.W., MELCHIN-GER A.E., LUBBERSTEDT, T., XIA, X.C., STAM, P., ZABEAU, M. & KUIPER, M. 1999. Two high-density AFLP linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. *Theor. appl. Genet.* 99: 921-935.
- XU, Y., ZHU, L., XIAO, J., HUANG, N. & MCCOUCH, S.R. 1997. Chromosomal regions associated with segregation distortion of molecular markers in F2, backcross, double haploids, and recombinant inbred populations in rice (*Oryza sativa* L.). *Mol. Gen. Genet.* 253: 535-545.
- ZABEAU, M. & VOS, P. 1993. *Selective restriction fragment amplification: a general method for DNA fingerprinting*. European Patent Application number: 92402629.7, Publication number 0 534 858 A1.
- ZAMIR, D. & TADMOR Y (1986) Unequal segregation of nuclear genes in plants. *Bot Gazette* 147: 355-35.

**Déterminisme génétique de quelques composés biochimiques de la graine de café vert impliqués dans la qualité à la tasse. Etude d'un croisement entre *Coffea pseudozanguebariae* et *C. liberica* var. *dewevrei*.**

**Résumé :**

La diversité et le déterminisme génétique de l'accumulation dans les grains du café vert de composés biochimiques majeurs impliqués dans la qualité ont été étudiés. Les composés concernés sont : les acides chlorogéniques, la caféine, la trigonelline et le saccharose. Deux types de matériel végétal ont été analysés. Le premier correspond à des accessions spontanées des espèces cultivées : *Coffea arabica* Linné et *C. canephora* Pierre. Elles ont permis de comparer la diversité biochimique naturelle à celle des cultivars. Le second est une descendance hybride, issue d'un croisement interspécifique entre une espèce sauvage, *C. pseudozanguebariae* Bridson (PSE) et une espèce anciennement cultivée, *C. liberica* var. *dewevrei* De Wild Th. Dur. (DEW). Dans cette étude nous avons abordé, la génétique quantitative, les corrélations phénotypiques et la localisation des QTLs des composés sur la carte génétique établie.

Le matériel spontané et le matériel hybride présentent une grande variabilité biochimique en comparaison aux cultivars. L'analyse en composantes principales pour étudier les corrélations phénotypiques constitue une approche originale pour préciser les voies de biosynthèse des acides chlorogéniques chez les caféiers.

La génétique quantitative des composés biochimiques révèle que la teneur en saccharose est un caractère additif (excepté chez les hybrides BC1 par PSE), très peu influencé par les effets environnementaux (86,5 % de la variance est d'origine génotypique). En revanche, l'hérédité de la teneur en trigonelline serait d'origine nucléaire et cytoplasmique. Les teneurs en acides chlorogéniques sont plus sensibles aux conditions environnementales : la part de variation d'origine génétique est de 43 à 78 %. L'additivité, avec ou sans transformation mathématique, est vérifiée pour les trois classes (CQA, DQA et FQA) et pour la majorité des isomères. Un gène majeur est responsable de l'absence de la teneur en 3-FQA.

La carte génétique établie sur les 62 hybrides backcross diploïdes ( $2x = 22$ ) [(PSE x DEW) x DEW], comprend 181 locus (167 AFLP, 13 RFLP et un marqueur morphologique) répartis sur 14 groupes de liaison et couvrant 1144 cM. Les distorsions de ségrégation affectent 30 % des marqueurs. Ces distorsions de types (3 : 1) et (1 : 3) suggèrent l'existence d'événements de conversion génétique.

La recherche de QTLs révèle l'existence de six régions de la carte génétique impliquées dans les variations des teneurs en acides chlorogéniques. Les teneurs en caféine et en trigonelline sont respectivement expliquées par quatre et un QTL. Les variations de teneur en saccharose sont associées à un marqueur non cartographié. D'autres QTLs d'intérêt agronomique ont été mis en évidence, comme le poids de cent grains (deux QTLs) et la durée du cycle de fructification.

Cette approche génétique de composés biochimiques de la graine de café vert a été discutée en rapport avec la physiologie et les voies de biosynthèse. Elle constitue un point de départ pour l'amélioration qualitative à la tasse de l'espèce de basse altitude *C. canephora*.

**Mots clés :** *Coffea*, croisement interspécifique, acides chlorogéniques, caféine, saccharose, trigonelline, cartographie génétique, AFLP, conversion génétique, Quantitative Trait Loci.

## Sigles des acides chlorogéniques

---

### Variables absolues

#### Acides chlorogéniques :

**CGA** : Ensemble des acides chlorogéniques (somme des trois classes)

**CGA1** : Ensemble des isomères possédant l'unité acide caféique (CQA + DQA)

#### Classe des CQA :

**3-CQA** : Acide caféyl-3 quinique

**4-&5-CQA** : Acide caféyl-4 quinique et en en acide caféyl-5 quinique

#### Classe des DQA :

**3,4-DQA** : Acide dicaféyl-3,4 quinique

**3,5-DQA** : Acide dicaféyl-3,5 quinique

**4,5-DQA** : Acide dicaféyl-4,5 quinique

#### Classe des FQA :

**3-FQA** : Acide férulyl-3 quinique

**4-FQA** : Acide férulyl-4 quinique

**5-FQA** : Acide férulyl-5 quinique

---

### Variables relatives

Ces variables correspondent à des rapports entre classes ou isomères.

Exemple :  $\% \text{ 3-CQA} = \text{3-CQA} / \text{CQA}$

---

**Chaîne de biosynthèse au verso →**

