

1997-1998

UNIVERSITE MONTPELLIER II
SCIENCES ET TECHNIQUES
DU LANGUEDOC

UNIVERSITE
DE
PERPIGNAN

ECOLE NATIONALE
SUPERIEURE AGRONOMIQUE
DE MONTPELLIER

DIPLOME D'ETUDES APPROFONDIES

BASES DE LA PRODUCTION VEGETALE

Option « Biotechnologie et Amélioration des Plantes »

MISE EN EVIDENCE ET CARACTERISATION DE TROIS SEQUENCES

REPETEES AU SEIN DE :

Coffea pseudozanguebariae Bridson et *Coffea liberica* var. *dewevrei* De Wild et Th. Dur.

Marie-Flore CELARIER

Laboratoire GeneTrop – Centre Orstom de Montpellier

Soutenu le 7 Septembre 1998 devant la commission d'examen :

A. BERVILLE
J.F BRIAT
F. CASSE
A. CHARRIER
M. COUMANS
A. DELCAMP
M. DELSENY

C. GRIGNON
S. HAMON
Ph. JOUDRIER
M. LEBRUN
J.J MACHEIX
N. MICHAUX-FERRIERE
D. MOUSAIN

M. NOIROT
G. PELTIER
A. PALLOIX
P. ROBIN
F. TARDIEU
F. WARENBOURG
J. WERY

REMERCIEMENTS

Je remercie *Serge Hamon* et *Michel Noirot* de m'avoir accueillie dans leur équipe, de m'avoir proposé ce sujet sur les séquences répétées de caféiers et de m'avoir permis de le réaliser

Je remercie également tous les membres de GeneTrop pour leur sympathie ; en particulier, *Marie-Christine Combes* pour son encadrement technique, *Philippe Barre*, *Philippe Lashermes*, *Isabelle Rimbault* et *Pierre Trouslot* pour leur aide occasionnelle mais efficace.

Je remercie tout spécialement les personnes du CIRAD qui ont contribué à cette étude : *Karine Boisvin* pour la banque d'ADN, *Marie-France Duval* pour les RFLP "sondes chaudes", *Marion Layssac* et *Angélique d'Hont* pour l'hybridation *in situ*.

La liste des remerciements ne serait pas complète sans citer *Jean-Louis Noyer* et *Franck-Christophe Baurens*, sur qui j'ai toujours pu compter dans toutes les circonstances. Je les remercie pour tous les conseils et toute l'aide qu'ils ont pu m'apporter.

DEA Bases de la Production Végétale

Mise en évidence et caractérisation moléculaire de trois séquences répétées au sein de *Coffea pseudozanguebariae* et *Coffea liberica* var. *dewevrei*.

Résumé

Afin d'étudier les possibilités offertes par l'hybridation interspécifique chez *Coffea canephora*, un croisement modèle entre deux espèces phylogénétiquement éloignées est étudié : *C. pseudozanguebariae* (PSE) × *C. liberica* var. *dewevrei* (DEW), proche de *C. canephora*. La quantité d'ADN nucléaire des deux espèces est respectivement de 1,13pg et de 1,43pg. Elle peut s'expliquer par des différences de constitution chromosomique principalement au sein de séquences non codantes, essentiellement répétées.

La construction d'une banque a été préférée à des méthodes plus sélectives de façon à couvrir le génome en entier. A partir d'une banque d'ADN génomique d'hybride F1 restreint aux sites GATC, par l'endonucléase Sau3A, 193 clones contenant des séquences répétées ont été retenus, après rétrohybridation. Vingt deux sondes sont totalement séquencées, 5 partiellement. Elles sont riches en AT et montrent souvent des sites GATC internes. La plupart ne présente pas d'homologie avec les séquences connues.

Trois sondes sont analysées de façon plus approfondie sur les espèces parentes PSE et DEW et sur 8 autres espèces choisies pour leur origine géographique et leur quantité d'ADN (*C. racemosa*, *C. sessiliflora*, *C. eugenioides*, *C. sp. "Moloundou"*, *C. canephora*, *C. congensis* et *C. heterocalyx*). La première est constituée de quatre séquences distinctes, dont une au moins répétée en tandem. La seconde révèle une forte homologie avec des séquences codant pour la reverse-transcriptase de rétrotransposon de type gypsy-like, tant au niveau nucléotidique qu'au niveau acides aminés. Elle est donc, le premier rétrotransposon découvert chez les caféiers. La dernière possède des profils RFLP sur l'ADN différents uniquement pour les deux espèces à forte quantité d'ADN.

Il est primordial de poursuivre l'étude du rétrotransposon pour préciser son nombre de répétitions, sa structure fine et sa capacité à transposer. Il est aussi nécessaire de poursuivre, selon la même stratégie, l'étude de la banque. La priorité devra être donnée à l'évaluation du nombre de répétitions des séquences par génome et à leur localisation sur les chromosomes. Les séquences répétées de la banque pourront, entre autres, contribuer aux études phylogénétiques, d'ores et déjà corrélées par les résultats RFLP obtenus.

Mots clés : *Coffea* • Séquences répétées • Rétrotransposon

And molecular characterization of three repeated sequences of *Coffea pseudozanguebariae* and *Coffea liberica* var. *dewevrei*.

Abstract.

In order to study the possibilities for interspecific hybridization using *Coffea canephora*, an interspecific model cross between two phylogenetically distant species was studied : *C. pseudozanguebariae* (PSE) × *C. liberica* var. *dewevrei* (DEW). The latter species belongs to the same group as *C. canephora*. The nuclear DNA contents of the two species are respectively 1,13 pg and 1,43 pg. This can be explained by differences in their chromosomal composition, mostly attributable to repeated sequences.

As it covers the whole genome, a genomic DNA library was constructed rather than using more selective methods. 193 clones of repeated sequences were selected by hybridization to F1 genomic DNA restricted by the Sau3A restriction enzyme (GATC). 22 probes were totally sequenced, 5 partially. They are rich in A-T residues, and often have internal GATC sites. The majority of them have no homology with known sequences.

Three probes were analysed in the parental species PSE and DEW and in 8 other chosen species chosen on the basis of their geographical origins and their DNA content (*C. racemosa*, *C. sessiliflora*, *C. eugenioides*, *C. sp. "Moloundou"*, *C. canephora*, *C. congensis* et *C. heterocalyx*).

The first probe was found to consist of 4 distinct regions. One of them is tandemly repeated. The second sequence is strongly homologous to the reverse-transcriptase of "gypsy-like" retrotransposons, in terms of both nucleotide sequence and deduced sequence of the protein it encodes. It is the first retrotransposon to be discovered in *Coffea*. The RFLP profiles identified using the third probe differ only for the 2 species with greater DNA content.

The study of this retransposon must be continued in order to determine their copy number, their ultrastructure and their capacity to transpose. The study of the genomic library must be continued too. The evaluation of copy number and *in situ* localization are of particular priority. The repeated sequences could contribute to phylogenetic studies, as results already suggested.

Key words : *Coffea* • Repeated sequences • Retrotransposon

Mise en évidence et caractérisation moléculaire de trois séquences répétées au sein de *Coffea pseudozanguebraia* et de *Coffea liberica* var. *dewevrei*

Résumé

Afin d'étudier les possibilités offertes par l'hybridation interspécifique chez *Coffea canephora*, un croisement modèle entre deux espèces phylogénétiquement éloignées est étudié : *C. pseudozanguebariae* (PSE) × *C. liberica* var. *dewevrei* (DEW), proche de *C. canephora*. La quantité d'ADN nucléaire des deux espèces est respectivement de 1,13pg et de 1,43pg. Elle peut s'expliquer par des différences de constitution chromosomique principalement au sein de séquences non codantes, essentiellement répétées.

La construction d'une banque a été préférée à des méthodes plus sélectives de façon à couvrir le génome en entier. A partir d'une banque d'ADN génomique d'hybride F1 restreint aux sites GATC, par l'endonucléase Sau3A, 193 clones contenant des séquences répétées ont été retenus, après rétrohybridation. Vingt deux sondes sont totalement séquencées, 5 partiellement. Elles sont riches en AT et montrent souvent des sites GATC internes. La plupart ne présente pas d'homologie avec les séquences connues.

Trois sondes sont analysées de façon plus approfondie sur les espèces parentales PSE et DEW et sur 8 autres espèces choisies pour leur origine géographique et leur quantité d'ADN (*C. racemosa*, *C. sessiliflora*, *C. eugenioides*, *C. sp. "Moloundou"*, *C. canephora*, *C. congensis* et *C. heterocalyx*). La première est constituée de quatre séquences distinctes, dont une au moins répétée en tandem. La seconde révèle une forte homologie avec des séquences codant pour la reverse-transcriptase de rétrotransposon de type gypsy-like, tant au niveau nucléotidique qu'au niveau des acides aminés. Elle est le premier rétrotransposon découvert chez les caféiers. La dernière possède des profils RFLP sur l'ADN, différents uniquement pour les deux espèces à forte quantité d'ADN.

Mots-clés *Coffea* • Séquences répétées • Rétrotransposon

Introduction

Deux espèces de caféiers sont principalement exploitées : *C. arabica* et *C. canephora* (Berthaud et Charrier 1988). L'*arabica* est cultivé sur les hauts plateaux d'Afrique et d'Amérique du Sud. Il produit un café aromatique de bonne qualité. En revanche, il est sensible aux nématodes et aux champignons tels *Hemileia vastatrix* responsable de la rouille orangée ou *Colletotrichum coffeanum* provoquant l'anthracnose des baies. Le *robusta* (*C. canephora*) est un caféier de basse altitude, vigoureux et résistant bien à ces maladies. Les grains, d'une moindre valeur commerciale, produisent un café plus amer, moins aromatique et plus riche en caféine (Barrel et Jacquet 1994). L'amélioration des qualités gustatives du café produit par *C. canephora* s'appuie sur la diversité génétique et physiologique des espèces du genre *Coffea*. Les caféiers diploïdes africains ($2n=2x=22$) se répartissent en trois groupes phylogénétiques (Cros 1994). Ils sont susceptibles de tous s'hybrider, toutefois, les hybrides de première génération (F1) sont souvent peu fertiles. Leur fertilité est corrélée positivement au nombre de bivalents formés lors de la méiose de ces hybrides (Louarn 1992). Ce nombre augmente d'autant plus que les parents ont une taille de génome proche et qu'ils appartiennent au même groupe phylogénétique (Charrier 1978; Louarn 1992; Barre 1997).

Les possibilités d'amélioration de *C. canephora* par hybridation interspécifique sont étudiées grâce à un croisement modèle entre deux espèces interfertiles parmi les plus éloignées phylogénétiquement : *C. pseudozanguebariae* (PSE) × *C. liberica* var.

dewevrei (DEW). DEW appartient au même groupe que *C. canephora* et diffère de PSE par la morphologie de l'arbre, la phénologie de la fructification, la composition biochimique des grains de café et la quantité d'ADN nucléaire (annexe 1). PSE possède une taille de génome de $2C=1,13\text{pg}$ et DEW de $2C=1,43\text{pg}$. La taille du génome est un caractère additif, aussi bien en F1 que dans les populations de rétrocroisement (Barre *et al.* en cours de publication). Chez ces individus, une relation linéaire négative existe entre la quantité d'ADN et le nombre de chromosomes de PSE (Barre *et al.* 1998). Le niveau de ploïdie ($2x$) et le nombre de chromosomes ($2n=22$) restent identiques chez les parents, les individus F1 et les individus de rétrocroisement. Cependant, la quantité d'ADN des individus issus de chaque rétrocroisement, varie d'un extrême à l'autre (de $1,13\text{pg}$ à $1,43\text{pg}$).

D'une manière générale, les variations de quantités d'ADN peuvent s'expliquer par des différences de constitution génomique. Elles portent principalement sur les séquences non codantes formées essentiellement de séquences répétées. Celles-ci couvrent une large proportion des génomes eucaryotes : 98% chez l'Homme (Davidson et Britten 1973; Schmid et Deninger 1975), 95% chez l'oignon (Thompson et Murray 1981). Même au sein des petits génomes comme celui d'*Arabidopsis thaliana*, elle atteint 20% (Thompson *et al.* 1995).

Les séquences répétées sont présentes à raison de 10^5 à 10^6 copies par génome (Lapitan 1992; Charlesworth *et al.* 1994). Elles possèdent une cinétique de renaturation très rapide avec un $Cot_{1/2} \approx 10^{-3}$ par rapport aux séquences uniques ($Cot_{1/2} \approx 10^3$). Leur unité de répétitions va d'un nucléotide à plusieurs kilobases. Elles sont organisées soit en tandem (à la queue-leu-leu) dans la même orientation, inversées ou en mélange, soit dispersées dans le génome (Flavell *et al.* 1977; Singer 1982).

Les séquences répétées en tandem de type ADN satellites (nano, micro, mini et midi-satellites) sont constituées d'un motif monotone répété n fois. Ces répétitions, d'un à quelques centaines de nucléotides, évoluent rapidement et comprennent de légères variations de séquences (Jeffrey *et al.* 1991). Les séquences en tandem comprennent également les séquences conservées comme les séquences télomériques (Fajkus *et al.* 1995) et celles codant pour les sous-unités ribosomiques (Lapitan 1992). Ces dernières comprennent un espaceur de gène (ITS) hypervariable très largement exploité en phylogénie. C'est le cas chez les caféiers (Lashermes *et al.* 1997).

Les séquences répétées dispersées regroupent les SINE's (Short Interspersed Elements) et les LINE's (Long Interspersed Elements). Les premières ont une taille de quelques centaines de paires de bases, comme la séquence Alu, mise en évidence chez l'homme (Deiniger *et al.* 1989) ou les séquences Alu-like, découvertes chez le bananier (Baurens *et al.* 1998). Les secondes, de plusieurs kilopaires de bases comprennent les transposons et rétrotransposons, qui peuvent s'exciser ou s'insérer dans le génome après répliation (Mac Clintock 1951; Gierl et Saedler 1992; Grandbastien 1992).

Toutes les séquences répétées jouent un rôle important au sein du génome en termes d'expression, d'évolution et de structure. Les séquences télomériques ont un rôle protecteur contre le vieillissement cellulaire (Zakian 1989; Harley *et al.* 1990; Hastie *et al.* 1990). Les transposons et rétrotransposons provoquent des duplications et des mutations participant à l'évolution du génome (Bennetzen 1996). L'ADN satellite, en particulier α -satellite, participerait à la mise en place d'un code de repliement chromosomique tridimensionnel (Vogt 1990).

Notre travail contribue à l'étude du croisement modèle PSE \times DEW. Une banque d'ADN génomique a été constituée. Trente sondes répétées ont été séquencées et comparées aux banques de données. Parmi elles, trois ont été

sélectionnées afin de déterminer les profils RFLP qu'elles engendrent et leur groupe d'appartenance (en tandem ou dispersées).

Matériels et méthodes

Matériel végétal

Les feuilles de l'individu F1, à l'origine de la banque d'ADN génomique, proviennent de caféiers de la collection de la station de Man (Côte d'Ivoire). Les autres espèces sont maintenues en collection dans les serres de l'ORSTOM à Montpellier (France). La température en serre est stabilisée à 24°C le jour et 18°C la nuit, l'humidité relative de 70%. En plus des espèces parentales PSE et DEW, 8 autres espèces ont été choisies selon leur origine géographique et leur quantité d'ADN (Tableau 1). Les feuilles et les racines sont récoltées sur des caféiers adultes. L'ADN est extrait à partir de feuilles récoltées sur 6 caféiers PSE et 6 caféiers DEW individuellement identifiés. L'ADN des 8 espèces représentant les trois groupes de caféiers diploïdes africains est extrait à partir d'un mélange de feuilles provenant d'une dizaine d'individus de chaque espèce. Les feuilles sont prélevées sans distinction d'âge mais dans un bon état phytosanitaire.

Banque d'ADN

Une banque d'ADN génomique est constituée à partir d'ADN d'un hybride F1 (PSE×DEW) restreint par l'enzyme Sau3A. Son site de restriction est à quatre paires de bases 5'-GATC-3'. Sau3A a été choisie pour sa fréquence élevée de coupures. Les fragments obtenus, entre 500 et 2500 pb, sont clonés dans des plasmides pUC19 ouverts par BamHI permettant ainsi, la reconstruction du site Sau3A après ligation. La transformation est réalisée dans des bactéries "super-compétentes" DH5 α (Baurens, 1997). Le taux d'efficacité est de $3,5 \times 10^7$ colonies transformées par microgramme d'ADN. Les plasmides sont extraits des bactéries à l'aide du kit Wizard®plusSV de Promega®. Les sondes sont amplifiées par PCR (Polymerase Chain Reaction) en utilisant les amorces universelles M13 et Reverse. Les conditions d'amplification sont : 4min à 94°C, 35 \times [1min30 94°C, 1min30 50°C, 2min 72°C], 4min 72°C. Des membranes avec l'ADN des sondes sont constituées et rétrohybridées avec de l'ADN d'individu F1 "cassé" mécaniquement. Les sondes sont considérées comme étant des séquences répétées si le signal d'hybridation

est intense. Sur les 4200 colonies bactériennes obtenues, 193 clones comprenant des séquences répétées constituent la banque proprement dite. Elles sont identifiées par le code pCaORS suivi d'un numéro.

Hybridation RFLP

Après extraction, puis lyse des noyaux selon le protocole de Barre (1997), l'ADN est précipité, puis repris dans du tampon TRIS-EDTA. La concentration en ADN des solutions est estimée sur gel d'agarose à 0,8%. Huit μ g d'ADN de chaque échantillon sont restreints par Sau3A, HaeIII ou Tru9I selon les recommandations du fournisseur. Sau3A (5'-GATC-3') sert de témoin et permet de retrouver la taille du fragment hybridé. HaeIII (GG $^\circ$ CC) et Tru9I (T $^\circ$ TAA) coupent fréquemment dans des zones riches en GC (HaeIII) ou en AT (Tru9I).

Les dépôts sont réalisés sur un gel à 1% d'agarose de dimensions 20cm \times 30cm. La migration a lieu dans du tampon Tris-Borate-EDTA (0,09M; 0,089M; 0,002M pH8), pendant 16h à 2,33V/cm. Après dépurination à l'HCl 0,25N pendant 30min puis équilibrage au NaOH 0,4N, un transfert d'ADN de type Southern est effectué sur une membrane de Nylon chargée positivement (Boehringer Mannheim). Les membranes sont hybridées par des sondes marquées à la digoxygénine-d-UTP par amplification avec déplacement de brin (Nick-translation). Le lavage final de stringence est réalisé à 65°C pendant 30min dans une solution de 0,1% SDS et 0,1X SSC. Des contrôles sont effectués en hybridant les membranes avec des sondes marquées au d-CTP 32 P par amorçage au hasard (random priming labelling). Après une exposition de 5 heures pour les sondes froides et d'une dizaine d'heures pour les sondes radioactives, les films (Boehringer Mannheim) sont révélés, puis analysés. Afin d'éliminer l'hypothèse d'une hybridation avec le plasmide et de vérifier l'efficacité de la dés hybridation, les trois membranes sont hybridées avec du pUC19 natif.

Tableau 1. Espèces de caféiers diploïdes africains avec leur quantité d'ADN (pg) et leur groupe géographique

| Afrique de l'Est | Hauts plateaux | Afrique centrale et Ouest | |
|-----------------------------|--------------------------|---------------------------|----------------------|
| 0,9pg à 1,2 pg | 1,4pg à 1,5pg | 1,4pg à 1,5pg | 1,7pg |
| <i>C.pseudozanguebariae</i> | <i>C.eugenioides</i> | <i>C.liberica</i> var | <i>C.humilis</i> |
| <i>C.racemosa</i> | <i>C.sp.</i> "Moloundou" | <i>Dewevrei</i> | <i>C.heterocalyx</i> |
| <i>C.sessiliflora</i> | | <i>C.canephora</i> | |
| | | <i>C.congensis</i> | |

Séquençage.

Trente sondes au total sont séquencées (Génomexpress S.A.) : 10 ont été choisies pour leur taux, supposé élevé, de répétitions (pCaORS006, 008, 012, 033, 035, 036, 038, 040, 054, 055). Les 20 autres ont été prises au hasard (pCaORS001, 002, 003, 004, 005, 007, 009, 010, 011, 013, 014, 015, 016, 017, 018, 019, 020, 021, 022, 023). L'analyse des séquences permet d'éliminer les bases correspondant au polylinker de pUC19 et de rechercher les sites internes de coupure des trois enzymes précédemment citées.

Les séquences sont comparées à celles des banques de données : EPLN regroupant les séquences nucléiques de plantes et la totalité des banques de séquences protéiques (Non-redundant Genbank CDS). FASTA3 (Pearson et Lipman 1988) est l'algorithme de recherche utilisé en libre accès sur le réseau Internet : <http://www2.ebi.ac.uk/fasta3>. Les résultats sont fournis sous forme de propositions associées à des probabilités. Plus la probabilité est faible, meilleure est l'homologie entre la séquence soumise et la séquence de la banque. Les logiciels fournissent aussi les alignements de nucléotides (ou d'acides aminés) et les références bibliographiques des séquences homologues.

Résultats

A partir de l'ADN génomique d'un hybride F1 (PSE×DEW), une banque de 193 clones sélectionnés pour contenir des séquences répétées, a été réalisée. Trente sondes ont été séquencées. Dix ont été choisies pour leur taux, supposé élevé, de répétitions et 20 autres ont été prises au hasard.

Séquençage et homologie des séquences.

Les séquences des sondes de pCaORS002 à pCaORS022 et de pCaORS033, 035, 036, 038 sont connues en totalité. Le séquençage de pCaORS006, 013, 040, 054, et 055 est partiel tandis que celui de pCaORS001, 016 et 023 a échoué. Le séquençage est de bonne qualité, toutes les bases sont identifiées.

Les inserts ont une taille allant de 47 à plus de 925 paires de bases (pb) (Tableau 2). Elles possèdent, pour la plupart, des sites internes de coupure des trois enzymes. Les sites ∇ GATC peuvent avoir 3 origines différentes : i) soit ce sont des sites qui n'ont

pas été restreints par Sau3A; ii) soit ils correspondent, en réalité à des sites de coupure d'iso-schizomères de Sau3A comme par exemple MboI ou Xho II ; iii) soit ce sont des sites reconstitués après religation de plusieurs fragments formant des séquences chimériques. Ne possédant aucun moyen de distinguer ces trois hypothèses, tous les sites ∇ GATC sont considérés *a priori* comme sites Sau3A et les séquences comme chimériques. Les séquences non chimériques ont une taille variant de 14 à plus de 818 pb. Les séquences étudiées sont plus riches en A-T qu'en G-C.

Aucune des séquences n'a montré d'homologie forte avec des séquences présentes dans la banque EPLN (score FASTA $<10^{-2}$), à l'exception de pCaORS0017 et 17-1 qui semblent homologues à des séquences de rétrotransposon (score FASTA de 0,0024 pour le rétrotransposon R1e de *Zea mays*, 0,043 pour un rétrotransposon gypsy/Ty3 d'*Ananas comosus*).

Sur les trois types de membranes RFLP, les 9 sondes (pCaORS006, 008, 012, 017, 033, 035, 036, 038, et 040) s'hybrident avec l'ADN de tous les individus PSE, DEW et avec l'ADN des autres espèces de caféiers. L'intensité du signal est variable. Aucun polymorphisme intraspécifique n'apparaît au sein des 6 individus DEW. Pour les 6 individus PSE, les profils RFLP sont identiques pour 5 d'entre eux. L'individu H58 diffère à cause d'une restriction incomplète de son ADN.

Tableau 2. Principales caractéristiques des sondes séquencées

| Sondes | Séquences entières | Taille des fragments (pb) | Sites Hae III | Sites Tru 91 | Pourcentage en AT (%) |
|-----------|--------------------|---------------------------|---------------|--------------|-----------------------|
| pCaORS002 | oui | - | - | - | - |
| 2-1 | | 33 | 0 | 0 | 58 |
| 2-2 | | 14 | 0 | 0 | 50 |
| pCaORS003 | oui | 146 | 1 | 1 | 52 |
| pCaORS004 | oui | 265 | 0 | 3 | 64 |
| pCaORS005 | oui | 112 | 0 | 1 | 62 |
| pCaORS006 | non | >923 | - | - | - |
| 6-1 | | 164 | 0 | 0 | 62 |
| 6-2 | | 61 | 1 | 0 | 60 |
| 6-3 | | 344, 171 | 2 | 0 | 59 |
| 6-4 | | 183 | 0 | 0 | 65 |
| pCaORS007 | oui | - | - | - | - |
| 7-1 | | 153 | 0 | 2 | 64 |
| 7-2 | | 140 | 1 | 0 | 55 |
| pCaORS008 | oui | - | - | - | - |
| 8-1 | | 128 | 1 | 2 | 71 |
| 8-2 | | 473 | 0 | 1 | 61 |
| pCaORS009 | oui | 70 | 0 | 0 | 53 |
| pCaORS010 | oui | 186 | 0 | 2 | 63 |
| pCaORS011 | oui | - | - | - | - |
| 11-1 | | 53 | 0 | 1 | 60 |
| 11-2 | | 185 | 0 | 0 | 53 |
| pCaORS012 | oui | - | - | - | - |
| 12-1 | | 336 | 0 | 4 | 62 |
| 12-2 | | 185 | 2 | 2 | 60 |
| 12-3 | | 35 | 0 | 0 | 54 |
| pCaORS013 | non | >701 | - | - | - |
| 13-1 | | 136 | 0 | 0 | 68 |
| 13-2 | | 25 | 0 | 0 | 60 |
| 13-3 | | 206,334 | 0 | 1 | 68 |
| pCaORS014 | oui | 254 | 0 | 1 | 66 |
| pCaORS015 | oui | - | - | - | - |
| 15-1 | | 29 | 0 | 0 | 52 |
| 15-2 | | 132 | 0 | 4 | 63 |
| pCaORS017 | oui | - | - | - | - |
| 17-1 | | 116 | 0 | 0 | 65 |
| 17-2 | | 43 | 0 | 0 | 62 |
| pCaORS018 | oui | 90 | 0 | 0 | 60 |
| pCaORS019 | oui | 120 | 0 | 0 | 60 |
| pCaORS020 | oui | 182 | 0 | 1 | 58 |
| pCaORS021 | oui | - | - | - | - |
| 21-1 | | 45 | 0 | 0 | 58 |
| 21-2 | | 58 | 0 | 0 | 57 |
| 21-3 | | 132 | 0 | 3 | 70 |
| 21-4 | | 143 | 0 | 1 | 60 |
| pCaORS022 | oui | 104 | 0 | 0 | 53 |
| pCaORS033 | oui | - | - | - | - |
| 33-1 | | 410 | 0 | 6 | 65 |
| 33-2 | | 163 | 0 | 0 | 52 |
| pCaORS035 | oui | - | - | - | - |
| 35-1 | | 66 | 0 | 0 | 66 |
| 35-2 | | 202 | 0 | 0 | 68 |
| pCaORS036 | oui | - | - | - | - |
| 227 | | 227 | 3 | 0 | 67 |
| 623 | | 623 | 0 | 6 | 61 |
| pCaORS038 | oui | 444 | 0 | 0 | 65 |
| pCaORS040 | non | >818 | 1 | 5 | 60 |
| pCaORS054 | non | >448 | 0 | 3 | 61 |
| pCaORS055 | non | - | - | - | - |
| 55-1 | | 349 | 0 | 3 | 65 |
| 55-2 | | 216 | 1 | 1 | 62 |

Etude particulière de trois sondes

Trois sondes sont analysées de manière plus approfondie : i) pCaORS006 est choisie pour son profil d'hybridation RFLP en échelle, reflet d'une séquence répétée en tandem; ii) pCaORS17 montre une forte homologie avec des séquences connues (rétrotransposon); iii) pCaORS38 possède des profils d'hybridation à une seule bande sur les trois membranes.

Sonde pCaORS006

La taille de cette sonde est supérieure à 923 pb (Tableau 2). Elle possède trois sites Hae III et trois sites Sau3A internes. *A priori*, elle est constituée d'au moins quatre séquences différentes associées en une chimère (Fig. 1).

GATCATTCTTCTTTTGAGTTTCTTGACACATTGGTCCAT
GGATGACCCCCACACCCCTTTGGCTTTGCTTAGTTAGCTCTT
TCATCTATTGACACGCATGAGTTTCATCAGAAATTATTTA
CACTACCTAGCATACTTGCAACATGTCATCACTTCATTATA
[164 pb]
GATCAAGCTTACGAAATTCGTAAGGAATGATAGGCCTCTG
AGTGATGAATCCTTTTTGTGA [61 pb]
GATCAATTCGAACCTATCTCTGGCCTCGTCACCTATAAAC
TTTGGCAATGGGGTAGGTTCTTGATAGGTTGAGCCTCATG
TGACGGCCCCACCTCCCCCTAGGGCGTACTCCAGGTTTCG
GCAGGTCGCTTGCCCAACTCTTGCCAGGACTCACTCACTC
GTATCACGTGCATACATCCATAGACAATAAATAACGACAC
AATCAAACTTAGAATTTACATCGATGCATAATCAATATA
CCAAGTCTCTATATACCCCAACAGGTTCAAAGTGATACA
ATCCAAGTACAATCATCCTAATCGAGTGATAATTGCGAGTACAAG
TCAAAAGTCAAAACAAC T*****
****AGTATGAAAAGTTCATAAAAGTGAGCCATAA TACTTC
AAGTAGCAAATCTCAAAGTGAGCGAGTG TAAAAGTTTCA
AAAGAAACAATAATCCAATAAACAATAATCACTCTAAAAGT
ATAAGGATACGGATAGCTCTCAGGAGCCAAATCCCATTC
CTTCATTAGAGCTT [344pb***171 pb]
GATCTAGTGACTTGGATATCTTTGTTAGTTTACAGAATTC
GTCACTTGGGAGGCTAATTGGGTTCTATGAATCTAAATGT
GCCTGCTCTTTATTAGCTAGGAATGCTAGTTCTATTTCAA
ATATTTGTAGTTGGAAAATTTCTCCACCATATTTCTTTCTT
CCGTAATTA CTGCAGATC [183 pb]

Fig. 1. Séquence pCaORS006 avec les sites enzymatiques Sau3A et HaeIII. Légende : les sites de coupure Sau3A sont représentés en gras, ceux de HaeIII en italique. ** : nucléotides non séquencés

L'hybridation de la sonde, non digérée, sur les trois membranes révèle l'existence d'une échelle de plusieurs bandes sur les trois membranes avec l'ADN des individus DEW. Ce type de profil peut correspondre à la présence de séquences répétées en tandem. La sonde étant chimérique, le fragment responsable de ce profil est inconnu.

Les différents fragments sont donc séparés par électrophorèse sur gel d'agarose "low melting" à 1%, après amplification de pCaORS006 et digestion par Sau3A. Quatre niveaux de bandes sont observés : autour de 650 pb, de 180 pb (associant deux bandes) et vers 60 pb. Ces quatre niveaux correspondent aux quatre fragments différents déterminés par séquençage (figure1). La bande à 650 pb comprendrait le fragment central (6-3) partiellement séquencé (344 pb + 171 pb = 515 pb). Les deux bandes proches de 180 pb regrouperaient les fragments 6-1 (164 pb) et 6-4 (183 pb). La bande de 60 pb environ contiendrait en plus du fragment 6-2 de la sonde, les fragments amplifiés du polylinker de pUC 19 (60 pb et 39 pb environ). En dehors du fragment central, il est difficile de séparer les autres fragments les uns des autres. De ce fait, 6-3 est hybridé en premier sur les trois membranes. Les résultats montrent que ce fragment est effectivement responsable de l'organisation en échelle sur l'ADN de DEW. Pour les autres espèces, les profils ne sont pas aussi structurés.

La comparaison de pCaORS006 ne révèle aucune homologie avec les séquences de la banque EPLN, de même que les comparaisons fragment par fragment (tableau 3).

Tableau 3. Comparaison des différents segments aux banques de données

| Nom du fragment | Taille en pb | Meilleur score FASTA | Description de la séquence homologue | Pourcentage d'identité (%) | Nombre de nts. identiques |
|-----------------|--------------|----------------------|---|----------------------------|---------------------------|
| 6 | 923 | 1,2 | ARNm de <i>Limnanthes douglasii</i> (1486 pb) | 66 | 106 |
| 6-1 | 168 | 0,47 | ARNm de <i>Limnanthes douglasii</i> (1486 pb) | 67 | 94 |
| 6-2 | 65 | 0,032 | ADN chromosomique d' <i>Arabidopsis thaliana</i> (56890 pb) | 69 | 61 |
| 6-3 | 344 | 0,45 | ARM d'acyltransférase de <i>Solanum tuberosum</i> (4153 pb) | 60 | 178 |
| 6-4 | 183 | 0,47 | ADN chromosomique d' <i>Arabidopsis thaliana</i> (74805 pb) | 64 | 90 |

nts : nucléotides

Sonde pCaORS017

pCaORS017 a une taille de 159 pb. Elle comporte un site interne Sau3A mais aucun site HaeIII ou Tru9I. La présence du site Sau3A permet à nouveau, de faire l'hypothèse de la nature chimérique de la sonde: 17-1 avec 116 pb et 17-2 avec 43 pb (Fig. 2).

GATCTTAAGAAAATTGAAGGTACAGTGAATTGGCCAAA
CCTGAAATGTGAAACAACCTGAAGGGATTCTGGGATTGA
CTAGATATTATAGGAAATTTGTGAGGAGTTATGGATC [116
pb]
ATTACTTGAAATACAATCGCTCGAATTTTCGGGTGCAATG
ATC [43 pb]

Fig. 2 Séquence pCaORS017 avec les sites enzymatiques Sau3A

La comparaison avec la banque EPLN de pCaORS017 dans son intégralité révèle une homologie forte avec des séquences de rétrotransposons. Cette homologie est conservée avec 17-1 mais pas avec 17-2 (Tableau 4).

L'homologie de 17 et de 17-1 à des séquences de rétrotransposons se retrouve chez le maïs (R1e et Ty3/gypsy), l'ananas (Ty3/gypsy), la tomate (gypsy-like) et un transposon de lys. Les scores FASTA

s'échelonnent de 0,0059 à 0,98. Afin de déterminer la partie du transposon ou du rétrotransposon dans laquelle se situe la séquence 17-1, une recherche des phases ouvertes de lecture (orf: open reading frame) est réalisée sur le serveur internet : <http://www.infobiogen.fr/services>.

La traduction proposée des 159 nucléotides de pCaORS017 en 53 acides aminés et des 116 nucléotides de 17-1 en 38 acides aminés dans le sens direct et avec l'orf 1 est reportée sur la figure 3.

La comparaison de la séquence protéique obtenue à partir de la traduction de 17 et de 17-1 avec les séquences protéiques de la banque CDS donne des pourcentages d'homologie importants avec des reverse-transcriptases de rétrotransposons (Tableau 5).

Les homologies tant au niveau nucléiques que protéiques laissent supposer que pCaORS017 correspond à une séquence de rétrotransposon.

GAT CTT AAG AAA ATT GAA GGT ACA GTG AAT TGG CCA AAA CCT GAA AAT GTG AAA CAA CTG AAG
D L K K I E G T V N W P K P E N V K Q L K
GGA TTT CTG GGA TTG ACT AGA TAT TAT AGG AAA TTT GTG AGG AGT TAT GGA TCA TTA CTT GAA
G F L G L T R Y Y R K F V R S Y G S L L E
ATA CAA TCG CTC GAA TTT TCG GGT GCA ATG ATC
I Q S L E E S G A M I

Fig. 3. Séquence de pCaORS017 (17-1 et 17-2) et sa traduction à partir de la première orf. Légende : la séquence de 17-2 est soulignée

Tableau 4. Comparaison des différents segments aux banques de données nucléotidiques

| Nom du fragment | Taille en pb | Meilleur score FASTA | Description de la séquence homologue | Pourcentage d'identité (%) | Nombre de nts identiques |
|-----------------|--------------|----------------------|---|----------------------------|--------------------------|
| 17 | 159 | 0.0024 | Rétrotransposon R1e de <i>Zea mays</i> (4124 pb) | 64 | 125 |
| 17-1 | 116 | 0,0059 | Retrotransposon R1e de <i>Zea mays</i> (4124 pb) | 65 | 104 |
| 17-2 | 43 | 1.2 | ADN chromosomique d' <i>Arabidopsis thaliana</i> (74945 pb) | 73 | 34 |

nts : nucléotides

Tableau 5. Comparaison des différents segments aux banques de données protéiques

| Nom du fragment | Taille en aa* | Meilleur score FASTA | Description de la séquence homologue | Pourcentage d'identité (%) | Nb d'aa identiques |
|-----------------|---------------|----------------------|---|----------------------------|--------------------|
| 17 | 53 | 1.9 e -06 | Polyprotéine, reverse transcriptase (149aa) | 50 | 40 |
| 17-1 | 38 | 5.5 e-07 | Polyprotéine, reverse transcriptase (149aa) | 53 | 38 |

* aa: acides aminés

Sonde pCaORS038.

pCaORS038 a une taille de 444 paires de bases. Elle ne possède aucun site de restriction Sau3A, HaeIII, ou Tru9I (Fig. 4).

```
GATCGCAGACATTTCTGGAACACCTCTAACAGAGGGAC
AAATAGTCAATTTTGAAAGAACTTATTGTCAACCTCTTT
CAGATATGAATCTATCTGATTCAGAAGGGAAGAACTTG
CATCAGTATCGCAATTTCAATTCAAACATGGGTTTGATT
CACACCCCATATTCTGAGAAAATTTACCATCCGAAAA
GAGGAAAAAACGGATTCCTTTGTCTAAAGAAATGCGTGC
AAAAAGGGCAGATGTATAGAACTTTTCAACGAGATAGT
GCTTTTCAACTCTCTCAAAATGGAATCTATTCCAAACA
TATATGCCATGGTTCCTTACTTCGACAGGGTACAAATAT
CTAAATTTGATATTTTAGATACTTTTTCAGACCTATTGC
GATACTAAGTAGCAGTCAAAAATGGTATCCATTTTTC
ACGATATTATGCATGGATC [ 444pb]
```

Fig. 4. Séquence pCaORS038 avec les sites enzymatiques Sau3A

La séquence est hybridée sur les trois membranes. L'ADN de tous les individus et de toutes les espèces s'hybride. Quelle que soit la membrane, les profils obtenus sont simples. Ils comportent une seule bande pour tous les individus. Tous les profils de la membrane Sau 3A sont à une bande, au niveau correspondant à la taille de la sonde soit 444 pb. L'intensité du signal est plus importante chez PSE que chez DEW. Sur la membrane Hae III, les profils d'hybridation des PSE diffèrent légèrement de ceux de DEW. Le fragment hybridé (>2000pb) est plus petit et plus intense chez PSE. Toutes les espèces ont un profil à une bande de taille plus proche de celle de PSE. Seules les espèces *C. humilis* et *C. heterocalyx* ont, en plus, une seconde bande vers 400 pb (Fig. 5). Ce sont deux espèces, phylogénétiquement éloignées, et qui possèdent les quantités d'ADN les plus importantes du genre.

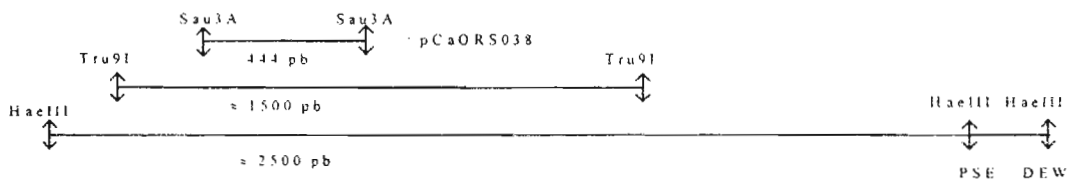


Fig. 6. Cartographie des sites de restriction

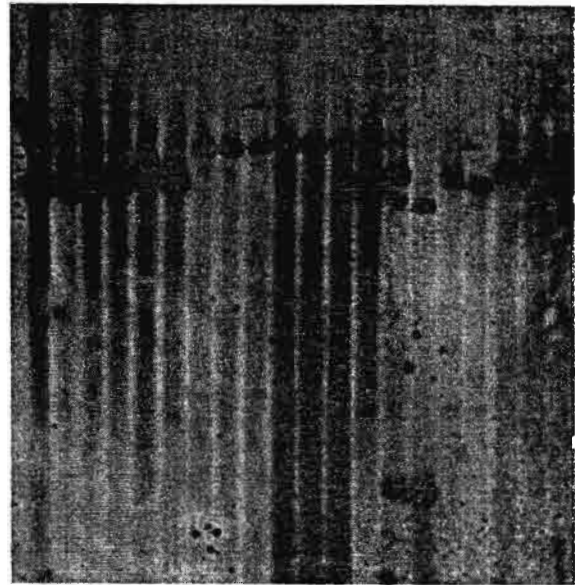


Fig. 5. Hybridation de pCaORS038 sur la membrane HaeIII. De gauche à droite : 6 PSE, 6DEW, *C. sp Mouloundou*, *C. humilis*, *C. heterocalyx*, *C. congensis*, *C. canephora*, *C. eugenioides*, *C. sessiliflora* et *C. racemosa*

Les profils d'hybridation sur la membrane Tru 9I sont constitués d'une seule bande de 1500 pb environ.

Toutes ces informations permettent de cartographier les sites de restriction Tru9I et HaeIII de part et d'autre de la sonde pCaORS038 (Fig. 6). Cependant, pour préciser exactement leur emplacement, il faudra réaliser des doubles ou triple restrictions.

pCaORS038 ne montre aucune homologie forte avec des séquences connues. Le meilleur score FASTA est de 0,14 avec un gène d'*Anthirrinum majus* de 4269 nucléotides. L'identité est de 66% sur 102 nucléotides.

Discussion

Les 30 sondes séquencées n'ont pas d'homologie forte avec les séquences des banques de données, à l'exception de pCaORS017. D'après les hybridations RFLP, il ne semble pas y avoir de variabilité intraspécifique, ni de séquence spécifique à PSE ou à DEW. Trois sondes sont étudiées de façon plus approfondie : i) pCaORS006 (>923pb) est une séquence chimérique. Un de ses fragments (6-3) possède un profil RFLP en échelle révélant une séquence répétée en tandem; ii) pCaORS017 (149pb) est homologue à une reverse-transcriptase de rétrotransposon de type gypsy-like; iii) pCaORS038 (444pb) génère des profils RFLP à une seule bande permettant la cartographie des sites de restriction. Le choix de la banque, la qualité moléculaire des sondes, leur apport dans la structuration du groupe des caféiers et les possibilités de poursuites d'étude font l'objet de notre discussion.

Choix de la stratégie

Plusieurs stratégies de recherche de séquences répétées ont été envisagées. Il était possible d'orienter la recherche sur un type de séquences précis tel que les séquences télomériques ou les transposons. Ce pouvait être l'étude sur des centromères ou des télomères séparés par électrophorèse en champ pulsé après restriction par des enzymes ne coupant pas dans ces zones telles que Bal31 (Ganal *et al.* 1991). Ce pouvait être aussi, en utilisant l'amplification par PCR du génome grâce à des amorces consensus telles que les amorces Alu-like (Baurens *et al.* 1998) ou des amorces comprises dans les régions conservées du gène de la reverse-transcriptase de rétrotransposon (Pearce *et al.* 1997). Une autre possibilité consistait à digérer l'ADN génomique extrait avec une ou plusieurs enzymes (Harrison et Heslop-Harrison 1995; Irifune *et al.* 1995). L'ADN restreint est mis à migrer sur un gel d'agarose. Les bandes nettes qui apparaissent après restriction, sont

récupérées puis clonées. Ces stratégies n'ont pas été adoptées, car elles restreignent trop le champ d'étude à une catégorie précise de séquences répétées. L'absence de données sur les séquences répétées des caféiers demandait que les recherches soient les plus larges et les plus complètes possibles. C'est pourquoi, nous avons préféré une méthode plus lourde, nécessitant un investissement à long terme mais couvrant tout le génome : la création d'une banque génomique d'ADN enrichie en séquences répétées. Cette approche permet théoriquement de cloner différents types de séquences répétées réparties aléatoirement dans tout le génome. Une banque d'ADN génomique de l'hybride F1 (PSE×DEW) a donc été constituée pour mettre en évidence des séquences répétées.

Qualité moléculaire des sondes obtenues

Sur les 193 sondes retenues parce que supposées répétées après rétrohybridation, 30 sont séquencées. Leur taille varie de 47pb à plus de 925pb. Les séquences les plus petites ne sont pas exploitables comme pCaORS002 qui n'a que 47pb et un site interne Sau3A. De plus, aucune redondance n'est relevée.

La présence de sites Sau3A internes dans de nombreuses séquences peut avoir trois origines distinctes. Il pourrait s'agir de sites Sau3A non restreints à la suite d'une mauvaise restriction. Ce pourrait également être des sites Sau3A reconstitués après religation de plusieurs fragments dans le plasmide, constituant des séquences chimériques. Cependant, le taux élevé de sondes chimériques semble surprenant. Une autre explication est que tous les sites GATC ne sont pas méthylés sur les mêmes bases. L'enzyme Sau3A ne reconnaît que les sites GATC méthylés sur la guanine. Avec une méthylation différente, ce sera l'un des isoschizomères comme MboI et XhoII qui sera capable de restreindre l'ADN, et non Sau3A. Statistiquement, l'absence de sites MboI ou XhoII ou des autres isoschizomères sur la totalité des sondes séquencées est peu probable, puisque, théoriquement, leur fré-

quence de coupure est identique à Sau3A. Une nouvelle restriction des sondes par Sau3A ne permet pas de déterminer quelle est la bonne hypothèse. En effet, même s'il est considéré que la restriction du départ est totale, le taux de méthylation de l'ADN est modifié au cours des étapes de construction de la banque et il est impossible de distinguer les sondes chimériques des autres. S'il est supposé que la méthylation de l'ADN est conservée par la lyophilisation et les techniques d'extraction, elle est modifiée lorsque l'ADN, inséré dans le plasmide, se trouve dans la bactérie (Russo *et al.* 1996).

Le parti de considérer les séquences avec des sites GATC internes comme des séquences chimériques a alors été pris. Ceci implique de réaliser successivement une étude sur la séquence entière, puis sur les différents fragments. La comparaison des séquences entières et fragmentées avec les banques de données permet de favoriser une hypothèse par rapport à une autre. En ce qui concerne pCaORS017, l'hypothèse d'une sonde non chimérique semble être favorisée. La séquence est de petite taille. La comparaison des séquences 17 et du fragment 17-1 révèle une meilleure homologie pour la séquence en entier que pour le fragment, pour la séquence nucléotidiques, et une homologie similaire pour les séquences d'acides aminés. pCaORS006 semble plutôt correspondre à une sonde chimérique. Sa taille est importante et il n'existe aucun rapport significatif entre l'homologie obtenue avec la sonde en entier ou avec un des fragments. L'homologie très faible de pCaORS006 en entier avec une séquence de *Limnanthes douglasii* est encore bien plus basse que celle obtenue avec 6-1 pour la même séquence.

L'étude des sondes en entier puis des fragments les composant pose le problème de la séparation fine des bandes de taille similaire. Le fragment de plus haut poids moléculaire est souvent récupéré. Les fragments de petites tailles vont se confondre entre eux (fragments 6-1 et 6-4) ou se confondre avec les

amplifications du polylinker de pUC19 (6-2). Le mélange de fragments de sonde et de polylinker ne pose pas de problème au niveau des hybridations, puisqu'il a été démontré que pUC19 ne s'hybridait pas avec l'ADN des caféiers. Cependant, une séparation en gel d'acrylamide augmenterait sans doute la résolution des petits fragments.

Les sondes testées ne sont pas homologues à de l'ADN ribosomique. D'une part, lors de l'hybridation des sondes avec l'ADN de F1, il n'y a pas eu de signal extrêmement intense, synonyme d'une hybridation avec de l'ADNr. D'autre part, l'enzyme Sau3A ne coupe sans doute pas à l'intérieur de l'ADNr. Malgré le très grand nombre de répétitions de cette séquence, elle n'est pas clonée, car la taille du fragment formé, après restriction, est alors, trop importante. Toutefois pour s'assurer de l'absence de séquences ribosomiques dans la banque, il faudrait hybrider toutes les sondes avec de l'ADN ribosomique.

La banque d'ADN génomique construite avec l'ADN de F1 est donc un outil de bonne qualité qui pourra largement être utilisée par la suite.

Homologie des sondes

Les sondes séquencées montrent, en général, très peu d'homologie pour des séquences connues. Bien que les banques contiennent encore peu de séquences répétées de végétaux, il est possible que les séquences étudiées soient relativement spécifiques des caféiers. Comme attendu, elles semblent, pour la plupart non codantes, à l'exception de pCaORS017.

La séquence pCaORS017 possède une homologie de séquences aussi bien au niveau des nucléotides que des acides aminés avec des rétrotransposons. Cette homologie en acides nucléiques n'est pas très forte, mais se retrouve sur plusieurs rétrotransposons de diverses plantes (maïs, ananas, tomate) confortant l'hypothèse sur sa nature. De plus, la séquence nucléotidique traduite présente des homologies fortes avec des reverse-trans-

criptases. Ainsi, pCaORS017 appartiendrait à une séquence codant pour une reverse-trans-criptase d'un rétrotransposon de type "gypsy-like". La famille Ty3/gypsy est, avec la famille Ty1/copia l'une des familles de rétrotransposons à LTR (long terminal repeat) la plus étudiée. La famille Ty3/gypsy est ubiquiste, elle est présente chez les animaux, les végétaux et les champignons. Les rétrotransposons de ce type s'insèrent préférentiellement à proximité ou à l'intérieur de séquences répétées. Ils possèdent plus d'homologie avec les rétrovirus que les rétrotransposons de type Ty1/copia (Grand-bastien 1992; Bennetzen 1996).

Implication des séquences répétées dans la structure et l'expression du génome.

Si, pendant longtemps, il a été considéré que l'ADN répété n'avait pas de fonction dans le génome, il s'avère que certaines séquences répétées participent à l'explication de la taille, de la structuration et de l'évolution des génomes. Les séquences de type transposon / rétrotransposon, en s'insérant au hasard dans un gène peuvent conduire à des mutations génétiques importantes. C'est l'exemple, très célèbre et longtemps incompris, de la couleur des grains de maïs modifiée selon les insertions de transposons (Mc Clintock 1951). Le rôle des séquences répétées dans l'expression du génome peut également dépendre de l'agencement des gènes. Si, au cours d'une transposition, un gène se trouve dans une zone hétérochromatique, riche en séquences répétées, son expression va être modifiée, comme cela a été montré chez la drosophile (Eissenberg *et al.* 1995). Les séquences répétées de type ADN satellite, localisées près des centromères, jouent probablement un rôle dans l'appariement chromosomique lors des divisions cellulaires successives. Des différences de constitution très importantes quantitatives et / ou qualitatives, au sein des centromères, pourraient entraîner les difficultés d'appariements chromosomiques observées

lors des hybridations interspécifiques (Roizès *et al.* 1994). Les séquences répétées protègent aussi l'ADN contre le vieillissement cellulaire, c'est le cas spécifique des séquences télomériques, mais aussi des autres séquences répétées, sièges plus importants de mutations. De ce fait, les séquences répétées évoluent plus rapidement que les séquences codantes et sont largement utilisées comme marqueur moléculaire, par exemple pour des études phylogénétiques (Karp et Edwards. 1998) ou pour déterminer l'origine parentale des chromosomes d'un hybride (Barre 1997; Calderini *et al.* 1997). Les nombreuses implications des séquences répétées au sein des génomes au niveau de la taille, de la structure et de l'appariement des chromosomes, montrent l'importance d'étudier ce type de séquences chez les caféiers.

Le travail mené ici, a permis l'obtention d'un outil d'étude de bonne qualité : une banque génomique riche en séquences répétées. Il a permis également, d'étudier trois séquences répétées dont une homologue à un rétrotransposon, premier découvert chez les caféiers. Même si aucune des séquences étudiées ne peut-être considérée comme entièrement responsable de la différence de quantité d'ADN entre PSE et DEW, il est indispensable à ce stade d'approfondir l'étude, d'une part de pCaORS017 et des autres sondes séquencées, et d'autre part, de poursuivre l'exploitation de la banque.

En ce qui concerne la séquence homologue au rétrotransposon, il faudrait augmenter la taille de la séquence connue. Ce peut être par synthèse d'amorces externes permettant d'amplifier les génomes parentaux et F1, s'il existe des séquences identifiées proche. Une autre possibilité consisterait à recréer une banque d'ADN génomique de F1 avec une enzyme à coupure moins fréquente (enzyme à 6 paires de bases). En hybridant les nouvelles sondes avec pCaORS017, la ou les séquences repérées complèteront pCaORS017. Il sera également intéressant d'évaluer le nombre de répétitions de cette séquence, sa répartition sur le génome par hybridation *in situ* pour en déterminer la structure fine et l'organisation,

son éventuelle traduction en protéine, de même que sa capacité à transposer. Les travaux d'hybridation RFLP de pCaORS0017 et d'hybridation *in situ* ont d'ores et déjà été commencés.

Pour les autres séquences, il est primordial d'évaluer le nombre de copies puisqu'aucune ne s'hybride uniquement sur l'une ou l'autre des espèces. Le nombre de répétitions varie certainement d'une espèce à l'autre. Pour s'en assurer, il faut évaluer le nombre de copies par séquence et par espèce. Pour cela, plusieurs techniques sont envisageables : par Dot-Blot (Pearce *et al.* 1997), par PCR compétitive (Baurens *et al.* 1996) ou par la réalisation d'une double échelle de dilution à la fois de l'ADN génomique et de la sonde, puis en comparant l'intensité du signal d'hybridation avec comme référence de base la sonde (Calderini *et al.* 1997). Il sera également intéressant de localiser les séquences les plus intéressantes par hybridation *in situ*.

Il est important également de poursuivre l'exploitation de la banque d'ADN. La même stratégie peut être suivie : séquençage systématique, comparaison avec les banques de données, hybridation RFLP, évaluation du nombre de copies dans le génome et localisation sur les chromosomes par hybridation *in situ*.

Les séquences répétées ainsi identifiées peuvent contribuer aussi à des études phylogénétiques. Les résultats RFLP déjà obtenus corroborent la classification en trois groupes des caféiers diploïdes africains établies à partir de l'ADN ribosomique (Lashermes *et al.* 1997). Ainsi, les profils de *C.racemosa*, *C.sessiliflora* et PSE sont similaires. Il existe des similitudes entre les profils de *C.canephora*, *C.congensis*, *C.heterocalyx*, *C.humilis* et DEW. Les profils de *C.eugenioides* et *C.sp* "Moloundou", sont très souvent semblables et se distinguent très nettement de ceux des autres espèces.

Si dans un premier temps, l'étude des séquences répétées se porte sur les parents du

croisement, il sera intéressant dans un second temps, d'analyser la transmission des séquences répétées dans la descendance F1 et de rétrocroisements. A plus long terme, l'impact des séquences répétées dans l'appariement chromosomique, et dans les problèmes de fertilité lors des croisements interspécifiques pourront être étudié par cytogénétiques sur les cellules mères de pollen.

Références bibliographiques

- 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). Thèse Doct., ENSAM 117p.
- Barre P, Layssac M, D'Hont, Louarn J, Charrier A, Hamon S and Noirot M (1998) Relationship between parental chromosomal contribution and nuclear DNA content in the coffee interspecific hybrid : *C.pseudozanguebariae* × *C.liberica* var. *dewevrei*. *Theor Appl Genet* 96 : 301-305.
- Barre P, Louarn J, Duperray C, Hamon S, Charrier A, and Noirot M. Additivity of nuclear 2C DNA content in interspecific coffee hybrids and random distribution of chromosomes in backcrosses with both species. *Theor Appl Genet.* (soumis).
- Barrel M et Jacquet M (1994) La qualité du café, ses causes, son appréciation, son amélioration. Plantation, recherche, développement. juillet-août 1994 : 5-13.
- Baurens FC (1997) Identification par PCR des espèces impliquées dans la composition génomique des cultivars de bananier, à l'aide de séquences répétées. Thèse Doct., Univ. P.Sabatier Toulouse III.
- Baurens FC, Noyer JL, Lanaud C, Lagoda PJJ (1996) Use of competitive PCR to assay copy number of repetitive elements in banana. *Mol Gen Genet* 253: 57-64.
- Baurens FC, Noyer JL, Lanaud C, Lagoda PJJ (1998) Inter-Alu PCR like genomic profiling in banana. *Euphytica* 99 : 137-142.
- Bennetzen JL (1996) The contributions of retroelements to plant genome organization, function and evolution. *Trends in Microbiology* 4 (9) : 347-353.
- Berthaud J and Charrier A (1988) Genetic resources of Coffea. In : *Coffee* 4 : 1-40, Clarke R.J et Macrae R. Elsevier Applied Science, Londres

- Calderini O, Pupilli F, Paolocci F and Arcioni S (1997) A repetitive and species-specific sequence as a tool for detecting the genome contribution in somatic hybrids of the genus *Medicago*. *Theor Appl Genet* 95 : 734-740.
- Charlesworth B, Sniegowski P and Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature (London)* 371 : 215-220.
- Charrier A (1978) La structure génétique des caféiers spontanés de la région malgache et leurs relations avec les caféiers d'origine africaine (*Coffea*). Thèse Doct., Mémoire ORSTOM, Paris n°87, 223p.
- Cros J (1994) Implications phylogénétiques des variations de l'ADN chloroplastique chez les caféiers (genre *Coffea* L. et *Psilanthus* Hook. f.). Thèse, Univ. Montpellier II 160p.
- Davidson EH and Britten RJ (1973) Organization, transcription, and regulation in the animal genome. *Quart Rev Biol* 48 (4) : 565-613.
- Deininger PL (1989) SINE's, short interspersed repeated DNA elements in higher eucaryotes. In *Mobile DNA*, edited by M. Howe and D. Berg, ASM press Washington, D.C.
- Eissenberg JC, Sarah CR, Elgin R and Paro R (1995) Epigenetic regulation in *Drosophyla* : a conspiracy of silence. In *Chromatin structure and gene expression*, edited by Sarah CR Elgin R. Oxford University Press.
- Fajkus J, Kovarik A, Kralovics R and Bezdek M (1995) Organization of telomeric and sub-telomeric chromatin in the higher plant *Nicotiana tabacum*. *Mol Gen Genet* 247 : 633-638.
- Flavell RB, Rimpau J and Smith DB (1977) Repeated sequence DNA relationships in four cereal genomes. *Chromosoma* 63 : 205-222.
- Ganal MW, Lapitan NLV and Tanksley SD (1991) Macrostructure of the Tomato telomeres. *The Plant Cell* 3 : 87-94.
- Gierl A and Saedler H (1992) Plant-transposable elements and gene tagging. *Plant Mol Biol* 19 : 39-49.
- Grandbastien MA (1992) Retroelements in higher plants. *Trends in Genet* 8 (3) : 103-108.
- Harley CB, Futcher AB and Greider CW (1990) Telomeres shorten during ageing of human fibroblasts. *Nature* 345 : 458-460.
- Hastie ND, Dempster M, Dunlop MG, Thompson AM, Green DK and Allshire RC (1990) Telomere reduction in human colorectal carcinoma and ageing. *Nature* 346 : 866-868.
- Harrison GE and Heslop-Harrison JS (1995) Centromeric repetitive DNA sequences in the genus *Brassica*. *Theor Appl Genet* 90 : 157-165.
- Irifune K, Hirai K, Zheng J, Tanaka R and Morikawa H (1995) Nucleotide sequence of a highly repeated DNA sequence and its chromosomal localization in *Allium fistulosum*. *Theor Appl Genet* 90: 312-316.
- Jeffreys AJ, MacLeod A, Tamaki K, Neil DL and Monckton DG (1991) Minisatellite repeat coding as a digital approach to DNA typing. *Nature* 354 : 204-209.
- Karp A and Edwards KJ (1995) Molecular techniques in the analysis of the extent distribution of genetic diversity. In *Analysis, characterization and conservation of plant genetic resources*. Report of an IPGRI workshop 9-11 oct. 1995. Rome (Italy).
- Lashermes P, Combes MC, Trouslot P and Charrier A (1997) Phylogenetics relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. *Theor Appl Genet* 94 : 947-955.
- Lapitan NLV (1992) Organization and evolution of higher plant nuclear genomes. *Genome* 35 : 171-181.
- 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), 200p.
- Mac Clintock B (1951) Cold Spring Harbor Symp Quant Biol 16 : 13-47.
- Pearce SR, Harrison G, Heslop-Harrison PJS, Flavell AJ and Kumar A (1997) Characterization and genomic organization of Tyl-copia group retrotransposon in rye (*Secale cereale*). *Genome* 40: 617-625.
- Pearson WR and Lipman DJ (1988) *Proc Natl Acad Sci*. 85: 2444-2448.
- Roizès G, Marçais B et Yurov Y (1994) Les centromères des chromosomes de mammifères. *Médecine/sciences* 10 : 282-295.
- Russo VEA, Martienssen RA and Riggs AD (1996) Epigenetic mechanisms of gene regulation. Cold spring harbor eds.
- Schmid CW and Deininger PL (1975) Sequence organization of human genome. *Cell* 6 (3) : 345-358.
- Singer MF (1982) Highly repeated sequences in mammalian genomes. *Int Rev Cytol* 76 : 67-112.

Thompson WF, and Murray MG (1981) The nuclear genome : structure and function. In Biochemistry of plants. Academic Press, New York 1-77.

Thompson H, Schmidt R, Brandes A, Heslop-Harrison P, and Dean C (1995) Repeated sequences within the Arabidopsis genome. Plant genome III.

Vogt P (1990) Potential genetic functions of tandem repeated DNA sequence blocks in the human genome are based on highly conserve "chromatin folding code". Hum Genet 84 : 301-336.

Zakian VA (1989) Structure and function of telomeres. Annu Rev Genet 23:579-604.

ANNEXE 1

Caractéristiques physiologiques, génétiques et phénologiques de *C. pseudozanguebariae* (PSE) et *C. liberica* var. *dewevrei* (DEW), d'après Louarn (1992) et Barre (1997).

| Espèces | Origine géographique | Quantité d'ADN (pg) | Feuilles | | Fleurs | | | Fruits | | | Phénologie | | Caféine (% MS) |
|---------|--------------------------------|---------------------|--------------|---------------|---------------|-----------|-------------------|--------|--------|--------------|------------|-----------------------|----------------|
| | | | largeur (cm) | longueur (cm) | inflorescence | pédoncule | nombre de pétales | taille | disque | couleur | floraison* | fructification (mois) | |
| PSE | Afrique de l'Est | 1,13 | 3,6±0,5 | 9,0±1,2 | uni | long | 5-6-7-8 | petit | petit | violet foncé | 7-8 | 2,5-3,5 | <0,02 |
| DEW | Afrique Centrale et de l'Ouest | 1,43 | 10,1±1,2 | 23,6±1,9 | pluri | - | 5-6 | moyen | petit | rouge | 6 | 10-11 | 0,5-1,8 |

* : nombre de jours après la pluie déclencheuse

MS : matière sèche des graines