

MINISTÈRE DE L'AGRICULTURE ET DE LA PÊCHE

CENTRE INTERNATIONAL D'ÉTUDES SUPÉRIEURES EN SCIENCES

AGRONOMIQUES

- MONTPELLIER SUP AGRO -

THÈSE

Présentée pour l'obtention du titre de

**DOCTEUR EN SCIENCES**

École doctorale: Biologie des Systèmes Intégrés, Agronomie et Environnement  
Spécialité: Ressources Phytogénétiques et Interactions Biologiques

Par

**MAHÉ Laetitia**

Titre de la thèse

**Contribution à l'amélioration génétique de la résistance des  
caféiers (*Coffea arabica* L.) à la rouille (*Hemileia vastatrix*)**

**De l'étude des hybrides interspécifiques naturels de Nouvelle-Calédonie  
à la cartographie d'un locus de résistance**

Soutenue le mardi 9 janvier 2007 devant le jury composé de

Rapporteur	Véronique LEBFÈVRE, Directeur de Recherche, INRA Avignon
Rapporteur	Daniel PRAT, Professeur, Université Claude Bernard Lyon 1
Examineur	Claire NEEMA, Professeur, INA/PG laboratoire de pathologie végétale
Examineur	André CHARRIER, Professeur, SupAgro de Montpellier
Examineur	Michel DELSENY, Directeur de Recherche, CNRS Perpignan
Directeur de thèse	Philippe LASHERMES, Directeur de Recherche, IRD Montpellier

## Remerciements

- *A Philippe Lashermes, pour son soutien durant ces trois années de recherche, son encadrement, nos discussions scientifiques, pour m'avoir permis de me rendre à des conférences et présenter mon travail de recherche. Je lui suis infiniment reconnaissante pour avoir partagé ses connaissances, son expérience, et pour sa disponibilité.*
- *A Daniel Le Pierrès, pour m'avoir appris la biologie des caféiers et à reconnaître les hybrides interspécifiques de Nouvelle-Calédonie.*
- *A Marie-Christine Combes, l'arme secrète de cette équipe qui porte sur ses épaules la matière de tous projets, qui m'a apportée aide et amitié. Marie-Christine ne fait pas partie de l'organigramme des officiels de la thèse mais sa présence, son investissement ont fait beaucoup pour mon travail.*
- *A l'ensemble des personnes de l'équipe de Véronique Lefebvre de l'INRA d'Avignon et en particulier à Patrick Signoret pour son aide précieuse.*
- *Aux membres de l'équipe de Nouvelle-Calédonie pour leur investissement et leur financement. J'y associe Vitor Varzêa du CIFC du Portugal, grâce à lui l'évaluation des plantes à la rouille a été possible.*
- *Aux membres de l'équipe « résistance des plantes », notamment, Michel Nicole, Diana Fernandez, Benoît Bertrand, Philippe Marmey, Anne-Sophie Petitot, Anne-Claire Lecouls, Agnès Galzi, Sylvie Pounhet et Claire Guilhaumon pour les moments de détente ou de travail que nous avons partagés.*
- *Et, quelque soit l'humeur du jour pendant ces trois années, étaient présents: Myriam R. (pour l'aide à la bibliographie), Juan-Carlos, Elijha, Sorho, Romain, Arnaud, Myriam C (pour les chocolats), Christine, Laurence, Valérie, Léandro, Edgardo.*
- *Aux membres de la petite entreprise: l'association CONTACT avec lesquels nombre de projets se sont montés.*
- *A ma famille pour son soutien.*
- *Enfin je dédie ce travail à celui qui m'a permis de réaliser cette thèse et qui m'a accompagné et soutenu chaque jour: Hervé Masson.*

Avant-Propos	6
--------------	---

CHAPITRE I Revue bibliographique	9
----------------------------------	---

<b>1 LE CAFEIER</b>	<b>10</b>
1.1 Les caféiers : une grande famille	10
1.1.1 Phylogénie des caféiers du genre <i>Coffea</i>	10
1.1.2 Caractéristique et origine des caféiers	10
1.2 <i>C. arabica</i> , génétique et amélioration	12
1.2.1 Les ressources génétiques chez <i>C. arabica</i>	13
1.2.2 Les ressources génétiques chez les espèces diploïdes	13
1.2.3 Les hybrides interspécifiques	14
1.3 Les objectifs de l'amélioration	14
<b>2 LA RESISTANCE A LA ROUILLE ORANGÉE</b>	<b>15</b>
2.1 Biologie et physiologie des rouilles orangée du caféier	16
2.1.1 Le genre <i>Hemileia vastatrix</i>	16
2.1.2 Caractérisation de la résistance du caféier à la rouille orangée	17
2.2 La lutte contre la rouille	18
2.2.1 La lutte chimique	18
2.2.2 Les sources de résistance naturelles et artificielles	18
2.2.3 Les recherches génétiques sur la résistance à la rouille	20
<b>3 POLYPLOIDIE ET INTROGRESSION</b>	<b>21</b>
3.1 Les polyploïdes	21
3.1.1 Généralités sur les polyploïdes	21
3.1.2 Formation des polyploïdes	21
3.2 L'hybridation introgressive comme source de diversité	23
3.2.1 L'introgression entre espèces	24
3.2.2 L'introgression entre <i>C. arabica</i> et les espèces diploïdes	24
<b>4 COMPARAISON DES GENOMES</b>	<b>27</b>
4.1 La comparaison des génomes	27
4.2 L'histoire des études comparatives	28
4.3 Le synténie : implications et utilisations	29
4.3.1 Evolution de la structure des génomes	29
4.3.2 Connaissance sur l'évolution des génomes	30
4.3.3 Amélioration de la cartographie génétique et physique	30
4.3.4 Clonage d'un gène par utilisation de l'orthologie	31
4.4 De nouvelles plantes modèles pour la synténie	32
4.4.1 Synténie entre la tomate et <i>Arabidopsis</i>	32
4.4.2 Utilisation de la cartographie comparée pour le clonage de gène de résistance	33
4.4.3 Les défis de la génomique comparative	36
<b>5 LES OBJECTIFS DU TRAVAIL DE RECHERCHE</b>	<b>37</b>

CHAPITRE II	39
-------------	----

<b>6 Introgressive hybridization between the allotetraploid <i>Coffea arabica</i> and one of its diploid ancestors <i>C. canephora</i> in an exceptional sympatric zone in New Caledonia</b>	<b>41</b>
6.1 Abstract	41
6.2 Introduction	41
6.3 Materials and Methods	43
6.3.1 Sampling	43
6.3.2 Determination of nuclear DNA content	43
6.3.3 AFLP assay	43
6.3.4 Microsatellite assay	44
6.3.5 DNA chloroplast identification	44
6.3.6 Data analysis	44
6.4 Results	45
6.4.1 Molecular authentication of interspecific hybrid plants collected in New Caledonia	45
6.4.2 Gene flow between <i>C. arabica</i> and <i>C. canephora</i>	47

6.4.3	Genetic variation of hybrid plants	48
6.5	Discussion	48
<b>7</b>	<b>A new source of resistance against coffee leaf rust from New-Caledonian natural interspecific hybrids between <i>Coffea arabica</i> and <i>C. canephora</i></b>	<b>54</b>
7.1	Abstract	54
7.2	Introduction	54
7.3	Materials and methods	55
7.3.1	Plant material	55
7.3.2	Molecular marker analysis	56
7.3.3	Rust inoculation test	56
7.4	Results and discussion	57
<b>CHAPITRE III</b>		<b>59</b>
<b>8</b>	<b>Development of sequence characterized DNA markers linked to leaf rust (<i>Hemileia vastatrix</i>) resistance in coffee (<i>Coffea arabica</i> L.)</b>	<b>61</b>
8.1	Abstract	61
8.2	Introduction	61
8.3	Materials and methods	63
8.3.1	Plant material	63
8.3.2	Evaluation of rust resistance	64
8.3.3	Development of SCARs from AFLP markers	64
8.3.4	BAC library analysis	65
8.3.5	Analysis of SCAR and SSR markers	65
8.4	Results	66
8.4.1	Conversion of AFLP markers into SCARs	66
8.4.2	Development of BAC-derived markers	67
8.4.3	Identification of SSR markers	67
8.4.4	Genetic Mapping of the $S_H3$ region	67
8.4.5	Confirmation of the leaf rust-linked markers in an advanced segregating population	68
8.5	Discussion	68
<b>CHAPITRE IV</b>		<b>71</b>
<b>9</b>	<b>Comparison between a coffee single copy chromosomal region and <i>Arabidopsis</i> duplicated counterparts evidenced high level synteny between the coffee genome and the ancestral <i>Arabidopsis</i> genome</b>	<b>73</b>
9.1	Abstract	73
9.2	Introduction	73
9.3	Materials and methods	76
9.3.1	BAC resources	76
9.3.2	<i>Arabidopsis</i> homologous sequences	76
9.3.3	<i>Arabidopsis</i> predicted genes present on syntenic <i>Arabidopsis</i> fragments.	77
9.3.4	Coffee orthologs	77
9.3.5	Hybridization experiment	77
9.3.6	BAC fingerprinting and contig assembly	77
9.3.7	Genetic mapping	78
9.4	Results	78
9.4.1	Anchoring to the <i>Arabidopsis</i> genome	78
9.4.2	Analysis of targeted <i>Arabidopsis</i> regions	79
9.4.3	Development of COS markers	79
9.4.4	Contig assembly	80
9.4.5	Validation of contig assembly	80
9.4.6	Synteny between coffee and syntenic segments of <i>Arabidopsis</i>	81
9.5	Discussion	81
<b>10</b>	<b>Cartographie génétique de la région chromosomique <math>S_H3</math> sur <i>Solanum lycopersicum</i></b>	<b>85</b>
10.1	Une population de lignées introgressées pour la cartographie génétique chez la tomate	85
10.2	Détermination des orthologues tomates	86
10.3	Hybridation sur des lignées introgressées quasi-isogéniques	86

10.4	Transfert des marqueurs tomate chez le caféier	87
CHAPITRE V: Discussion générale		88
<b>11</b>	<b>SYNTHESE DES PRINCIPAUX RESULTATS</b>	<b>89</b>
11.1	Etude de la genèse des hybrides et des flux de gène entre <i>Coffea arabica</i> et <i>Coffea canephora</i> à l'intérieur de zones sympatriques en Nouvelle Calédonie	89
11.1.1	Diversité et formation des hybrides	90
11.1.2	Les flux de gènes	92
11.1.3	Nouvelle diversité d'hybrides interspécifiques pour l'amélioration de <i>Coffea arabica</i> à la rouille.	92
11.2	Etude du facteur de résistance $S_{H3}$ à <i>Hemileia vastatrix</i> agent de la rouille.	94
11.2.1	Développement de marqueurs liés au locus $S_{H3}$	95
11.2.2	Cartographie génétique et physique d'un locus de résistance à <i>Hemileia vastatrix</i>	96
11.3	Développement de deux stratégies de génétique comparative	97
11.3.1	Détermination de la microsyténie entre <i>Arabidopsis</i> et le caféier.	97
11.3.2	Localisation de la région chromosomique homéologue $S_{H3}$ chez <i>Solanum lycopersicum</i> .	99
<b>12</b>	<b>PERSPECTIVES</b>	<b>101</b>
12.1	Etude des facteurs affectant l'évolution de l'allotétraploïde <i>C. arabica</i> dans les régions sympatriques de Nouvelle Calédonie.	101
12.2	Vers une connaissance plus approfondie du locus $S_{H3}$	102
12.2.1	Affinement de la cartographie physique	102
12.2.2	Clonage du ou des gènes de résistance au locus homéologue à $S_{H3}$ chez <i>C. arabica</i>	103
12.2.3	Etude des locus de résistance à la rouille au sein des caféiers.	104
12.3	La génomique comparative avec le caféier	104
12.3.1	Génomique comparative de la région chromosomique $S_{H3}$ avec d'autres espèces	104
12.3.2	Evolution des génomes	105
Références bibliographiques		106
Annexes		130
Liste des abréviations		131
Liste des figures		132
Liste des tables		134

## **Avant-Propos**

Le café est un produit de consommation international, se rangeant parmi les cinq exportations agricoles les plus importantes des pays en voie de développement (Food and Agriculture Organization, <http://apps.fao.org>). La production et le traitement du café emploient plus de 25 millions de personnes dans le monde entier (O'Brien et Kinnaird 2003), ce qui en fait un enjeu économique majeur. La production commerciale de café repose principalement sur deux espèces étroitement liées: *Coffea arabica* et *Coffea canephora*, qui comptent respectivement pour 70 et 30% de la production de café dans le monde (International Coffee Organization, <http://www.ico.org>). *Coffea arabica* ( $2n=4x=44$ ) est l'unique espèce allotétraploïde du genre contenant deux sous génomes provenant d'espèces diploïdes ( $2n=2x=22$ ), *C. eugenioides* et *C. canephora*.

La pression phytosanitaire importante au niveau des zones d'exploitation du café constitue une contrainte sérieuse en caféiculture. Cette pression s'exerce tout particulièrement sur *C. arabica* dont les variétés les plus cultivées se sont avérées sensibles à la plupart des maladies et parasites du caféier, tel que la rouille orangée, les nématodes ou encore le scolyte des baies. L'amélioration génétique de *C. arabica* est basée principalement sur l'exploitation des ressources génétiques des caféiers diploïdes. Elle consiste à introduire au sein des variétés cultivées, des gènes d'intérêt à partir des espèces diploïdes. Les hybrides interspécifiques, artificiels ou naturels, sont des relais pour le transfert de caractères agronomiques par rétro-croisements sur *C. arabica*.

Le développement de lignées résistantes, notamment à la rouille, a fait l'objet de programmes de sélection généalogique qui ont duré plus de quarante années. Or, en comparaison de variétés traditionnelles, la majorité des lignées sélectionnées présentent une proportion importante de défauts des grains et une baisse de la qualité à la tasse. L'analyse des résultats de cette sélection montre qu'il est difficile de parvenir à conserver le ou les caractères introgressés tout en réduisant les effets défavorables de l'introgession. Le transfert de nouvelles résistances dans les variétés cultivées rend nécessaire le contrôle des fragments chromosomiques hérités du parent sauvage pour réduire l'effet négatif de l'introgession sur la qualité du café. Aux contraintes de sélection s'ajoute le contournement par les agents pathogènes de la résistance des plantes. L'amélioration de *C. arabica* requiert aujourd'hui la recherche de nouvelle source de résistance en particulier pour la résistance à la rouille et le développement de marqueurs pour la mise en place de programme de sélection assistée par marqueur.

Dans ce contexte, la cartographie des gènes de résistance à la rouille au sein du genre *Coffea*, et l'identification au sein des espèces sauvages de caféiers de nouvelles sources de résistance, en particulier vis-à-vis de la rouille, constituent les objectifs majeurs des programmes d'amélioration de la principale espèce d'intérêt agronomique, *C. arabica*.



## **CHAPITRE I Revue bibliographique**

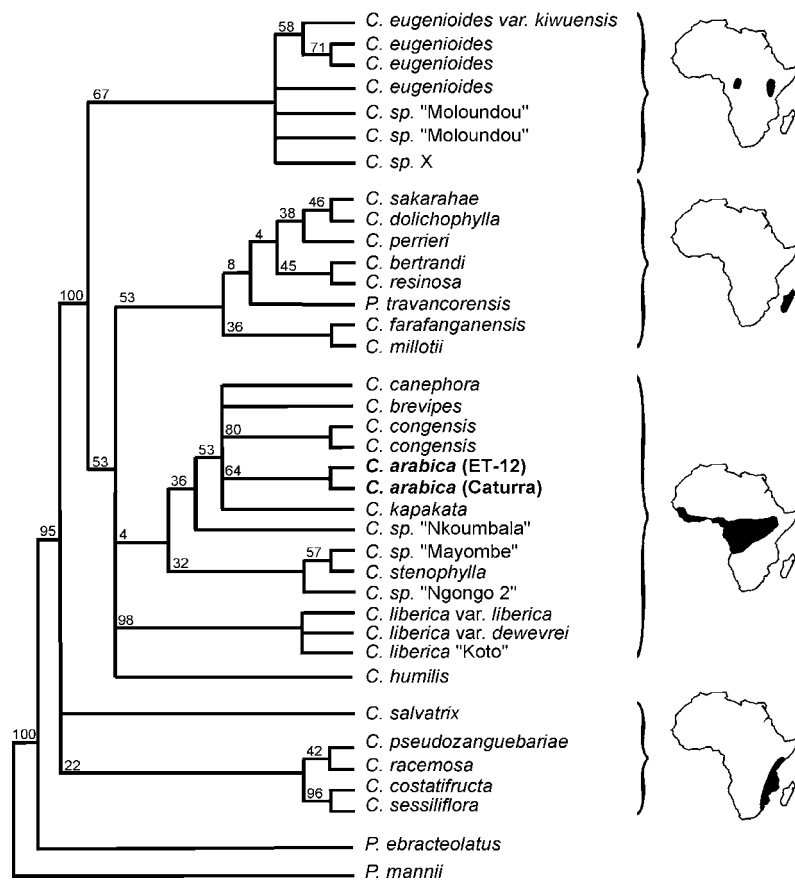


Figure 1. Classification phylogénétique de 32 taxons de caféiers, sur la base du polymorphisme de l'ADN nucléaire ribosomal et distribution géographique des principaux groupes de la classification (Lashermes et al. 1997).

## 1 LE CAFEIER

### 1.1 Les caféiers : une grande famille

Les caféiers appartiennent à la famille des *Rubiaceae* qui comporte plus de 6000 espèces. Ce sont des plantes pérennes présentant une taille très variable qui peut aller de petits arbustes jusqu'à des arbres d'environ dix mètres. Dans cette famille, les caféiers constituent la tribu *Coffeae* caractérisée par une placentation dite « cofféenne », où la partie ventrale de l'albumen des graines forme une invagination. Aujourd'hui nous considérons deux genres : le genre *Coffea* L. subdivisé en deux sous-genre *Coffea* et *Baracoffea*; et le genre *Psilanthus* spp. lui même divisé en deux sous-genre *Psilanthus* et *Afrocoffea*. Des différences au niveau de la position des fleurs et du mode de développement sont à l'origine de cette séparation (Leroy 1980; Bridson 1987; Bridson et Verdcourt 1988). Le genre *Coffea* sous genre *Coffea*, qui comprend plus de 113 taxons (Davis *et al.* 2006), est essentiellement représenté par les deux principales espèces cultivées : *C. arabica* et *C. canephora*. Dans notre étude, la terminologie caféière s'appliquera au genre *Coffea* sous genre *Coffea*.

#### 1.1.1 Phylogénie des caféiers du genre *Coffea*

Le développement des outils d'analyse moléculaire a permis de structurer les espèces du sous genre *Coffea*. Les variations de l'ADN chloroplastique et nucléaire de taxons représentatifs du genre *Coffea* sous-genre *Coffea* ont été étudiées. Les analyses phylogénétiques des données moléculaires à partir des deux génomes analysés (e.g. nucléaire et chloroplastique) suggèrent une explosion radiative récente et l'existence d'au moins 5 groupes phylogénétiques relativement peu différenciés. Ces groupes d'espèces de caféiers (*Coffea* sous-genre *Coffea*) recouvrent les trois ensembles biogéographiques: zone guinéo-congolaise (2 groupes), Afrique de l'Est (1 groupe), et Madagascar, l'archipel des Comores et les îles Mascareignes (1 groupe) (Figure 1).

#### 1.1.2 Caractéristique et origine des caféiers

Les espèces du sous genre *Coffea* sont toutes diploïdes ( $2n = 2x = 22$ ) à l'exception de l'espèce *C. arabica* qui est tétraploïde ( $2n = 4x = 44$ ). L'analyse des caryotypes des 11

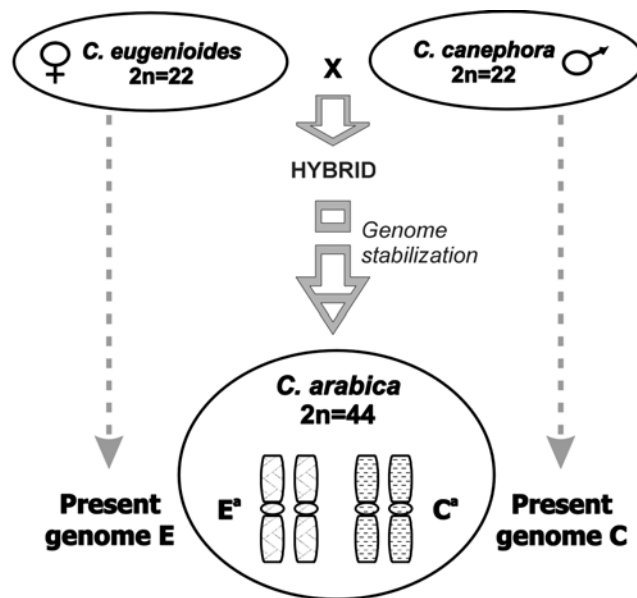


Figure 2. Origine de l'espèce allotetraploïde *Coffea arabica*

chromosomes du génome de caféier, rapportée par Sybenga (1959), met en évidence l'existence d'un génome de base composé de 11 chromosomes ( $n=11$ ) présentant une taille moyenne réduite (1 à 3  $\mu\text{m}$ ) et peu variable. En revanche la quantité d'ADN nucléaire est très variable entre les différentes espèces de caféier, avec des valeurs comprises entre 0,95 et 1,78 pg par noyau. *Coffea arabica* avec le double de chromosomes présente 2,61 pg d'ADN (Cros *et al.* 1995), pour un nombre de base approximatif de 950Mb (Anthony et Lashermes 2005).

Une origine amphidiploïde est suggérée pour l'espèce *C. arabica* (Carvalho 1950; Grassias et Kammacher 1975). Par des approches moléculaires et notamment l'utilisation de marqueurs chloroplastiques et ribosomiques (Berthou *et al.* 1983; Lashermes *et al.* 1995; 1997; Cros *et al.* 1998), il a été possible de confirmer la présence d'une forte affinité entre *C. arabica* et des espèces du groupe des diploïdes originaires d'Afrique centrale et de l'ouest (*C. eugenioides*, *C. canephora*, *C. congensis*, *C. brevipes* et *Coffea* sp. « Mouloundou »). Une étude plus approfondie appuyée sur des analyses RFLP (en utilisant des sondes spécifiques de locus) et cytogénétiques (par hybridation de l'ADN génomique des espèces potentielles avec les chromosomes d'*arabica*) a permis de démontrer la présence de deux lots de chromosomes Ea et Ca, qui proviendraient respectivement de formes ancestrales des espèces actuelles *C. eugenioides* (parent femelle, Génome Ea) et *C. canephora* (parent mâle, génome Ca).

Ainsi l'hybride ancestral tétraploïde pourrait résulter de croisements interspécifiques faisant intervenir des diplo-gamètes ( $2n$ ) dont la formation aurait été favorisée par des conditions climatiques particulières (Figure 2). Il est possible aussi que la polyploïdisation ait eu lieu après la formation d'un hybride interspécifique diploïde suivie d'un doublement chromosomique. Les variations climatiques importantes pendant le quaternaire et leurs implications sur la distribution des différentes espèces de caféiers rendent difficiles l'interprétation de l'origine géographique de *C. arabica*. Toutefois, Lashermes *et al.* (1999) ont suggéré le versant nord-ouest du rift africain comme zone d'origine possible de *C. arabica*, zone géographique où les espèces parentales *C. canephora* et *C. eugenioides* se seraient chevauchées. Par la suite, l'espèce *C. arabica* aurait colonisé la région comprenant le sud du Soudan, le Nord du Kenya et le sud-ouest de l'Ethiopie qui est considéré comme son centre de diversité.

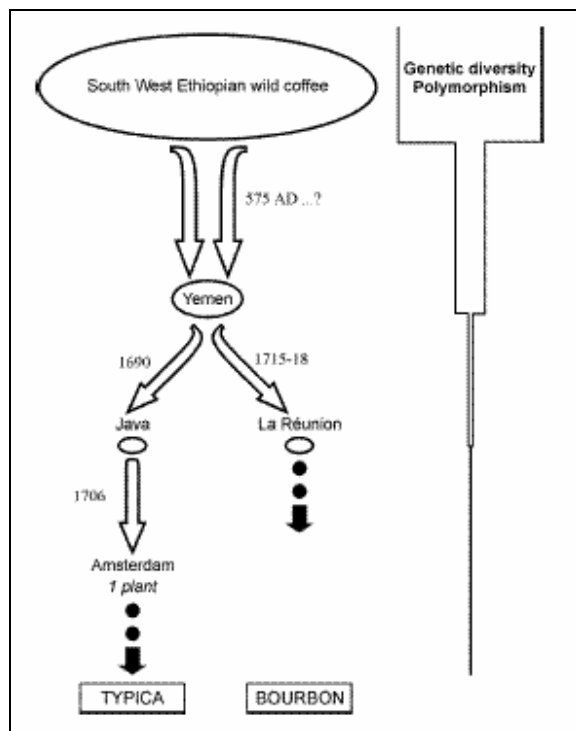


Figure 3. Représentation schématique des principales étapes dans l’histoire de la culture du caféier extrait de Anthony et al. 2002.

## 1.2 *C. arabica*, génétique et amélioration

L'espèce *C. arabica* diffère des espèces diploïdes par son mode de reproduction. Considérée comme espèce autogame, *C. arabica* présente cependant un faible pourcentage de pollinisation croisée (5 à 15%). En revanche les espèces diploïdes sont dans une très grande majorité autostériles à fonctionnement allogame stricte. L'allogamie semble résulter d'un système d'incompatibilité génétique. Chez *C. canephora*, le système d'incompatibilité est de type gamétophytique, contrôlé par un locus S multi-allélique (Berthaud, 1980; Lashermes *et al.* 1996b).

Les principales variétés de *C. arabica* cultivées en Amérique latine (Caturra, Catuai, Mundo Novo) hautement productives donnent un café réputé de bonne qualité. Elles furent sélectionnées à partir de l'étroite base génétique dispersée au début du XVIII<sup>ème</sup> siècle (Figure 3) par les puissances coloniales. Cependant, en conséquence du nombre extrêmement réduit d'individus fondateurs à l'origine des caféiers en plantation à travers le monde, ainsi que du mode de spéciation de *C. arabica*, les variétés cultivées présentent une très faible diversité génétique (AFLP et SSR : Anthony *et al.* 2002a; AFLP: Steiger *et al.* 2002; RAPD: Sera *et al.* 2003; Chaparro *et al.* 2004). Cette caractéristique rend la culture de *C. arabica* particulièrement vulnérable aux aléas.

L'amélioration génétique de l'espèce *C. arabica* est basée principalement sur l'exploitation des ressources génétiques des caféiers diploïdes. Elle consiste à introduire plus de variabilité génétique au sein des variétés cultivées de l'espèce, par introgression d'ADN des espèces diploïdes. Le croisement et l'obtention d'hybrides interspécifiques entre l'espèce tétraploïde *C. arabica* et les espèces diploïdes ne constituent pas un obstacle au transfert de gènes par voie sexuée. Après doublement chromosomique du parent diploïde, des hybrides interspécifiques, et intergénériques, tétraploïdes relativement fertiles peuvent être obtenus (Van der Vossen 1985; Carvalho 1988; Le Pierrès 1995; Charrier et Eskes 1997; Couturon *et al.* 1998). L'analyse génétique de ces hybrides et de leurs descendants a montré une fréquence importante de recombinaisons génétiques entre génomes et un niveau élevé d'introgression (Lashermes *et al.* 2000b; Herrera *et al.* 2002a et b; Prakash *et al.* 2002). Le caféier *C. arabica* est sélectionné essentiellement par la méthode généalogique et par rétro-croisements. Bien que la sélection généalogique nécessite plus de trente années pour l'obtention de lignées chez le caféier (e. g. un cycle d'une génération dure au moins 4 ans), elle s'est avérée

Tableau 1. Exemples de caractères d'intérêt agronomique identifiés au sein d'espèces diploïdes de caféiers ou de formes sauvages de *Coffea arabica*.

<b>Caractères d'intérêt agronomique</b>	<b>Sources de diversité</b>	<b>Références</b>
<b>Sources de résistance aux maladies et ravageurs</b>		
- <i>Perileucoptera coffeela</i>	<i>C. racemosa</i>	Carvalho <i>et al.</i> 1988
- <i>Hemileia vastatrix</i>	<i>C. canephora</i> ,	Berthaud et Lourd, 1982 Kushalappa et Eskes, 1989 Montagnon et Leroy, 1993
	<i>C. liberica</i> , <i>C. pseudozanguebariae</i> ,	Rodrigues Jr. <i>et al.</i> 1975
	<i>C. eugenioides</i>	Rodrigues Jr., 1990
- <i>Colletotrichum kahawae</i>	<i>C. canephora</i>	Van der Vossen, 1997
- <i>Meloidogyne exigua</i>	<i>C. canephora</i> , <i>C. racemosa</i>	Bertrand <i>et al.</i> 1999; Anthony <i>et al.</i> 2002b
- <i>Meloidogyne paranaensis</i>	Formes sauvages d' <i>arabica</i> , <i>C. canephora</i>	Anthony <i>et al.</i> 2002b
- <i>Meloidogyne incognita</i>	Formes sauvages d' <i>arabica</i> , <i>C. canephora</i>	Anzuetto <i>et al.</i> 2001
- <i>Meloidogyne arabicida</i>	Formes sauvages d' <i>arabica</i> , <i>C. canephora</i>	Anthony <i>et al.</i> 2002b
<b>Stérilité mâle</b>	Formes sauvages de <i>C. arabica</i>	Dufour <i>et al.</i> 1997
<b>Durée du cycle de maturation des fruits</b>	<i>C. congensis</i> , <i>C. canephora</i>	Données non publiées
<b>Composition biochimique des graines</b>	<i>C. pseudozanguebariae</i>	Mazzafera et Carvalho, 1992
<b>Adaptation aux sols hydromorphes</b>	<i>C. congensis</i>	Charrier et Vigreux 1972
<b>Tolérance à la sécheresse</b>	<i>C. racemosa</i>	Guerreiro-Filho, 1972
<b>Tolérance aux basses températures</b>	<i>C. congensis</i>	Ahmad et Vishveshwara, 1980
<b>Rusticité</b>	<i>C. canephora</i>	Capot, 1977
<b>Productivité</b>	<i>C. canephora</i> , <i>C. arabica</i> , <i>C. liberica</i>	Capot, 1977; Carvalho <i>et al.</i> 1978.
<b>Granulométrie des grains</b>	<i>C. canephora</i> , <i>C. arabica</i>	Le Pierrès, 1995



efficace pour améliorer la production, la vigueur et la résistance aux maladies de *C. arabica*, tout en conservant les caractéristiques du café arabica (Charrier et Eskes 1997). Les hybrides interspécifiques, artificiels ou naturels, servent de relais pour le transfert de caractères agronomiques par rétro-croisements avec *C. arabica*.

Les caféiers subspontanés du centre primaire de diversité de *C. arabica* constituent une importante ressource génétique (Chaparro *et al.* 2004; Moncada 2004). Toutefois, la diversité génétique présente au sein de ce matériel végétal reste limitée, et l'exploitation des ressources génétiques que représentent les espèces diploïdes de caféiers apparaît incontournable. Ces deux principales ressources représentent un réservoir de diversité génétique essentiel pour l'amélioration de *C. arabica* dans une perspective de développement durable (Tableau 1) (Carvalho, 1988; Anthony *et al.* 1999).

De nombreuses variétés lignées (e. g. Catimor, Sarchimor, ICATU, variété Columbia) ont été créées. Au cours des dernières années, une stratégie de valorisation plus rapide de ces lignées sélectionnées a été de créer des variétés hybrides (Van der Vossen 2001). Les premières variétés hybrides ont été obtenues par croisement en conditions de pollinisation contrôlée (manuelle) entre 2 lignées ou 2 pools de lignées de caféiers (e. g. Ruiru 11). Plus récemment, des variétés hybrides F1, diffusées par multiplication végétative *in vitro* (embryogenèse somatique) d'arbres hybrides, ont été proposées (Bertrand *et al.* 1999; Etienne *et al.* 2002).

### 1.2.1 Les ressources génétiques chez *C. arabica*

Les principales collections actuelles de ressources génétiques des caféiers sont issues des prospections réalisées à partir du début des années 60. En ce qui concerne l'espèce *C. arabica*, les plus importantes prospections ont été réalisées entre 1964 et 1966 en Ethiopie (FAO 1969; Guillaumet et Hallé 1967). Actuellement il existe 10 collections majeures pour les caféiers arabica dans un nombre égal de pays appartenant aux trois continents, Afrique, Amérique, Asie.

### 1.2.2 Les ressources génétiques chez les espèces diploïdes

La collection d'espèces diploïdes fut motivée par la destruction rapide des écosystèmes forestiers africains. Comme résultat des prospections réalisées entre 1960 et 1987,

20000 caféiers sylvestres ont été recensés (Anthony 1992). Néanmoins, la plus grande partie de la diversité génétique se retrouve au sein des collections de caféiers diploïdes dont les plus importantes sont celles de la Côte d'Ivoire et Madagascar (Anthony *et al.* 1999). La diversité génétique des espèces diploïdes représente ainsi un intérêt agronomique majeur (Berthaud et Charrier 1988). Notamment, l'espèce *C. canephora* dont la mise en culture reste récente (e. g. il y a un siècle environ), est particulièrement intéressante. Les cultivars de cette espèce, en dépit d'une qualité à la tasse moindre comparée à celle du café Arabica, se caractérisent par une bonne adaptation à basse altitude, une importante vigueur, une productivité élevée et la résistance vis à vis de divers agents pathogènes du caféier.

### 1.2.3 Les hybrides interspécifiques

Un autre matériel génétique important est constitué par les hybrides spontanés qui sont apparus lorsque des populations de *C. arabica* et de caféiers d'espèces diploïdes comme *C. canephora* ont été cultivées ensembles. L'Hybride de Timor (HdT), notamment, a été identifié au sein d'une plantation de *C. arabica* (plantée en 1927) sur l'île de Timor en Indonésie (Bettencourt 1973). Les analyses génétiques et moléculaires, de cet hybride et de ses descendants, ont démontré que l'HdT est issu d'un croisement interspécifique spontané entre *C. arabica* et *C. canephora* (Bettencourt 1973; Gonçalves et Rodrigues 1976; Lashermes *et al.* 2000b). Les descendants de l'HdT, largement diffusés à travers le monde, ont été très exploités dans les programmes d'amélioration des caféiers ces 50 dernières années. Ils constituent la principale source de résistance aux maladies et ravageurs du caféier, en particulier, vis-à-vis de la rouille orangée, de *Colletotrichum kahawae* et des nématodes à galles du genre *Meloidogyne* (Bertrand *et al.* 2001; Van der Vossen 2001).

## 1.3 Les objectifs de l'amélioration

Les objectifs de sélection propres à la culture de *Coffea arabica* sont directement liés aux problèmes biologiques associés aux systèmes de culture, aux contraintes économiques ainsi qu'aux exigences commerciales.

La culture de *C. arabica* présente différentes caractéristiques :

- Caractéristiques relatives aux zones de culture

D'une part, certaines zones de haute montagne difficiles d'accès ne permettent pas l'application des traitements fongiques. D'autre part, la faible disponibilité en terres agricoles nécessite une intensification de la culture et donc la recherche de gains de productivité. Dans cette optique, les producteurs souhaitent globalement multiplier les variétés naines et vigoureuses (e. g. cv Caturra) afin de cultiver à haute densité et sans ombrage.

- Caractéristiques relatives aux exigences des consommateurs

La demande en café gourmet ou encore café « terroir » se fait croissante. Par ailleurs, les consommateurs soutenus par les producteurs, exercent une forte pression commerciale par l'aspiration à une caféiculture "plus biologique" : un café exempt de résidus de produits phytosanitaires et une réduction des intrants au niveau des plantations. Le consommateur aspire aussi à plus de choix.

Ces contraintes multiples soulignent la nécessité de développer de nouvelles variétés, présentant la ou les qualités suivantes :

- Une résistance aux maladies existantes
- Une productivité au moins égale à celle des meilleures lignées actuelles
- Une qualité organoleptique comparable ou supérieure aux meilleurs *C. arabica*
- Une rusticité assurant une bonne stabilité de production
- Une fertilité parfaite

Notre étude sur l'amélioration de *C. arabica* porte sur sa résistance face au pathogène *Hemileia vastatrix* responsable de la rouille des feuilles.

## **2 LA RESISTANCE A LA ROUILLE ORANGÉE**

La rouille des feuilles ou rouille orangée est l'une des plus graves maladies du caféier causée par le champignon *Hemileia vastatrix*. Elle l'est par son étendue géographique; d'abord confinée en Afrique, elle s'est étendue en Asie, et a atteint le continent latino

américain en 1970, puis s'y est répandue partout en moins d'une quinzaine d'années. Elle concerne tous les pays producteurs de café, exceptés Hawaï et l'Australie. Elle l'est aussi par l'importance des dégâts, souvent illustrée par le cas extrême de Ceylan où la caféiculture a disparu du fait des attaques de cet agent pathogène (Berkeley 1869). Les pertes de production dues à la rouille sur l'arabica sans traitement sont estimées à 30% (Monaco 1977). Elle l'est enfin par les problèmes posés pour la combattre par la voie des traitements chimiques coûteux et difficiles à exécuter dans les zones montagneuses où se cultive *C. arabica* qui est l'espèce la plus vulnérable et la plus cultivée. Pour toutes ces raisons, la rouille orangée du caféier a toujours été une source de préoccupation pour les pays producteurs d'Arabica et tout particulièrement pour les pays latino-américains dont l'économie est très dépendante du café et où le matériel végétal cultivé avait jusqu'à présent une base génétique très étroite (puisque résultant d'une seule introduction de *C. arabica*).

## **2.1 Biologie et physiologie des rouilles orangée du caféier**

### 2.1.1 Le genre *Hemileia vastatrix*

La rouille orangée est une maladie des caféiers causée par un champignon biotrophe basidiomycète appartenant à l'ordre des Urédinales et la famille des Pucciniacées. Le genre *Hemileia* se caractérise par l'absence d'écidiospores et de pycnidiospores, les téléspores et les basidiospores sont peu fréquentes (Gopalkrishnan 1951). Les urédospores sont équinoles dorsalement et lisses ventralement. Les téléspores et les basidiospores n'infectent pas le caféier; Ceci laisse supposer qu'il existe un hôte intermédiaire de la rouille orangée (Kushalappa et Eskes 1988; 1989). Toutefois, jusqu'à aujourd'hui aucun hôte secondaire n'a pu être révélé. En l'absence d'hôte secondaire, le cycle de multiplication d'*Hemileia vastatrix* se décline en urédospores qui contaminent la surface des feuilles, produisant un tube germinatif à l'extrémité duquel se différencie un appressorium. Celui-ci va alors émettre un hyphe infectieux qui pénètre par les stomates pour coloniser la cavité sous-stomatique et produit une cellule mère de l'haustorium. De toute évidence, il n'existe pas de parasexualité chez *H. vastatrix*. Le plus probable est que seule la mutation soit à l'origine de la variation du pouvoir pathogène des rouilles (McIntosh et Watson 1982).



Figure 4. Développement des urédospores à la face abaxiale d'une feuille de caféier

Tableau 2. Liste des génotypes de caféier pour l'étude des races d'*H. vastatrix* du caféier (Bettencourt 1981; Bettencourt et Rodriguez 1988)

Génotype des résistances	Hôtes différentiels	Groupes physiologiques
S <sub>H</sub> ?	Matari	β
S <sub>H</sub> 1	Dilla & Alghe	α
S <sub>H</sub> 1,4	134/4- S 12 Kaffa	I
S <sub>H</sub> 1,5	87/1- Geisha	C
S <sub>H</sub> 1,6	H 468/23	5
S <sub>H</sub> 1,2,5	1006/10-KP532(pl31)	L
S <sub>H</sub> 1,3,5	H 153/2	Z
S <sub>H</sub> 1,4,5	635/3- S 12 Kaffa	W
S <sub>H</sub> 1,4,6	H 539/8	8
S <sub>H</sub> 1,5,6	H 538/29	7
S <sub>H</sub> 1,2,3,5	H 150/8	V
S <sub>H</sub> 1,2,4,5	HW 17/12	O
S <sub>H</sub> 1,3,4,5	H 148/5	U
S <sub>H</sub> 1,2,3,4,5	HW 18/21	S
S <sub>H</sub> 2,5	DK & 6	D
S <sub>H</sub> 2,3,5	34/13- S 353 4/5	H
S <sub>H</sub> 2,4,5	H 152/3	Y
S <sub>H</sub> 2,5,6	H 537/18	6
S <sub>H</sub> 2,3,4,5	H 147/1	T
S <sub>H</sub> 2,3,5,6	H535/10	9
S <sub>H</sub> 3,5	S288-23	G
S <sub>H</sub> 3,4,5	H 151/1	X
S <sub>H</sub> 3,5,6	H 581/17	10
S <sub>H</sub> 4	S12 Kaffa	γ
S <sub>H</sub> 4,5	110/5- S4 Agaro	J
S <sub>H</sub> 4,5,6	H 583/5	11
S <sub>H</sub> 5	Bourbon	E
S <sub>H</sub> 5,6	H 440/7	4
S <sub>H</sub> 5, 7 ou S <sub>H</sub> 5, 7, 9 <sup>a</sup>	7960/15	Sans désignation
S <sub>H</sub> 5,6,9	H 419/20	3
S <sub>H</sub> 5,8	H 420/2	2
S <sub>H</sub> 5,6,7,9	H420/10	1
S <sub>H</sub> 6,?	832/1- H. Timor	A
S <sub>H</sub> ?	H. Kawisari	M
No S <sub>H</sub>	C. racemosa	F
S <sub>H</sub> 6,?	H420/10**	1

<sup>a</sup> Ce génotype attend confirmation

Les effets environnementaux sur le développement de la maladie sont la pluie, la température, la charge fructifère, l'inoculum résiduel, la période de récolte (Bock 1962; Chalfoun 1979; Kushalappa et Eskes 1989). Dans la nature, les facteurs limitants pour la germination et la pénétration sont la sécheresse prolongée et les vents qui sèchent rapidement la face inférieure des feuilles après les rosées et les brouillards. Les attaques de l'*H. vastatrix* débutent par l'apparition de petites taches translucides sur les feuilles des arbustes infectés. Ces taches, dont le diamètre s'accroît, jaunissent et après quelques jours, se forment à la surface inférieure du limbe, de petits groupes de corps granuleux de couleur orangé rouge, fait d'un grand nombre d'urédospores (Figure 4) (Silva *et al.* 1999). La couleur orangée des urédospores (abaxial) vient de la présence dans leur protoplasme de nombreuses gouttelettes huileuses fortement colorées (lipochromes) (Roger 1951). Une attaque sévère se traduit par une défoliation de l'arbre. Une observation en coupe, montrera un mycélium intercellulaire abondant dans le parenchyme lacuneux. Les hyphes émettent de fins filaments qui percent les parois cellulaires et se gonflent à leurs extrémités pour former des haustoria ovoïdes et piriformes (Silva *et al.* 1999).

### 2.1.2 Caractérisation de la résistance du caféier à la rouille orangée

Les premiers travaux sur la rouille orangée furent réalisés par Mayne en 1932 sur la variété Kent's. Le travail de Mayne permit d'identifier 4 races de rouille, qui posèrent les bases de l'étude initiée au Portugal par Oliviera dans les années 50 au CIFIC (Centro de Investigações das ferrugens do Cafeeiro). L'école portugaise a mis en évidence une résistance de type verticale tel que définie par Van der plank (1978). Cette résistance hôte de type qualitatif (gène - pour - gène) (Flor 1971) repose sur la présence chez le caféier d'un petit nombre de facteurs majeurs dominants  $S_H$ . Les génotypes de caféier ont aussi été classés en groupes de résistance ou groupes physiologiques par rapport à leurs interactions avec les différentes races du pathogène de la rouille (Bettencourt *et al.* 1965; Bettencourt et Rodriguez 1988; Tableau 2). La plupart des échantillons incluant les cultivars traditionnels *Bourbon* et *Typica* mis à l'épreuve appartiennent au groupe E ( $S_{H5}$ ) sensibles à la plupart des races. Les individus présents dans le groupe A (ex : hybride du Timor) expriment une résistance face à toutes les races de rouille identifiées à ce jour (D'Oliveira et Rodrigues 1961) et les autres se répartissent dans divers groupes (Tableau 2).

Les géotypes utilisés pour classer les groupes de résistance sont appelés différentiels. Les différentiels sont constitués de variétés de *C. arabica* et *C. canephora* et pour le plus grand nombre d'hybrides spontanés ou artificiels collectés dans le monde. Les races sont celles collectées sur le même lieu ou non que les différentiels. Un minimum de 18 différentiels et 45 races ont ainsi été obtenus de différents pays et sont maintenus par le CIFC (Vàrzea *et al.* 2002; Vàrzea *et al.* 2005). Sur les neuf facteurs de résistance identifiés, *C. arabica* porte les facteurs S<sub>H1</sub>, S<sub>H2</sub>, S<sub>H4</sub> et S<sub>H5</sub>, *C. canephora* : S<sub>H6</sub>, S<sub>H7</sub>, S<sub>H8</sub> et S<sub>H9</sub> et *C. liberica* S<sub>H3</sub> (Noronha-Wagner et Bettencourt 1967; Bettencourt et Noronha-Wagner 1971; Bettencourt *et al.* 1980). Il est ainsi possible à partir de la collection des races d'*Hemileia vastatrix* de déterminer les groupes physiologiques auxquels appartient un caféier quelle que soit son origine.

## 2.2 La lutte contre la rouille

La race II, la plus fréquemment rencontrée, représente 58% de tous les isolats examinés pour la virulence de 32 pays, suivi de la race I (14%), III (9%) et XV (4%) (Eskes 1989).

### 2.2.1 La lutte chimique

Une méthode de lutte chimique contre la rouille à partir de chaux soufrée fut d'abord mise en place par Marshall Ward à Ceylan, puis cette méthode a été re-développée et utilisée en Inde, au Kenya et plus récemment au Brésil, en utilisant principalement des fongicides à base de cuivre (Rodriguez *et al.* 1975). La lutte chimique contre la rouille orangée est un moyen performant, et reste fréquemment utilisé en Amérique latine (Avelino 2002). Toutefois dans un contexte économique difficile (un coût de 10 à 20% du prix de production) et face à la demande du marché pour des produits écologiquement propres, la recherche de variétés résistantes est plus que jamais d'actualité. De plus, dans certaines régions la localisation géographique des cultures d'*arabica* en zone montagneuse rend la lutte chimique impossible.

### 2.2.2 Les sources de résistance naturelles et artificielles

La recherche de plants résistants pour l'amélioration de la résistance des variétés aux attaques de la rouille orangée remonte à 1875. Les premières prospections de plants

Tableau 3. Hybrides interspécifiques spontanés ou artificiels avec *Coffea arabica*

Espèces parentales (femelle x male)	Variétés développées	S ou A*	Années	Provenance
<i>C. arabica</i> x <i>C. liberica</i>	Kalimas	S	1899	Indonésie
<i>C. arabica</i> x <i>C. liberica</i>	Kawisari	S	1899	Indonésie
<i>C. arabica</i> x <i>C. canephora</i>	Bogor Prada Hybride	S	1914	Indonésie
<i>C. arabica</i> x <i>C. canephora</i>	L'hybride du Timor	S	1917	L'île du Timor
<i>C. arabica</i> x <i>C. canephora</i>	Devamachy	S	1930	Inde
<i>C. arabica</i> x <i>C. liberica</i>	S26, S288	S	1932	Inde
<i>C. arabica</i> x <i>C. dewevrei</i>	C387	S	1935	Brésil
Kent's x S288	S795	A	1946	Inde
<i>C. arabica</i> x <i>C. canephora</i> dupliqué	Arabusta	A	1947	Afrique
<i>C. arabica</i> x arabusta	Icatu	A	1955	Brésil
Caturra Vermelho x Hybride du Timor	Catimor, Cauvery	A	1980	Portugal
Villa Sarchi x Hybride du Timor	Sarchimor	A	1980	Portugal
Catuai x Catimor	Cavimor	A	1980	Portugal
Caturra Amarelo x Sarchimor	Cachimor	A	1980	Portugal
Caturra Amarelo a Amarelo x Hybride du Timor	Colombia	A	1980	Colombie

\*S ou A : Hybrides Spontanés ou Artificiels



résistants se sont d'abord portées sur *C. liberica* introduit à Java à partir du Liberia en 1875, puis sur des hybrides spontanés entre *C. arabica* et *C. liberica*, enfin sur *C. canephora* et les hybrides spontanés avec *C. arabica* à partir de 1900 (Tableau 3). Parallèlement, la recherche de nouvelles sources de résistance directement chez *C. arabica* a été conduite en Ethiopie (avant 1900) où se situe le bassin d'origine probable d'*H. vastatrix* (Eskes 1989). Finalement, il s'est avéré que les hybrides spontanés entre *C. arabica* et *C. liberica* d'abord, puis avec *C. canephora* donnèrent les meilleurs résultats.

L'Inde a eu une longue histoire d'amélioration de l'arabica pour la résistance à la rouille dominée par l'apparition répétée de nouvelles races physiologiques de rouille orangée (Carvalho *et al.* 1989). Ceci est attribué aux conditions climatiques chaudes et humides dans les régions productrices de café principalement de l'Inde méridionale, qui sont favorables à une période de latence courte (temps de l'infection à la sporulation) et à la multiplication pendant toute l'année du champignon pathogène. Les dérivés indiens de Devamachy (un hybride spontané de *C. arabica* x *C. canephora* semblable à Hybride du Timor) et des sélections comme S288, qui ont été développées à partir d'hybrides interspécifiques (*C. canephora* 2n x *C. liberica*), puis croisés avec *C. arabica*, ont montré une résistance élevée à la rouille tout en associant de bons rendements et une qualité satisfaisante (Ramachandran et Srinivasan 1979). Toutefois, c'est l'hybride spontané « hybride du Timor », dont les graines envoyées au CIFC, en 1955, furent décrites comme très résistantes à toutes les races de rouille, qui est aujourd'hui à la base des programmes d'amélioration des arabicas à la rouille orangée dans le monde. L'exploitation de la résistance de l'hybride de Timor, en croisement avec des variétés commerciales sensibles à la rouille a abouti à l'obtention de variétés résistantes mises à la disposition des producteurs. Il s'agit principalement des Catimors.

Si certains programmes d'amélioration d'arabica ont réussi (ex : la variété Colombia descendant de l'HdT), l'adaptation des hybrides aux conditions édapho-climatiques et/ou aux conditions de culture des stations d'essais de quelques pays se révèlent souvent difficile. Ce problème d'adaptation est à l'origine d'une productivité faible parfois associée à une qualité inférieure du café - boisson (Gonçalves et Rodriguez 1976), ce défaut se retrouve, notamment en Inde, en Tanzanie, au Kenya, en Angola, au Brésil, au Costa Rica et en Colombie (Bettencourt, 1973; Monaco 1977; Vishveshwara

1971). De plus, récemment, l'observation de résistance incomplète a été notée chez la variété Colombia (dérivant de l'hybride du Timor), qui est pourtant une variété multilignée constituée d'individus appartenant aux différents groupes physiologiques de résistance à cet agent pathogène. Face aux limites intrinsèques de l'hybride de Timor (adaptation aux différents milieux et qualité à la tasse) et à la grande capacité d'*Hemileia* à muter, la recherche de nouvelles sources de résistance s'avère plus que nécessaire.

### 2.2.3 Les recherches génétiques sur la résistance à la rouille

Prakash *et al.* (2004) ont défini un groupe de liaison pour le gène  $S_{H3}$  de résistance à la rouille orangée, qui provient d'une hybridation introgressive naturelle de *C. arabica* avec *C. liberica*. Ils ont trouvé 21 marqueurs d'AFLP groupés dans un seul groupe de liaison de 6.3 cM qui co-segrège avec  $S_{H3}$ . Ce travail est le plus avancé aujourd'hui pour l'étude des gènes de résistance à *Hemileia vastatrix*. La construction et la caractérisation récentes d'une bibliothèque de BAC de *C. arabica* (Noir *et al.* 2004) est une avancée importante pour la réalisation d'une carte physique intégrée du génome du caféier, qui aura beaucoup d'applications telles que la cartographie précise et le clonage de gènes de résistance (Lashermes *et al.* 2005). Cette bibliothèque de BAC contient plus de 88.000 clones avec des tailles d'insertion de 96 - 275 Kb, représentant environ 8 équivalents dihaploïdes de génome de *C. arabica*, avec une couverture égale des deux (Ca et Ea) sous génomes. Le développement de ces outils et l'étude préliminaire sur le  $S_{H3}$  ouvrent la voie pour la localisation génétique et physique d'un locus de résistance à la rouille.

### Qu'est ce qu'un polyploïde ?

Les espèces polyploïdes contiennent plus de deux génomes

- Un génome est l'ADN total pour un jeu de chromosome ( $x$ ), comme trouvé dans les gamètes matures ( $n$  est le nombre de chromosome par gamète)
- Un noyau cellulaire diploïde contenant deux génomes, comme trouvés dans les cellules fécondées
- Un noyau cellulaire polyploïde contenant plus de deux génomes. Deux types de base de polyploïdes sont reconnus : les autopolyploïdes, qui contiennent plus de deux génomes génétiquement identiques; et les allopolyploïdes, qui combinent les génomes de plus d'une espèce ancestrale (ex : un allotetraploïde contient quatre génomes de deux différentes espèces diploïdes ancestrales)

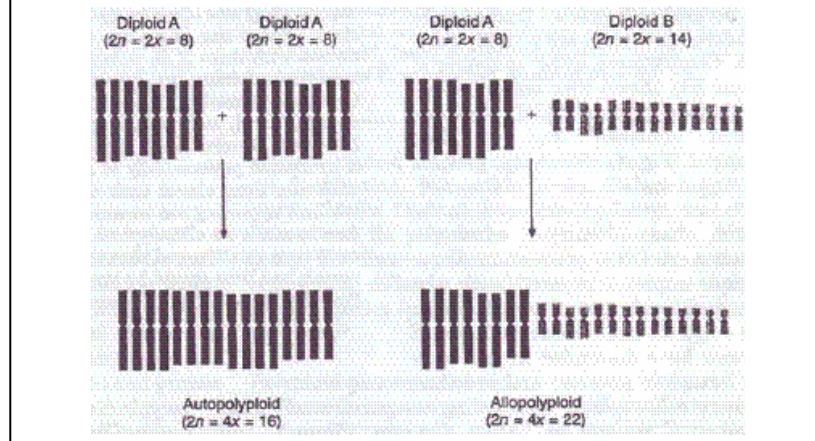


Figure 5. Qu'est-ce qu'un polyploïde ? d'après Leitch et Bennett (1997).

### 3 POLYPLOIDIE ET INTROGRESSION

#### 3.1 Les polyplœides

##### 3.1.1 Généralités sur les polyplœides

Une espèce polyplœide contient plus de deux jeux complets de chromosomes. Le nombre de chromosomes de l'organisme est un multiple ( $>2$ ) du nombre chromosomique de base (Figure 5).

##### 3.1.2 Formation des polyplœides

La formation des polyplœides peut se faire de plusieurs façons : par doublement génomique, production de gamètes non réduits et polyspermie. Le doublement génomique apparaît aussi bien chez les plantes que les animaux et se traduit par un échec de division de cellules suite au doublement des chromosomes lors de la mitose (Hermsen 1984). De même, la production de gamètes non réduits est due à un échec de division des cellules pendant la méiose. La production de gamètes femelles non réduits est fréquemment trouvée chez les animaux et les plantes, de même que la production de pollen non réduit chez les plantes. Le sperme non réduit, cependant, semble jouer un rôle mineur dans la polyplœidisation chez les animaux, peut-être parce que les spermatozoïdes diploïdes sont en concurrence défavorable avec les spermatozoïdes réduits, à une exception près (Alves *et al.* 1999). La polyspermie est également connue chez les plantes et les animaux et est, par exemple, le mécanisme le plus commun menant aux triploïdes humains (Uchida and Freeman 1985), qui représentent 1-3% des conceptions (McFadden *et al.* 1993).

De nombreuses plantes présentent des traces d'événements de polyplœidisation dans leur génome. La polyplœidisation s'est produite chez au moins 70% des plantes angiospermes (Masterson 1994), et chez 95% des ptéridophytes (Grant *et al.* 1981). Des plantes considérées diploïdes comme le maïs, *Arabidopsis*, ou des espèces de *Brassica* sont en fait des espèces polyplœides ou paléopolyplœides, ayant évolué à partir d'un ancêtre polyplœide (Leitch et Bennett 1997). Toutes ces duplications rendent complexe la relation existant entre les gènes. Afin de pouvoir distinguer les gènes au sein d'un

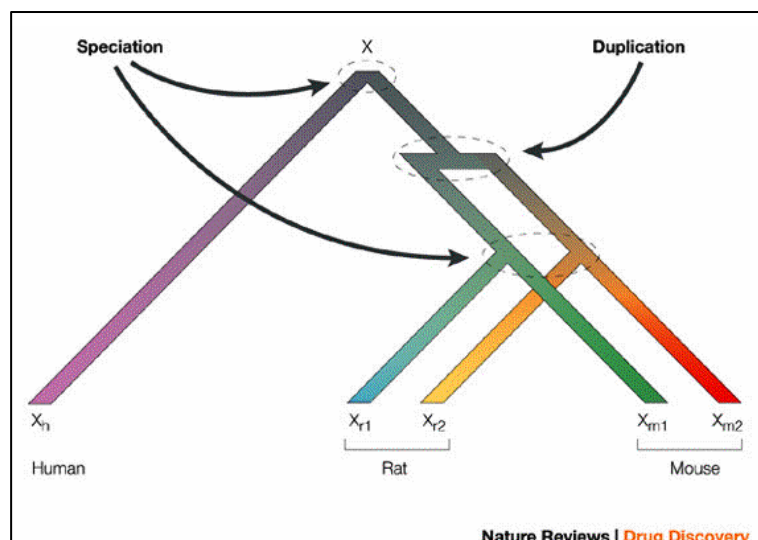


Figure 6. Evolution des relations entre gènes et terminologie.  
 ([http://www.nature.com/nrd/journal/v2/n8/slideshow/nrd1152\\_bx2.html](http://www.nature.com/nrd/journal/v2/n8/slideshow/nrd1152_bx2.html))

organisme et entre organismes, une terminologie a été mise en place (Fitch 2000). Les gènes sont dits homologues quand ils descendent d'un même gène ancestral (avec quelques divergences dues à l'évolution de la séquence nucléotidique) et sont opposés aux gènes dits analogues, gènes qui sont similaires mais ne descendant pas d'un même gène ancestral. La duplication et la divergence des séquences sont des mécanismes de l'évolution des gènes qui, avec la spéciation, entraînent la formation de séquences homologues (Figure 6). Parmi les gènes homologues, on distingue les gènes dits orthologues et paralogues. Des gènes dits orthologues sont des gènes similaires appartenant à des espèces différentes mais descendant d'un ancêtre commun (Figure 6; Xh, Xr et Xm), c'est-à-dire que ce sont des gènes qui dérivent d'un événement de spéciation et non de duplication alors que des gènes dits paralogues sont des gènes issus d'un événement de duplication au sein du génome d'une espèce (Xr1 et Xr2 ou Xm1 et Xm2). Ces gènes évoluent de façons différentes mais de trop grandes modifications de leur séquence entraîneraient une perte de fonction. C'est pourquoi, la structure de ces gènes (découpage intron/exon) est fortement conservée entre gènes orthologues ou paralogues bien que la séquence évolue différemment (mutations, insertions/délétions). La conservation des positions des introns au sein des séquences génomiques a été étudiée pour plusieurs gènes ou familles de gènes paralogues ou orthologues (Sahrawy *et al.* 1996; Proudhon *et al.* 1996; Lunn 2003). Carels et Bernardi (2000) ont montré que la position des introns était généralement conservée entre gènes paralogues et orthologues chez les plantes mais par contre la taille de ces introns peut être très différente.

Comme l'alloploïdie résulte de la fusion de génomes qui sont suffisamment homologues, et que les chromosomes homoéologues (chromosomes partiellement homologues issus d'un événement ancestral tel que la polyploidie) peuvent s'apparier pendant la méiose, les premières générations d'alloploïde doivent subir une forte sélection pour un appariement bivalent exclusif. Ceci est démontré dans les alloploïdes modernes par l'observation de l'appariement des homoéologues qui sont très inférieurs à ce que l'on pourrait prévoir pour des associations de chromosomes formées dans les hybrides entre les diploïdes parentaux existants. Un des meilleurs exemples de ce phénomène se situe chez le blé, où les tétraploïdes et les hexaploïdes montrent la formation de bivalents exclusifs; Il en est de même chez l'allotetraploïde *C. arabica* (Le Pierrès 1995; Lashermes *et al.* 2000a). Les études génétiques classiques ont

révéle, il y a de cela des décennies, l'existence de deux gènes dominants, Ph1, et à un moindre degré Ph2, qui sont responsables du comportement méiotique diploïde lors de l'appariement chez le blé polyploïde (Feldman 1993; Sears 1976; Riley et Chapman 1958). Les données cytogénétiques moléculaires récentes suggèrent que la fonction de Ph1 serait d'agir en tant que « chef d'orchestre local » pour assurer l'association spécifique des centromères entre les chromosomes homologues (Martinez-Perez *et al.* 2001). Récemment, Griffiths *et al.* (2006) proposent le gène *cdc2* qui a un rôle dans la condensation de chromosomes comme candidat pour Ph1. Les espèces allopolyploïdes de *Triticum dicoccoides*, de *Triticum timopheevi* et *Triticum araraticum* ne possèdent pas d'activité Ph1, mais montrent également un comportement méiotique disomique un appariement homoéologue faible (Griffiths *et al.* 2006). Sur cette base, Feldman *et al.* (1997), propose que des facteurs autres que Ph1 sont impliqués dans l'appariement méiotique chez le blé polyploïde (Ozkan *et al.* 2001). Chez *C. arabica*, la recherche d'un gène similaire à Ph1 n'a pas encore été conduite.

### **3.2 L'hybridation introgressive comme source de diversité**

L'hybridation introgressive correspond au transfert et à la stabilisation de matériel génétique entre les hybrides par un backcross avec l'un des parents (Arnold 2004). Les hybrides fertiles issus de croisements naturels sont fécondés par l'une des espèces parentes. L'espèce finit par récupérer les parties de génome qui lui permettent de s'adapter. L'exemple de l'espèce d'*Helianthus annuus* du Texas appelé *spp. texanus* illustre bien ce procédé. Cette plante est issue du croisement naturel entre l'espèce *H. annuus* et *H. debilis* *spp. cucumerifolius*. Cet hybride possède le génome de *H. annuus* ainsi que trois petites régions chromosomiques qui appartiennent à *H. debilis* (Kim et Rieseberg 1999). Ce mode d'hybridation a permis l'adaptation d'espèces à de nouvelles niches écologiques comme l'espèce *H. anomalus* adaptée au sable des dunes et qui possède un génome correspondant à la mosaïque des génomes parentaux (Ungerer *et al.* 1998, Schwarzbach *et al.* 2001). De même, la nature allotétraploïde de *C. arabica* lui permet de coloniser les régions de haute altitude en comparaison de ces diploïdes géniteurs. Le gain de diversité par le flux de gènes est aussi une source de nouveaux individus pour l'amélioration de l'espèce.

### 3.2.1 L'introggression entre espèces

Les études développées sur l'introggression se sont concentrées sur les meilleures méthodes pour insérer des caractères agronomiques importants des espèces sauvages vers un parent cultivé. Pour transférer un gène à travers les barrières reproductrices, l'allèle doit pouvoir se recombinaison dans un nouveau fond génétique (Barton et Hewitt 1985). En conséquence, l'introggression réussie, en serre ou dans la nature, dépendra en partie de l'architecture génétique des barrières reproductrices. Si beaucoup de gènes contribuent à la non-viabilité de l'hybride, alors une grande partie du génome peut être récalcitrant à l'introggression en raison de la présence de gènes « indésirables » (Whittemore et Schaal 1991; Rieseberg et Wendel 1993), en particulier si les taux de recombinaison sont bas. Cependant, si les barrières reproductrices sont dépendantes d'un système génétique simple, la majeure partie du génome devrait être perméable à l'introggression. Seuls les caractères étroitement liés à la stérilité ou aux gènes ayant un impact sur la viabilité seront difficiles à introgresser.

### 3.2.2 L'introggression entre *C. arabica* et les espèces diploïdes

Les problèmes inhérents à l'introggression interspécifique tels que l'instabilité hybride, la stérilité, les ségrégations non-Mendéliennes, et les faibles taux de crossing-over intergénomiques constituent des limites importantes (Stebbins 1958). L'apparition des hybrides spontanés entre *C. arabica* et les espèces relatives diploïdes telles que *C. canephora* et *C. liberica* est fréquente, particulièrement quand ces espèces se trouvent en proximité directe (Cramer 1957). En conditions artificielles, le succès de la création hybrides interspécifiques viables semble dépendre du sens des croisements. Quand *C. arabica* est utilisé en tant que parent maternel, les croisements réussis sont obtenus avec un grand nombre d'espèces diploïdes. En revanche, les tentatives pour produire des hybrides avec *C. arabica* comme parents mâles ont été des échecs (Carvalho et Monaco 1968; Le Pierrès 1995). Les hybrides interspécifiques directs sont habituellement triploïdes et plutôt vigoureux. Tandis que les hybrides hexaploïdes (ex. obtenus par la duplication des hybrides triploïdes) et les hybrides tétraploïdes (ex. résultant de l'hybridation directe ou réciproque entre *C. arabica* et les parents diploïdes auto-tetraploïdisés) semblent raisonnablement fertiles (Berthaud, 1978; Le Pierrès, 1995), les hybrides triploïdes sont quasi-stériles. La méiose dans des hybrides triploïdes est très



irrégulière (Krug et Mendes 1940; Chinnappa 1968; Kammacher et Capot 1972) avec 7.8 à 14.4 univalents dans les associations de chromosomes à la métaphase I, conduisant à des distributions d'anaphase très déséquilibrées. Néanmoins, les gamètes viables se produisent de temps en temps et ces hybrides ont été avec succès rétrocroisés à *C. arabica* (Orozco 1976).

En plus du développement et de la fertilité de l'hybride, l'introgession de gènes d'une espèce à l'autre exige que l'appariement et la recombinaison entre les chromosomes homologues se produisent dans les hybrides. Le comportement du génome de *C. canephora* et son interaction avec le génome de *C. arabica* ont été étudiés dans les hybrides interspécifiques tétraploïdes (*C. canephora* 4x x *C. arabica*) résultant du croisement entre *C. arabica* et un *C. canephora* tétraploïdisé à la colchicine. Le comportement du génome de *C. canephora* dans les hybrides interspécifiques tétraploïdes (*C. canephora* 4x x *C. arabica*) a été étudié avec succès en analysant la ségrégation et la cosegrégation des marqueurs spécifiques de *C. canephora* dans deux descendances BC1 (Herrera *et al.* 2002a). Pour presque tous les locus analysés, la ségrégation des allèles de *C. canephora* transmis par les hybrides (*C. canephora* 4x x *C. arabica*) est conforme à la ségrégation aléatoire attendue de chromosomes et à l'absence d'une sélection. Les taux de recombinaison des segments de chromosome de *C. canephora* ont été estimés pour sept intervalles de marqueurs, et comparés aux taux de recombinaison précédemment observés chez *C. canephora* pour les intervalles équivalents de marqueurs. Les fréquences de recombinaison estimées entre les deux étaient semblables, suggérant que la recombinaison dans l'hybride (*C. canephora* 4x x *C. arabica*) ne serait pas limitée par la différenciation génétique entre les chromosomes appartenant aux différents génomes. *C. canephora* 4x x *C. arabica* a donc semblé particulièrement favorable à la recombinaison intergénomique et aux introgressions de gènes (Herrera *et al.* 2002a).

L'analyse des descendances résultant du croisement de *C. arabica* (parent femelle) et de plusieurs hybrides triploïdes *C. arabica* x *C. canephora* indique une réduction forte des allèles de *C. canephora*, signe d'une contre-sélection importante de l'introgession du matériel génétique du *C. canephora* dans *C. arabica* en utilisant des hybrides triploïdes en tant que parents mâles (Herrera *et al.* 2002b). En 2004, une seconde étude de Herrera *et al.* (2004) a été conduite sur des hybrides triploïdes interspécifiques entre le

tétraploïde *C. arabica* ( $2n = 44$ ) et une espèce diploïde ( $2n = 22$ ), *C. canephora* ou *C. eugenioides*. Les hybrides ont ensuite été rétrocroisés avec *C. arabica* (parent mâle cette fois ci). Le degré d'introgression chez les individus backcross 1 a été estimé à l'aide de marqueurs microsatellites spécifiques de l'espèce. Plusieurs marqueurs introgressés ont été observés chez tous les individus BC1. Néanmoins, alors que la fréquence de marqueurs introgressés au sein des BC1 dérivés d'hybrides triploïdes impliquant *C. canephora* était de l'ampleur attendue, dans le cas d'une ségrégation aléatoire des chromosomes et la formation de gamètes diploïdes, cette fréquence semblait significativement plus faible au sein des individus BC1 issus d'hybrides triploïdes dérivé du *C. eugenioides*. De plus, une comparaison des descendance réciproques entre *C. arabica* et les hybrides interspécifiques triploïdes (*C. arabica* × *C. canephora*), employés comme parent mâle ou femelle, a révélé un très fort biais selon l'orientation du croisement. Ainsi la comparaison des descendance réciproques entre *C. arabica* et les hybrides interspécifiques triploïde (*C. canephora* × *C. arabica*) utilisés comme le parent mâle ou femelle a indiqué un effet très fort de la direction des croisements. Quand des hybrides triploïdes sont employés comme donneur de pollen, l'introgression de gène est rigoureusement réduite. Cette grande réduction a été suggérée comme étant due aux mécanismes liés à la formation des gamètes diploïdes, puisque la présence des allèles de *C. canephora* dans les gamètes fonctionnels a été rapportée ni sensiblement favorable ni nuisible (Herrera *et al.* 2002b). La situation contrastée observée dans le croisement inverse suggère un comportement différent de la méiose mâle et femelle dans les hybrides triploïdes. La formation des gamètes diploïdes fonctionnels portant des introgressions de *C. canephora* serait favorisée pendant la méiose femelle ou empêchée pendant la méiose masculine. Des différences de sexe dans la fréquence de recombinaison ont été rapportées chez un grand nombre de plantes comprenant *Petunia hybrida* (Cornu *et al.* 1989), les espèces de *Brassica* (Kearsey *et al.* 1996) et les espèces d'*Arabidopsis* (Barth *et al.* 2001), les espèces de solanaceae (de Vicente *et al.* 1992).

Les hybrides interspécifiques triploïdes représentent un pont génétique pour le transfert des gènes à partir des espèces relatives diploïdes aux formes cultivées tétraploïdes (Vardi et Zohary 1967). Les connaissances actuelles indiquent que les flux de gènes dans les croisements interspécifiques interploïdies est un phénomène complexe impliquant les barrières naturelles à des niveaux pré et de post-zygotique.

## **4 COMPARAISON DES GENOMES**

### **4.1 La comparaison des génomes**

Tous les génomes actuels des plantes supérieures dérivent d'un même ancêtre commun. Leur évolution a entraîné de nombreuses divergences qui sont à la base de la diversité de formes, de fonctions et d'adaptation des organismes vivants. Ces divergences sont d'autant plus importantes que les espèces sont éloignées. Grâce aux progrès de la biologie moléculaire et de la génétique, il est aujourd'hui possible de comparer les génomes entiers ou en partie de différents organismes ou au sein d'un même organisme (polyploïde par exemple), ce qui a mené à l'essor d'un nouveau domaine au sein de la génomique : la génomique comparative. Celle-ci peut se faire à 2 échelles, selon que l'on s'intéresse à la macrosynténie ou la microsyténie.

Les études macrosyténiques se concentrent sur le génome par une analyse de larges régions (ex : groupes de liaison) en comparant l'ordre des gènes à partir de cartes génétiques. La macrosyténie se fonde sur la comparaison de la position et de l'ordre de marqueurs moléculaires communs et permet ainsi d'avoir une vision globale, sur l'ensemble du (ou des) génome(s) étudiés, des réarrangements et des conservations de blocs synténiques. Ainsi, les cartes génétiques de 2 ou plusieurs espèces sont alignées à l'aide de marqueurs orthologues (locus qui dérivent d'un même locus ancestral par spéciation) et l'unité de distance génétique utilisée est le centiMorgan, cM (pourcentage de recombinaison).

Les comparaisons microsyténiques se basent sur des segments plus petits, mais continus, de régions génomiques complètement séquencées dans lesquelles l'ordre et l'orientation des séquences codantes aussi bien que les sections d'ADN non-codantes sont étudiés. La comparaison peut être faite sur l'ensemble du génome lorsque ces organismes ont été entièrement séquencés ou sur une partie du génome (clones BAC (« Bacterial Artificial Chromosome), par exemple) et permet d'estimer la conservation de synténie au niveau des gènes pour l'ensemble des gènes présents sur le fragment de la séquence d'ADN. Ainsi, les cartes physiques sont comparées en utilisant les séquences d'ADN et la distance est mesurée en paires de bases d'ADN (distance réelle).

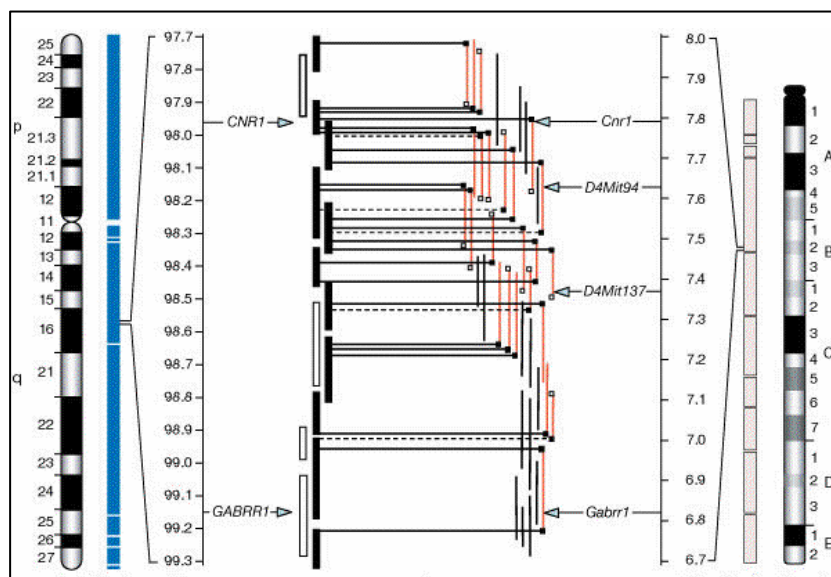


Figure 7. Alignement entre une partie du chromosome 6 de l'homme (*Hsa6*) et du chromosome 4 de la souris (*Mmu4*). (Gregory et al., 2002)

## 4.2 L'histoire des études comparatives

Les premières études de comparaison de génomes ont été faites chez les animaux en comparant des marqueurs moléculaires de cartes génétiques, de cartes d'hybrides de radiation et par hybridation sur chromosome (chromosome painting) entre le génome humain et des espèces proches, toutes étant des mammifères comme les rongeurs (Nadeau et Taylor 1984; Cheng *et al.* 1988; Watkins-Chow *et al.* 1997; O'Brien *et al.* 1999), les ruminants (Echard *et al.* 1994; Gautier *et al.* 2002; Schibler *et al.* 1998) ou le porc (Robic *et al.* 1999; 2001; Rink *et al.* 2002). Ces études ont révélé des conservations de blocs de marqueurs (macrosynténie) entre des espèces animales (Chowdhary *et al.* 1998; Gregory *et al.* 2002; Figure 7).

Chez les plantes, les cartes génétiques d'un grand nombre d'espèces végétales ont été construites, ce qui a permis le développement de la cartographie génétique comparée à partir de la fin des années 1980 (Schmidt 2000). Les premières expériences de comparaison de cartes génétiques ont été faites sur des espèces de la famille des *Solanaceae* (*L. esculentum*, *S. tuberosum*) en utilisant des marqueurs RFLP dérivés de la tomate (Bonierbale *et al.* 1988; Gebhardt *et al.* 1991; Tanksley *et al.* 1992; Livingstone *et al.* 1999). L'ensemble de ces études de comparaison de cartes génétiques a montré qu'il existait une bonne macrosynténie entre les différentes espèces de *Solanaceae*, même si quelques remaniements ont lieu au sein des génomes.

Depuis, d'autres études ont été faites sur d'autres familles notamment chez les *Brassicaceae* dont fait partie la plante modèle *A. thaliana*. La comparaison entre *A. thaliana* et des espèces de *Brassica* a montré que les macro et microcolinéarités sont très conservées entre les *Brassicaceae* et avec *Arabidopsis* (Cavel *et al.* 1998; Acarkan *et al.* 2000; Parkin *et al.* 2002; Grant *et al.* 2000). Chez le blé, Chao *et al.* (1989) ont montré la colinéarité entre les 3 génomes A, B et D du blé hexaploïde. Hulbert *et al.* (1990) ont ensuite démontré que les sondes développées chez le maïs pouvaient s'hybrider avec le sorgho, le millet et la canne à sucre. Ils ont ainsi produit le premier alignement des groupes de liaison entre maïs et sorgho mettant en évidence la similarité des 2 génomes. Cette conservation de l'ordre des gènes dans certaines régions chromosomiques ressort également des comparaisons entre blé et riz (Kurata *et al.* 1994), blé, orge et seigle (Devos *et al.* 1993), sorgho et maïs (Pereira *et al.* 1994), sorgho et canne à sucre (Guimaraes *et al.* 1997; Ming *et al.* 1998), riz et sorgho

## GRASS GENOMES

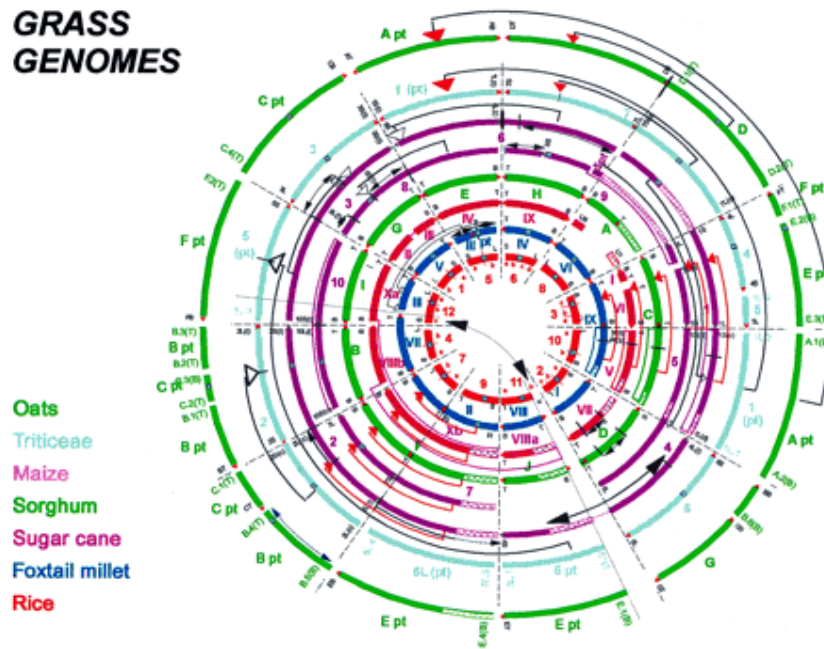


Figure 8. Une carte comparative consensus des graminées de Gale et Devos (1998)

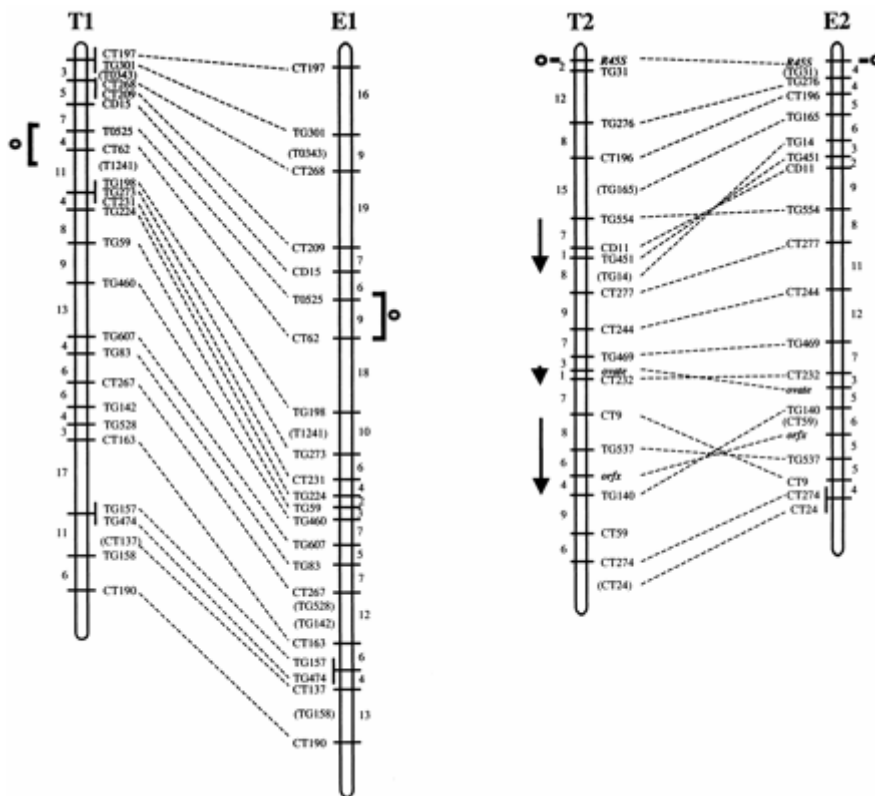


Figure 9. Carte moléculaire de liaison de deux chromosome de l'aubergine et comparaison avec les régions homéologues du génome de la tomate (Doganlar et al., 2002).

(Ventelon *et al.* 2001). Gale et Devos (1998) ont pu ainsi aligner les génomes de plusieurs espèces de monocotylédones; le blé, le maïs, le riz, l'avoine, le sorgho, la canne à sucre et le millet (Figure 8).

### 4.3 Le synténie : implications et utilisations

#### 4.3.1 Evolution de la structure des génomes

La première application de la cartographie comparée est l'étude de l'évolution de la structure des chromosomes. Ainsi la comparaison des cartes génétiques de la tomate et la pomme de terre a mis en évidence l'homologie des 12 groupes de liaison et la conservation de l'ordre des marqueurs. Cinq inversions paracentriques différencient les 2 espèces ont été mises en évidence (Bonierbale *et al.* 1988; Tanksley *et al.* 1992). De même, la comparaison entre les cartes génétiques du poivron, de la tomate, de la pomme de terre et de l'aubergine a permis d'identifier des mécanismes d'évolution du génome de cette famille (Livingstone *et al.* 1999; Doganlar *et al.* 2002; Figure 9). Une trentaine de cassures, comprenant des translocations, des inversions paracentriques et péricentriques, serait suffisantes pour expliquer les différences entre les 4 génomes.

Aussi la comparaison des cartes génétiques entre la tomate et le poivron a mis en évidence outre la conservation de larges régions colinéaires, des réarrangements chromosomiques. Ainsi des inversions paracentriques apparaissent comme le mécanisme premier d'évolution des chromosomes dans la famille des *Solanaceae* et que ces changements structurels ne sont pas constants entre les différentes paires de *Solanaceae* étudiées (Doganlar *et al.* 2002).

La comparaison entre *Arabidopsis* et des espèces de la famille des *Brassicaceae* a permis de mettre en évidence une conservation de la microsynténie avec quelques remaniements, insertion / perte de gène ou duplication de gènes (Acarkan *et al.* 2000, O'Neill et Bancroft, 2000; Parkin *et al.* 2002; Li *et al.* 2003). Environ 90 réarrangements ont été révélés entre *A. thaliana* et *B. nigra* avec des fragments conservés qui ont une longueur moyenne de 8 cM (Lagercrantz 1998).

#### 4.3.2 Connaissance sur l'évolution des génomes

La cartographie génétique apporte des éclaircissements importants sur la structure, l'organisation et sur le mode et la vitesse d'évolution des génomes de plantes (Schmidt, 2000), il devient de plus en plus clair que ce n'est que la comparaison directe des séquences d'ADN de génomes entiers qui pourra élucider les processus de l'évolution de ceux-ci et le rôle joué par les réarrangements chromosomiques dans la spéciation (Mitchell-Olds et Clauss 2002; Sankoff et Nadeau 2003). L'étude de la microsyténie requiert les séquences génomiques et leur annotation pour comparer les différents génomes. De nombreux programmes de séquençage de génomes ont été réalisés ou sont en cours afin de permettre les comparaisons de la microsyténie entre espèces. Le séquençage du génome de la levure (Goffieau *et al.* 1996), de la drosophile (Adams *et al.* 2000), de *Caenorhabditis elegans* (The *C. elegans* Sequencing Consortium, 1998), du fugu (Taylor *et al.* 2002), de l'homme (Dunham *et al.* 1999; The Chromosome 21 mapping and sequencing consortium 2000; International Human Genome Sequencing Consortium 2001; Craig Venter *et al.* 2001; Mungall *et al.* 2003) et d'*A. thaliana* (Lin *et al.* 1999; Mayer *et al.* 1999; Salanoubat *et al.* 2000; Tabata *et al.* 2000; Theologis *et al.* 2000) a été achevé, mais d'autres organismes sont en cours de séquençage (*M. truncatula*, le riz, le maïs, la tomate, le caféier etc.) ainsi qu'un nombre croissant de programmes de génomique d'autres espèces animales et végétales (séquençages d'ARNm, de séquences EST, de fragments d'ADN génomique). Le séquençage de génomes complets et de fragments de génomes (notamment de clones BAC) a permis la comparaison de cartes physiques entre différentes espèces aussi bien animales que végétales.

#### 4.3.3 Amélioration de la cartographie génétique et physique

L'exploitation la plus fréquente aujourd'hui des cartes alignées est l'identification de marqueurs d'une variété d'espèces pour une région génomique donnée. C'est particulièrement utile pour des expériences de cartographie fine ou de clonage positionnel. Ainsi par exemple, la relation de colinéarité entre *Lotus japonicus*, *Pisum sativum* et *Arabidopsis thaliana* a été établie et exploitée pour améliorer la cartographie physique d'un gène impliqué dans la formation de nodules de fixation d'azote chez les légumineuses. Après détermination de la région synténique, un marqueur AUX-1



d'*Arabidopsis* a pu être cartographié proche de la région cible chez *Lotus japonicus* (Stracke *et al.* 2004). De même la microsyténie établie entre la tomate et *Arabidopsis* a permis des améliorations des cartes génétiques et physiques. Ku *et al.* (2000), décrit le premier travail rapportant l'exploitation de microsyténie avec *Arabidopsis* pour des fins de clonage positionnel au sein d'une famille différente au sein du règne végétal. Cette étude a permis de construire une carte génétique à haute résolution pour la région du chromosome 2 de la tomate, co-segreant avec la mutation ovate. Toujours entre *Arabidopsis* et la tomate, la microsyténie a permis de réduire la distance entre les marqueurs encadrant le gène cible *Diageotropia (Dgt)* de 0.8 à 0.15 cM (Oh *et al.* 2002). Pareillement une importante zone synténique entre le chromosome 1 du riz et le chromosome 3 de l'orge (*Hordeum vulgare L.*), a permis de réaliser la saturation de la carte génétique contenant le gène de résistance de la rouille des feuilles de la région chromosomique de l'orge. Cette étude montre aussi clairement l'utilité des ressources génomiques du riz pour la recherche efficace d'EST orthologue de l'orge à des fins de saturation de la région génomique cible chez l'orge (Mammadov *et al.* 2005).

#### 4.3.4 Clonage d'un gène par utilisation de l'orthologie

L'identification d'orthologues putatifs pour des traits monogéniques ou quantitatifs couvrant différentes espèces est facilitée par les comparaisons de cartes (Paterson *et al.* 1995). Ceci a été documenté d'une manière élégante par la cartographie et le clonage des homologues du gène d'*Arabidopsis GAI* qui codent pour un modulateur de la réponse à la gibbérelline chez le maïs, le riz et le blé. Les gènes du blé et du maïs correspondant aux locus *Dwarf* (Peng *et al.* 1999). De même, le gène récessif de résistance contre les potyvirus chez la tomate (*pot-1*) a pu être cartographié et isolé grâce à la colinéarité établie entre la tomate (*L. esculentum*) et le piment (*Capsicum spp.*) et l'utilisation de l'orthologue du facteur d'initiation eukaryotique 4E (eIF4E) du Tabac. Cette démarche éclaire l'utilité de comparaison des cartes pour l'identification de la séquence orthologue du gène récessif de résistance qui a débouché sur une approche gène-candidat. Ce résultat va à contre courant de l'idée d'une absence de gène de résistance dans les régions synténiques. Ceci peut être expliqué par une évolution plus lente des gènes de résistance contre les potyvirus par rapport à la majorité des gènes de résistance clonés jusqu'à aujourd'hui, et peut donc être considéré comme une classe à part de gènes de résistance (Parrella *et al.* 2002; Ruffel *et al.* 2002; 2005).

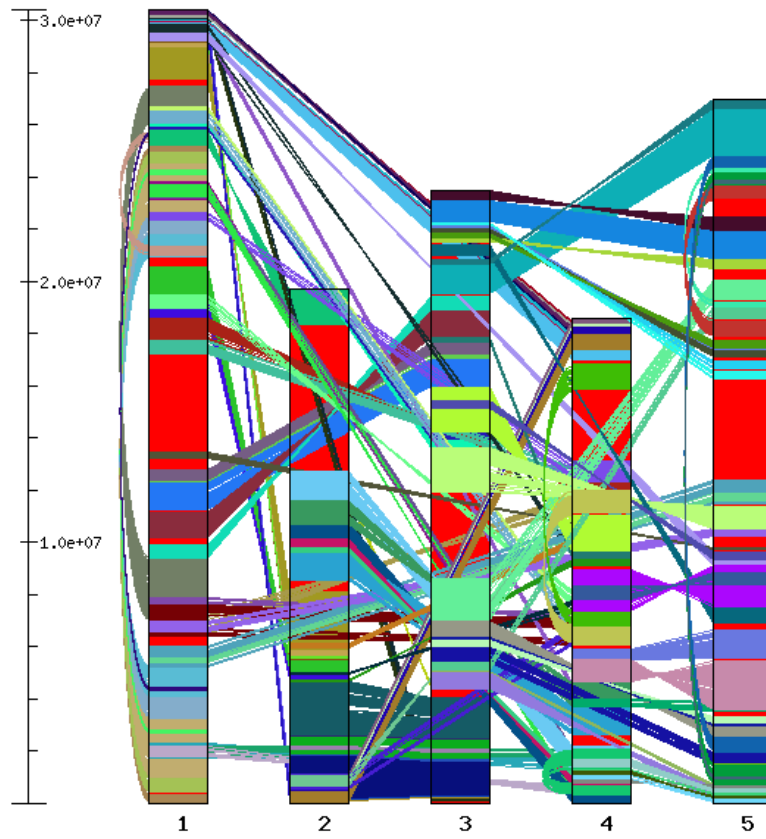


Figure 10. Segments dupliqués d'*Arabidopsis*

[http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis\\_genome\\_duplication.shtml](http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis_genome_duplication.shtml)

#### 4.4 De nouvelles plantes modèles pour la synténie

La disponibilité de la séquence complète du génome d'*Arabidopsis* a permis de comparer le génome de cette plante aux séquences de fragments d'ADN génomiques partiels de différentes espèces végétales. Par exemple entre *Arabidopsis* et des espèces de la famille des *Poaceae* (Van Dodeweerd *et al.* 1999; Mayer *et al.* 2001; Salse *et al.* 2002) ou des espèces appartenant à d'autres familles comme avec les *Fabaceae* (Grant *et al.* 2000; Foster-Hartnett *et al.* 2002; Yan *et al.* 2003) et les *Solanaceae* (Ku *et al.* 2000; Rossberg *et al.* 2001; Gebhardt *et al.* 2003). La macro- et microsyténie sont plus ou moins fortement conservées en fonction des temps de divergence entre les espèces comparées, et même si quelques réarrangements peuvent avoir eu lieu, le transfert d'information d'un organisme à un autre est toujours possible (Schmidt 2000; Bancroft 2001; Barnes 2002). Ainsi les paires les plus étudiées sont *Arabidopsis*-pomme-de-terre (Dominguez *et al.* 2003; Gebhardt *et al.* 2003), *Arabidopsis*-betterave (Dominguez *et al.* 2003), *Arabidopsis*-tournesol (Dominguez *et al.* 2003), *Arabidopsis*-*Prunus* (Dominguez *et al.* 2003), *Arabidopsis*-coton (Bowers *et al.* 2003) et *Arabidopsis*-tomate (Ku *et al.* 2000; Fulton *et al.* 2002; et SOL genomic networks <http://www.sgn.cornell.edu/>)

##### 4.4.1 Synténie entre la tomate et *Arabidopsis*

*Arabidopsis thaliana* est très utilisée comme plante modèle pour l'étude de plusieurs aspects de la biologie des plantes. Des analyses faites sur cette petite crucifère (125 Mbases et 5 paires de chromosomes) à l'aide de programmes « dot-plot » et de « blast » ont montré que le génome d'*Arabidopsis* était constitué d'une mosaïque de régions dupliquées (Figure 10; Blanc *et al.* 2000; Vision *et al.* 2000). Ces nombreuses duplications du génome semblent être le résultat d'au moins 3 événements de tétraploïdisation (Simillion *et al.* 2002; Vandepoele *et al.* 2002; Blanc *et al.* 2003; Ermolaeva *et al.* 2003; Raes *et al.* 2003).

La tomate et *Arabidopsis* appartiennent à deux familles différentes (*Solanaceae* et *Brassicaceae*, respectivement) qui ont divergé tôt dans la radiation des plantes dicotylédones (Figure 11). Il a été déterminé par évidence fossile, que les deux familles se sont séparées il y a plus de 90 millions d'années (MYA) (Gandolfo *et al.* 1998). Les comparaisons de l'ordre de l'ADN mitochondrial placent la divergence à 112-156 MYA

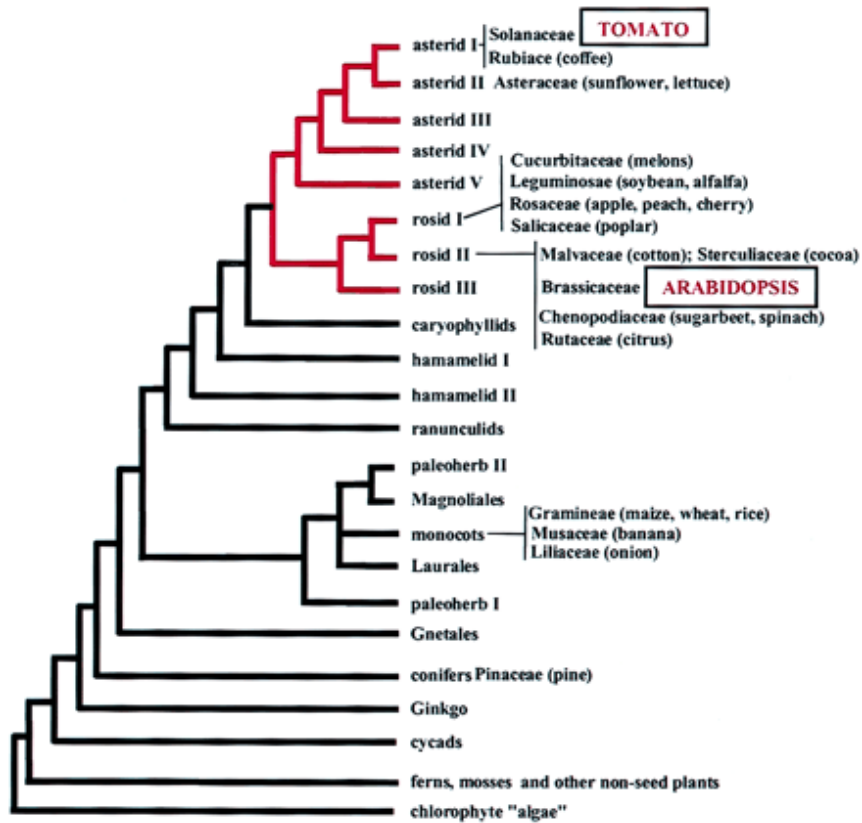


Figure 11. Dendrogramme dépeignant les rapports phylogénétiques des taxons des plantes supérieures. Les lignes rouges indiquent le clade contenant la tomate et l'*Arabidopsis*. Figure basée sur le schéma 2 dans la référence (Chase et al., 1993).

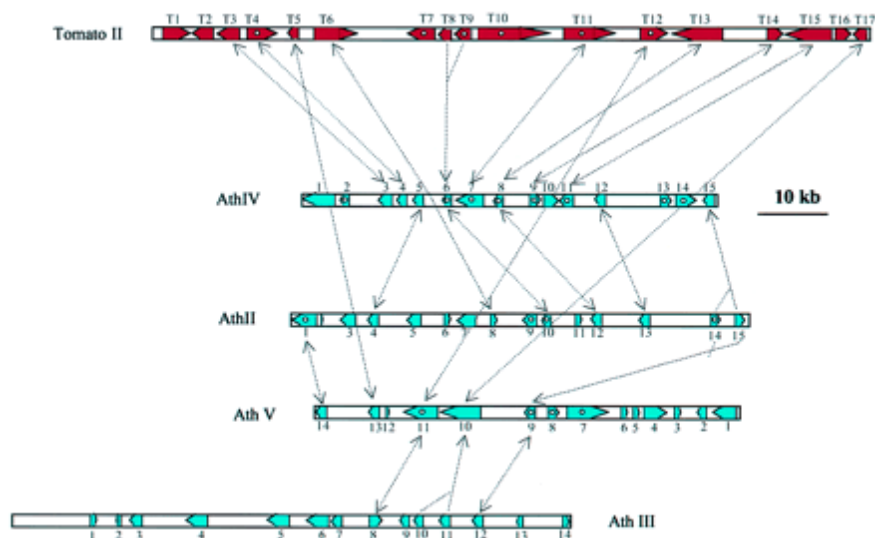


Figure 12. *ORF* commun entre la tomate II (dessus) et les *ORFs* des segments des chromosomes d'*Arabidopsis* 2-5.

(Yang *et al.* 1999). En raison de leur divergence ancienne, une comparaison de la tomate et des génomes d'*Arabidopsis* devrait fournir un aperçu de l'évolution des gènes et des génomes depuis la séparation des dicotylédones et des informations importantes concernant un grand nombre d'espèces (dont beaucoup de plantes cultivées) qui sont phylogénétiquement proches du clade tomate (Figure 11) comme la famille des Rubiacées.

Le séquençage d'une région de 105-kb du génome de la tomate a été réalisée et comparée à *Arabidopsis*. Plutôt qu'un modèle de colinéarité linéaire dans cette région, ils ont décrit des réseaux séparés de synténie qui reflètent les divers événements de duplication de génome (Figure 12) qui ont marqué l'évolution d'*Arabidopsis* depuis la divergence de son dernier ancêtre commun avec la tomate (Ku *et al.* 2000). L'utilisation d'*Arabidopsis* pour la cartographie comparée avec les *Solanaceae* doit donc prendre en compte les différentes duplications d'*Arabidopsis*.

Aujourd'hui les ressources moléculaires de la tomate se sont fortement développées. Ils existent des cartes très denses des chromosomes, plusieurs banques cDNA, YAC (Yeast Artificial Chromosome), BAC et d'autres bibliothèques de données génomiques (Martin *et al.* 1992; Budiman *et al.* 2000; Van der Hoeven *et al.* 2002; <http://sgn.cornell.edu>; <http://www.tigr.org>). Le développement de ressources importantes chez la tomate et le séquençage imminent de son génome en font une plante modèle pour des espèces proches.

#### 4.4.2 Utilisation de la cartographie comparée pour le clonage de gène de résistance

Kilian *et al.* (1997) étaient les premiers à essayer de cloner un gène sur la base de la microsyténie d'une région homeologue d'une autre espèce. Bien que les segments correspondants du chromosome de l'orge (*Hordeum vulgare*) et du riz (*Oryza sativa*) soient clairement homeologues de la région de résistance à la rouille *Rpg1* d'*Arabidopsis*, le gène cible n'a pu être identifié dans l'endroit prévu chez le riz. Cette étude illustre certaines des difficultés liées à l'application des cartes comparatives aux efforts d'isolement de gènes de résistance. En accord avec les travaux suggérant que les gènes de résistance puissent évoluer plus rapidement que d'autres types de gènes (Michelmore *et al.* 1998).

Les gènes de résistance aux maladies sont une classe importante de gènes au plan agronomique; ils sont de mieux en mieux caractérisés au niveau moléculaire. Tandis qu'une bonne conservation de l'ordre et de la fonction de gène a été observée pour les gènes qui régissent des caractères morphologiques et physiologiques dans beaucoup de familles végétales (Gale et Devos 1998), la correspondance structurale et fonctionnelle des gènes de résistance n'est pas encore examinée systématiquement entre les différentes taxons. Si les gènes de résistance suivent les modèles de conservation observés pour d'autres types de gènes des plantes, les locus de résistance de pathogènes identiques ou relatifs devraient être trouvés aux positions génomiques correspondantes chez les autres taxons. Cependant, si les changements moléculaires des locus de résistance sur l'ensemble des taxons divergents modifient rigoureusement la spécificité taxonomique mais pas la fonction, les gènes orthologues de résistance devraient se situer aux positions correspondantes dans des différents taxons, même si ces locus sont impliqués dans la reconnaissance de pathogènes très différents. Une autre alternative est que les gènes de résistance peuvent évoluer tellement rapidement qu'aucune conservation de l'ordre ou de la fonction ne sera évidente à travers des taxons. Les études menées à ce jour permettent d'éclairer le questionnement sur le maintien et l'évolution des gènes de résistance entre les différents taxons et la possibilité d'utiliser les comparaisons génétiques pour une approche d'isolement de gènes candidats.

Les positions génomiques des gènes de résistance aux maladies définis phénotypiquement et les homologues de gène de résistance ont été analysés entre *S. lycopersicum* (tomate), *Solanum* (pommes-de-terre) et *Capsicum* (piment). Les gènes de résistance se sont trouvés aux positions correspondantes entre deux génomes ou plus et ceci de façon plus fréquente que ce que l'on pourrait attendre par hasard (Grube *et al.* 2000); cependant, dans seulement deux cas, à chaque fois impliquant *Phytophthora* spp., les gènes se situaient aux positions correspondantes et ont gardé la spécificité pour des taxons différents du pathogène (Huang *et al.* 2005) ce qui a permis une approche locale de RGA en utilisant l'information génomique de la tomate pour isoler R3a chez la pomme de terre. En revanche, les gènes de résistance au virus de la mosaïque du tabac, et au spotted wilt virus de la tomate ont été cartographiés dans deux ou plus des *Solanaceae*, mais ne se sont pas trouvés en positions synténiques correspondantes. L'analyse et la synthèse détaillées conduites par Grube *et al.* (2000) sur toutes les données disponibles pour les gènes de résistance suggèrent que la spécificité

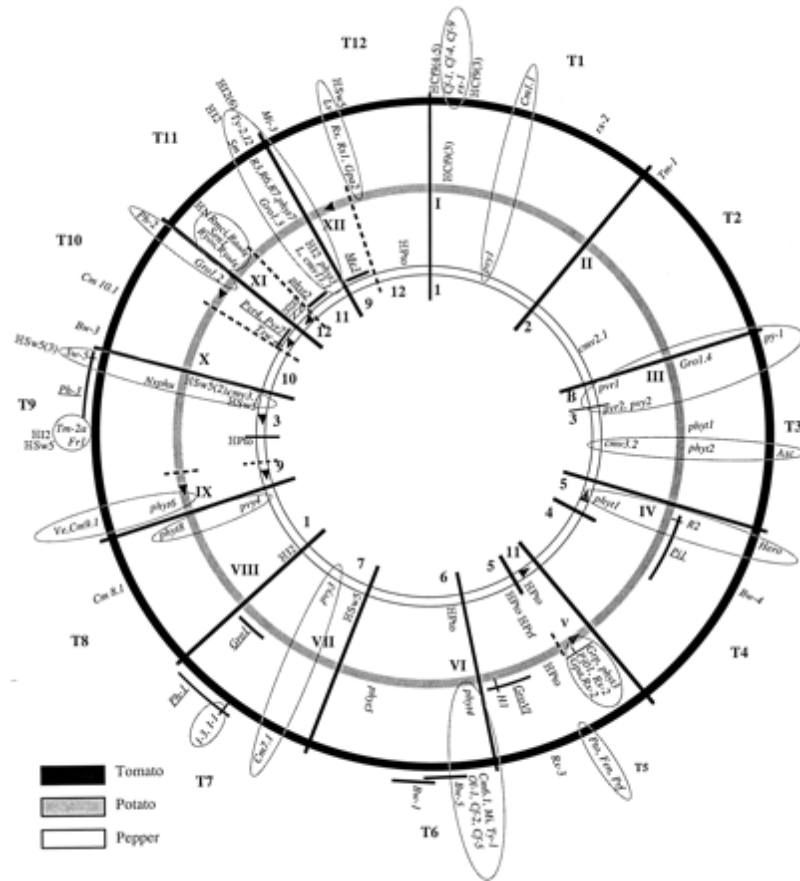


Figure 13. Organisation de gènes de résistance chez les solanacées. (Grube et al., 2000)

taxonomique des gènes de résistance des hôtes peut évoluer rapidement, alors que les fonctions générales des allèles de résistance (par exemple, déclenchement de la réponse de résistance) peuvent être conservées aux lieux homologues à l'intérieur de la famille des *Solanaceae* (Figure 13). Ainsi, l'observation d'une synténie pour les gènes de résistance est contrastée.

Par contre, dans une étude comparative sur les Poaceae, Leister *et al.* (1998) observe des variations très significatives en nombre de copies et position des analogues de gènes de résistance et que les classes spécifiques de gènes de résistance ne se cartographient pas à des positions similaires entre les différents genres ou même au sein d'espèces proches. Bien que cette expérience ne soit pas instructive au sujet de la conservation de la fonction des gènes de résistance, elle suggère qu'il y ait une conservation limitée de la position génomique de ces gènes. Toutefois des exceptions ont pu être observées pour lesquelles la conservation synténique est observée, par exemple entre le maïs et l'avoine, deux sites de résistance de la rouille (conférant la résistance aux agents pathogènes *Puccinia sorghi* et *P. coronata*) et les résistances au mildiou et du « streak mosaïc virus » chez le blé dont les gènes sont localisés sur des blocs homéologues, bien qu'ils puissent être séparés par 30 cM ou plus (Yu *et al.* 1996).

Le travail de Pan *et al.* (2000) nous permet de considérer la génétique comparative entre les gènes de résistance NBS-LRR (Nucleotide Binding Site-Leucine rich repeat) des Solanaceae et d'*Arabidopsis*. Dans cette étude, les séquences génomiques de la tomate montrant une homologie significative à des gènes conférant une résistance race - spécifique à des pathogènes, ont été identifiées systématiquement "en balayant" le génome et en utilisant une variété d'amorces basées sur des motifs NBS consensus. La considération des cartes génétiques comparatives de tomate et de pomme de terre a dévoilé des positions synténiques conservées des homologues de gènes de résistance. Par contre, cela n'a été le cas avec des homologues de gènes de résistance d'*Arabidopsis thaliana*. Ces données indiquent l'évolution remarquablement rapide des homologues de gènes de résistance pendant la diversification des familles des plantes (Pan *et al.* 2000)

Les gènes de résistance peuvent être situés dans des régions du génome pour lesquelles la synténie est mal préservée. Les positions chromosomiques des candidats de gènes de résistance semblent ne pas être préservées entre les espèces graminées (Peng *et al.* 1999). Les homologues du gène *RPM1* sont absents des génotypes sensibles



d'*Arabidopsis* (Leister *et al.* 1998; Foote *et al.* 1997). Dans plusieurs espèces, les gènes de résistance semblent être en région télomérique ou près du centromère. Par exemple, deux des gènes de résistance chez la laitue qui ont été localisés par hybridation *in situ* (FISH), ont été trouvés l'un dans une région télomérique et l'autre centromérique (Lagercrantz *et al.* 1996); *Rpg1* dans l'orge est télomérique (Kilian *et al.* 1997) alors que le gène *Mi* de la tomate est à la frontière de l'hétérochromatine centromérique (Kowalski *et al.* 1994). Il sera intéressant de voir si ces modèles se maintiennent pour les clusters de gènes de résistance dans beaucoup d'espèces parce que les réarrangements chromosomiques impliquent souvent des changements près du télomère et du centromère; la position chromosomique peut donc contribuer, au moins partiellement, au manque de synténie de quelques gènes de résistance. Aussi, la caractérisation des locus correspondant à un segment de *A. thaliana* qui porte le gène de résistance *RPM1* flanqué de deux gènes putatifs, *GTP* et *M4*, chez *B. napus*, a révélé la présence de six loci, dont seulement deux contiennent une copie de *RPM1* (Grant *et al.* 1998).

#### 4.4.3 Les défis de la génomique comparative

Un des défis de la génomique comparative est de distinguer les orthologues (c.-à-d. les gènes homologues avec un ancêtre commun qui ont été séparés par un événement de spéciation) des paralogues (c.-à-d. homologues dérivant d'un événement de duplication de gène); ce qui est particulièrement problématique dans de grandes familles de gènes telles que les gènes de résistance. Un deuxième problème concerne la possibilité d'identifier des segments conservés à l'aide de la cartographie génétique. Les familles multigéniques et les pseudogènes, par exemple, peuvent parfois rendre difficile l'identification du bon gène orthologue, surtout si les espèces comparées sont éloignées phylogénétiquement. En effet dans ce cas, il est difficile de distinguer l'érosion de l'identité de séquence due à la spéciation de celle due à la duplication. De plus, la distance entre deux marqueurs sur une carte génétique correspond à un pourcentage de recombinaison et ne reflète pas du tout le contenu en gènes et la séquence d'ADN de cette région du génome. Entre 2 marqueurs peuvent se cacher des centaines de gènes, et la macrosynténie entre deux espèces n'implique pas forcément l'existence de conservation des gènes et de leur ordre (microsynténie).

## 5 LES OBJECTIFS DU TRAVAIL DE RECHERCHE

La culture du caféier *Coffea arabica*, seule espèce polyploïde du genre *Coffea*, constitue une source de revenu importante et vitale pour l'économie de nombreux pays tropicaux. Toutefois, la très faible diversité génétique des variétés actuelles (i. e. les caféiers cultivés dans le monde dérivent d'un nombre extrêmement réduit d'individus fondateurs) et leur sensibilité à de nombreux parasites et ravageurs, rendent cette culture particulièrement vulnérable aux contraintes biologiques. Notamment, la rouille orangée due à un champignon basidiomycète, *Hemileia vastatrix*, constitue une contrainte agronomique majeure dans la plupart des pays producteurs. Les méthodes actuelles de lutte, essentiellement chimiques, représentent une large part des coûts de production, avec des conséquences néfastes pour l'environnement et la santé humaine. Dans une perspective de développement durable de la caféiculture, la création de nouvelles variétés présentant une résistance durable à la rouille et une bonne qualité pour le consommateur apparaît indispensable.

Les programmes d'amélioration visent à l'utilisation de sources de résistance à la rouille identifiées au sein des espèces diploïdes de caféiers, et notamment les espèces *C. canephora* et *C. liberica*, au travers d'un nombre très limité d'hybrides interspécifiques. Une diversification des sources de résistance apparaît nécessaire. Dans le cadre d'un programme de recherche et développement visant à relancer la filière « café » en Nouvelle-Calédonie (NC), des hybrides interspécifiques naturels entre *C. arabica* et *C. canephora* ont été identifiés dans des plantations abandonnées. Ces hybrides (HNC) montrent une bonne résistance aux races locales de rouille et pourraient constituer une nouvelle source(s) pour l'amélioration de *C. arabica*. Par ailleurs, dans la perspective d'obtenir une résistance durable des variétés améliorées, l'association de plusieurs gènes de résistance au sein d'un même génotype est recherchée. Toutefois, cette stratégie est difficile à mettre en œuvre par les méthodes de sélection conventionnelles et la qualité du café produit par les variétés introgressées n'est pas toujours préservée. La mise en place d'une sélection assistée par marqueurs apparaît donc comme un objectif prioritaire.

Les objectifs de cette thèse, financée par la Province Sud de Nouvelle Calédonie, étaient directement liés à une meilleure connaissance de cette nouvelle ressource génétique que constituent les hybrides HNC, et des bases génétiques et moléculaires de la résistance à

la rouille chez le caféier. Dans les approches entreprises, une large place a été donnée à l'utilisation de marqueurs génétiques moléculaires. Par ailleurs, nous avons évalué la possibilité d'utiliser les données développées sur les organismes modèles *Arabidopsis thaliana* et *Solanum lycopersicum* pour améliorer la cartographie physique d'une région chromosomique de résistance chez le caféier.

Cinq études principales ont été réalisées, les quatre premières font l'objet d'un chapitre rédigé sous la forme d'article scientifique :

- La première étude (Chapitre 2) s'intéresse au mode de formation des hybrides interspécifiques *C. arabica* x *C. canephora* et les flux de gènes entre ces deux espèces dans les conditions naturelles de Nouvelle-Calédonie. Cette analyse s'appuie sur des approches combinées d'outils moléculaires.
- Parallèlement la diversité génétique des hybrides naturels de NC et leur comportement vis-à-vis d'un échantillon représentatif de races de rouille connues se développant sur le caféier *C. arabica* ont été évalués (chapitre 2). Le potentiel de ces hybrides pour l'amélioration de *C. arabica* est mis en valeur.
- La troisième partie est consacrée à l'identification et au développement de marqueurs moléculaires (SCAR, SSR...) d'un gène de résistance à la rouille (i.e. S<sub>H3</sub>) pouvant être utilisés dans le cadre d'une sélection assistée par marqueurs (chapitre 3). Le développement d'une carte physique de la région chromosomique de *C. arabica* correspondant au locus S<sub>H3</sub> a été réalisé.
- La quatrième partie porte sur la mise en évidence et l'évaluation d'une microsynténie entre un segment chromosomique du caféier (région S<sub>H3</sub>) et des segments équivalents chez *Arabidopsis thaliana* (chapitre 4). Les implications évolutives et fonctionnelles de cette microsynténie sont discutées.
- Enfin, la cartographie de la région chromosomique correspondant au locus S<sub>H3</sub> a été réalisée sur *Solanum lycopersicum* (chapitre 4).

A l'issue de ces travaux, une synthèse des résultats obtenus (chapitre 5) est présentée et des perspectives de recherches sont exposées.

## CHAPITRE II

Développement des ressources naturelles pour  
l'amélioration de *C. arabica* : Etude des hybrides  
interspécifiques naturels de Nouvelle-Calédonie

Chapitre rédigé sous la forme d'un article accepté à GENOME

Introgressive hybridization between the allotetraploid *Coffea arabica* and one of its diploid ancestors *C. canephora* in an exceptional sympatric zone in New Caledonia

L. MAHÉ, D. LE PIERRÈS, M-C. COMBES and P. LASHERMES

## **6 *Introgressive hybridization between the allotetraploid Coffea arabica and one of its diploid ancestors C. canephora in an exceptional sympatric zone in New Caledonia***

### **6.1 Abstract**

In the past ten years, a lot of information has been gathered on genetic, epigenetic and genomic effects of polyploidy and especially the role of these attributes in the success of polyploid establishment and evolution. In contrast, information on gene flow between polyploid and their diploid relatives is very limited, even though they can be used to infer the selective forces that led to speciation in the first place. Here, we analyzed interspecific gene flow between the allotetraploid *Coffea arabica* and one of its ancestral diploid progenitors *Coffea canephora* in a sympatric zone of New Caledonia. The objectives were to investigate the processes governing hybrid formation and gene introgression in natural conditions. Occurrence of a large assortment of hybridization events between the two coffee species is clearly established. Although bi-directional mating was observed, only unidirectional introgression from *C. canephora* into *C. arabica* was noted in later generation hybrids. We proposed to consider a phase of stable coexistence between *C. arabica* and its diploid progenitor species at the time of *C. arabica* formation, supporting the suggestion of an important role for gradual, population processes governing the evolution of polyploidy in natural populations.

Most of the collected post-F<sub>1</sub> hybrid plants exhibited a high level of introgression (average 0.73 and 0.68 for SSR- and AFLP-based introgression indices respectively) suggesting no counter selection against the introgression of *C. canephora* genetic material. Altogether, the New Caledonian central mountains appear to be a highly favourable environment for introgressive hybridization and a genetic diversity center for *C. arabica*.

### **6.2 Introduction**

Genome introgression appears particularly important in plant evolution (Arnold and Hodges 1995; Arnold 1997). It occurs in regions where genetically differentiated taxa meet, mate and hybridize. These regions are auspicious for introgressive hybridization, which can be defined as the movement of genes between species mediated by backcrossing or more broadly defined as the transfer of genes between genetically distinguishable

populations (Rieseberg and Carney, 1998). Beneficial consequences have been proposed for introgression of genetic material including: increased genetic diversity (Beaumont et al. 2001; Sawkins et al. 2001, Tsukaya et al. 2003), and local adaptive evolution (Arnold 1997; Rieseberg and Carney 1998; Martin et al. 2006). However, introgression can also have a negative effect on the survival of certain plant species. For example Levin et al. (1996) showed how hybridisation might contribute to the extinction of rare plant species through demographic swamping and genetic assimilation by an abundant congener.

Among the 103 species, so far described (Davies et al. 2006), belonging to the subgenus *Coffea*, two species are of economic importance, *Coffea arabica* L. and *Coffea canephora* Pierre (Charrier and Berthaud 1985). *Coffea* species are diploid ( $2n=2x=22$ ), except *C. arabica* ( $2n=4x=44$ ) which is self-fertile and recognized as an allotetraploid (i.e.  $C^aE^a$  genome) formed by hybridization between two diploid species *C. eugenioides* (E genome) and *C. canephora* (C genome) (Lashermes et al. 1999). Low divergence between the two diploid constitutive sub-genomes of *C. arabica* and those of its modern-day progenitor species was demonstrated (Lashermes et al. 1999; Herrera et al. 2002a). At the present time, *C. arabica* is endemic to inter-tropical forests in the highlands of South West Ethiopia and the Boma Plateau of Sudan, and it is characterized by a low genetic diversity (Lashermes et al. 1999). In those environments, *C. arabica* appears as the only coffee tree species and gene exchange with diploid species can be ruled out. *C. canephora* originated in the lowland forests of tropical Africa, which stretch from guinea to Uganda and Angola (Charrier and Berthaud 1985).

In relation with the worldwide development of coffee growing, *C. arabica* trees are gathered occasionally with diploid species representatives in coffee-producing areas. Occurrence of spontaneous hybrids between *C. arabica* and diploid relative species such as *C. canephora* and *C. liberica* is from time to time reported, especially when these species grow in direct proximity (Cramer 1957). Although their genesis and evolution have never been investigated, those rare hybrids have been intensively used in coffee breeding programmes as the main source of resistance to pests and diseases. Recently, prospecting missions in New Caledonia (*Pacific island*) for genetically original material have revealed several areas of sympatric occurrence of both introduced species *C. arabica*, *C. canephora* and plants showing intermediate phenotypes (Charmetant and Le Pierres 1991; Le Pierres 1999). These coffee fields were abandoned in the 1940s and could be considered as neo-

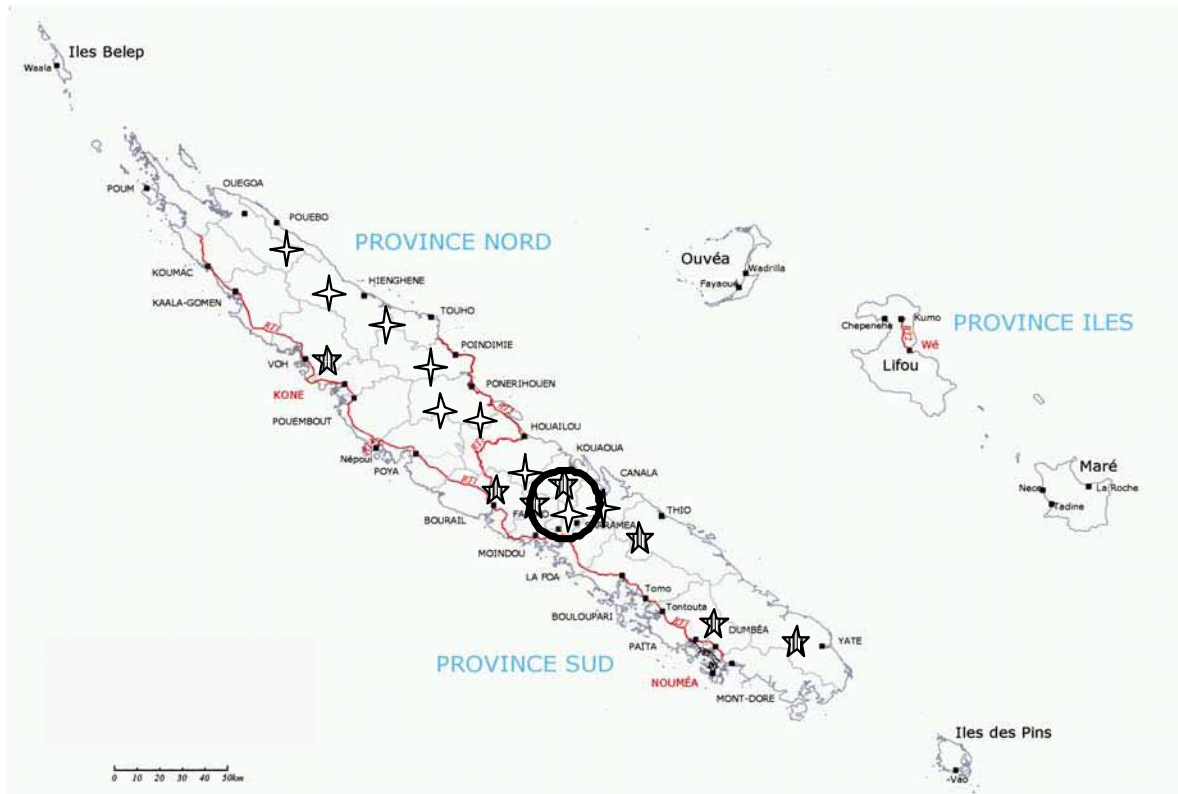


Figure 14. Localization of sites in New Caledonia where *C. arabica* (striped stars), *C. canephora* (white star) are reported. The sampled sympatric zone (Col d'Amieu) between *C. arabica* and *C. canephora* is identified by an open circle.



natural hybrid zone. Our objectives were to ascertain the presence of hybrid plants in those exceptional areas, to investigate the processes governing hybrid formation and gene introgression between two coffee species, *C. arabica* and *C. canephora*, and to evaluate the diversity of this genetic material.

## **6.3 Materials and Methods**

### **6.3.1 Sampling**

Leaves from adult plants were sampled at five sites around le Col d'Amieu in New Caledonia (Figure 14) : Friquet (F; 21°40', 165°45'), Szemmelweis (S; 21°39', 165°45'), Leonard (L; 21°39', 165°44'), Couli (C; 21°37', 165°49') and close to the Foa valley for Dogny (D; 21°39', 165°56). Phenotypically they were classified as 42 hybrids, 7 *C. canephora* plants, and 3 *C. arabica*. For comparison, three *C. arabica* varieties namely Bourbon, Typica and Le Roy, and a *C. canephora* accession (IF200) were also included as references. Hybrid plants were collected on the bases of their intermediate leaf size and *domatia* number between *C. canephora* (big and wide leaves with no *domatia*) et *C. arabica* (small and narrow leaves with *domatia*).

### **6.3.2 Determination of nuclear DNA content**

The total DNA amount in nuclei of collected coffee plants was assessed by flow cytometry (Dolezel et al, 1989). Segments from fresh leaves were analysed as reported by Herrera et al. (2002b). Each measurement was repeated at least twice. To determine the ploidy level of hybrids, the peak values were compared with the values determined (four replicates) for the diploid (*C. canephora*, acc. EC37) or tetraploid (*C. arabica*, acc. typica) parental species.

### **6.3.3 AFLP assay**

The AFLP protocol described by Lashermes et al. (2000) was followed. A total of 12 primer pairs were used in this analysis: E-ACG M-CTG, E-AAC M-CAA, E-ACA M-CAT, E-ACG M-CAA, E-ACA M-CTT, E-AAC M-CTT, E-ACG M-CTA, E-ACT M-CTT, E-ACT M-CTA, E-AAG M-CTA, E-ACG M-CAT, E-ACT M-CAA. The AFLP amplification products were designated in order of decreasing fragment size and according

to the primer combinations used. Only polymorphic clear bands were scored in a binary data matrix.

#### 6.3.4 Microsatellite assay

In total twenty-three SSR were selected for this analysis. Sixteen SSR loci revealed constantly distinct alleles between *C. arabica* and the cultivated accessions of *C. canephora* (Herrera et al. 2002a, Prakash et al. 2005). Those 16 SSR loci were further considered as species diagnostic SSR loci. Among the 23 SSR loci studied, 16 were previously mapped and are distributed on 7 of the 11 chromosomes of *C. canephora* (Lashermes et al. 2001). The PCR assay of specific SSR primer pairs was carried out as reported by Combes et al. (2000).

#### 6.3.5 DNA chloroplast identification

The simple cleaved amplified polymorphic sequence (CAPS) procedure developed by Cros et al. (1998) was used to determine the type (*C. arabica* vs *C. canephora*) of cpDNA present in the collected plants. A region (*atpB-rbcL*) of the chloroplast genome corresponding to the intergenic spacer region between *atpB* (ATP synthase beta subunit) and *rbcL* (RuBisCO large subunit) genes and the first 56 codons of *rbcL* was amplified using the primers E: 5' GGTCAAGTCCCTCTATCCC and F: 5' ATTTGAACTGGTGACACGAG. Then, the amplified products were sequentially digested by *AluI* and *Sau3A*. Conditions of amplification and digestion were as described by Lashermes et al. (1996a). Revelation was performed by electrophoresis on agarose gels. The *C. arabica atpB-rbcL* region is characterized by fragments of 269 bp and 81 bp in length, while *C. canephora atpB-rbcL* generates 3 fragments of 136, 117 and 81 bp, respectively.

#### 6.3.6 Data analysis

Introgression indices reflecting the proportion of *C. canephora* genome introgressed were calculated for all hybrid plants. The first index corresponds to the frequency of diagnostic SSR loci showing at least one canephora allele. The second index corresponds to the number of observed AFLP canephora marker alleles in a given hybrid plant divided by the estimated number of AFLP canephora marker alleles detected in a theoretical interspecific

hybrid *C. arabica* x *C. canephora* using the same set of primer combinations. This theoretical number of AFLP *canephora* markers was estimated as the mean number of AFLP markers differentiating the 7 accessions of *C. canephora* and the 3 accessions of *C. arabica* included in this study, multiplied by 0.85 to take into account that about 30% of AFLP markers are in the heterozygous state in *C. canephora* (Lashermes, unpublished data).

## **6.4 Results**

### **6.4.1 Molecular authentication of interspecific hybrid plants collected in New Caledonia**

The 42 putative hybrid plants collected (i.e. 5 sites in Central mountains; Figure 14) in the New Caledonian sympatric zone between *C. canephora* and *C. arabica* were analyzed using molecular markers and compared to collected samples and representative accessions of both coffee species *C. arabica* and *C. canephora*.

Twelve AFLP primer combinations, which generated clear patterns with coffee accessions in a previous study, were used (Herrera et al. 2004). The size of the AFLP fragments in these experiments ranged from approximately 40 to 400 bp. The number of scored amplified bands per sample and per gel varied between 15 and 30, depending on both the primer combination and the accession considered. As anticipated, the six samples of *C. arabica* (i.e. 3 collected samples and 3 accessions) appeared very similar (i.e. 2 polymorphic bands) while the number of polymorphic bands within the *C. canephora* samples (i.e. 7 collected samples and 1 accession) was very high (i.e. 143 polymorphic bands). A total of 410 AFLP fragments were scored in the group of putative hybrid plants. All the AFLP fragments (i.e. 267) observed in *C. arabica* accessions were identified within the group of putative hybrid plants. In addition, 153 AFLP fragments not detected in any of the accessions of *C. arabica* were identified. The low genetic diversity established in *C. arabica* allow to consider them further as *canephora* markers. In spite of the limited number of *C. canephora* accessions included in this study, 139 of these 153 marker bands were also observed in at least one of the *canephora* samples analyzed, confirming their *canephora* origin. Furthermore, the total number of *canephora* markers varied from 22 to 60 among the putative hybrid plants.

Table 4. List and characterization of identified first-generation hybrids.

Plant codes	Origin (site)	Ploidy level	CpDNA type	Presence of locus exhibiting two <i>C. canephora</i> alleles
<i>Leonard</i>				
L 1		4 X	<i>C. arabica</i>	
L 2		4 X	<i>C. arabica</i>	+
<i>Couli</i>				
C 23		4 X	<i>C. canephora</i>	
C 28		3 X	<i>C. arabica</i>	
C 30		4 X	<i>C. arabica</i>	
C 34		4 X	<i>C. arabica</i>	+
C 38		4 X	<i>C. arabica</i>	+
C 39		4 X	<i>C. canephora</i>	
C 40		3 X	<i>C. canephora</i>	
<i>Friquet</i>				
F 5		3 X	<i>C. canephora</i>	
F 6		3 X	<i>C. arabica</i>	
<i>Dogny</i>				
D 92		4 X	<i>C. arabica</i>	

The 23 SSR primer combinations selected for this study gave clear amplification profiles that could be easily scored in each of the 56 samples analyzed. All these SSR markers were polymorphic and the banding patterns resolved by each primer pair were in accordance with single locus variation and the allotetraploid structure of *C. arabica*. For 21 of the 23 SSR loci, identical alleles were observed within the six samples of *C. arabica*. Polymorphism was only observed at two SSR loci. In total, 153 alleles were identified within the group of putative hybrid plants. Of these 153 alleles, 32 are equivalent to alleles of *C. arabica* and 121 correspond most probably to alleles from *C. canephora*, although merely 84 were detected in the set of *C. canephora* samples included in this study. The number of SSR loci exhibiting canephora alleles in a given plant varied from 2 to 23 among the putative hybrid plants.

Altogether, both AFLP and SSR analyses clearly ascertain an interspecific origin (*C. arabica* x *C. canephora*) for the 42 plants presenting a “hybrid” phenotype collected in the New Caledonian central mountains.

The progeny generation of collected hybrid plants was determined based on the analysis of SSR loci (i.e. diagnostic loci) that exhibit alleles differentiating the two potential parental species, *C. arabica* and *C. canephora*. The 42 hybrid plants were assessed using the 16 diagnostic SSR loci. Since the arabica alleles are always observed in the hybrid plants, only the canephora alleles appeared informative. The plants exhibiting at least one canephora allele for all of the loci examined were considered as first-generation hybrids (i.e. F<sub>1</sub>), while the plants showing loci with only arabica alleles were considered as post-F<sub>1</sub> plants. The ability to distinguish among post-F<sub>1</sub> generations was not investigated since the expected canephora allele marker frequency distribution for one or more different backcross generations or for selfing can overlap, and because of the required assumption of random chromosome segregation. Among the 42 hybrid plants, 12 appeared as F<sub>1</sub> plants and 30 as post-F<sub>1</sub> plants (Table 4). The F<sub>1</sub> plants results were confirmed when using the 7 additional SSR loci. In addition, the 12 identified F<sub>1</sub> plants had values of the AFLP introgression index close to 1, confirming their F<sub>1</sub> status and a good fit of the data obtained by the two marker types. Furthermore, F<sub>1</sub> plants were identified at 4 of the 5 collected sites in association with post-F<sub>1</sub> plants. In contrast, only post-F<sub>1</sub> plants were detected at the Szemmelveisz site (S).

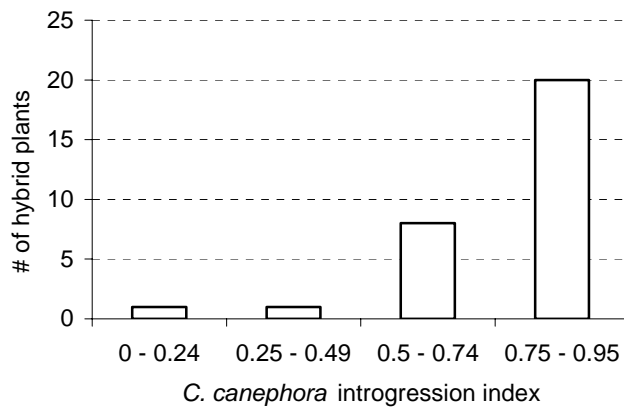


Figure 15. Distribution of post-F1 hybrid plants according to the *C. canephora* introgression index based on SSR data.

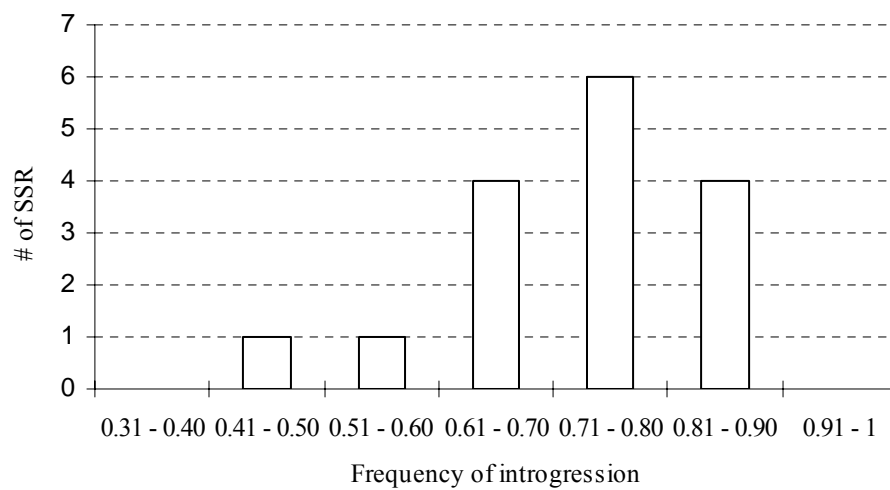


Figure 16. Distribution of SSR loci according to their frequency of introgression among the 30 post-F1 hybrid plants. For each individual post-F1 plant, the locus was considered introgressed when exhibiting at least one *C. canephora* allele.

The total DNA amount in nuclei of hybrid plants was assessed by flow cytometry. To ascertain their ploidy level, the 2C nuclear DNA contents were compared with the values determined for the two parental species *C. canephora* and *C. arabica*. Two groups of F<sub>1</sub> plants were identified (Table 4). One group of 8 plants exhibiting a DNA content equivalent to the species *C. arabica* were considered as tetraploid (4x). The second group involving 4 individuals exhibited a relative DNA content intermediate between the two parental species (i.e. *C. arabica* and *C. eugenioides*) and were considered as triploid (3x). Regarding the post-F<sub>1</sub> plants, a continuous and large variation in DNA content was observed. A tendency could be identified and resulted in a majority of plants closed to the tetraploid level. However, the possibility of plants involving aneuploid gametes from disturbed meiosis of F<sub>1</sub>'s precluded a precise categorization of post-F<sub>1</sub> plants.

#### 6.4.2 Gene flow between *C. arabica* and *C. canephora*

In order to unravel the dynamics of gene flows between the two species, the maternal affiliation of hybrid plants was investigated. Considering that cpDNA is inherited through the maternal parent in *Coffea* species (Lashermes et al. 1996a), the *atpB-rbcL* region was used as marker (Cros et al. 1998) to determine whether the cpDNA of the different hybrid plants originates from *C. arabica* or from *C. canephora*. Except for two plants that did not amplify, the maternal identities of hybrid plants were successfully established: 35 hybrids (80%) showed *C. arabica* chloroplast DNA while 5 presented *C. canephora* cpDNA. From these results we conclude that the F<sub>1</sub> hybrids were preferentially formed with *C. arabica* as female parent. However, hybrid plants exhibiting canephora-derived cpDNA were also observed among both the F<sub>1</sub> plants and the post-F<sub>1</sub> plants.

The level of *C. canephora* genome introgression in the post-F<sub>1</sub> plants was evaluated. Introgression indices reflecting the proportion of *C. canephora* genome introgressed were calculated based on either SSR or AFLP markers for the 30 inferred post-F<sub>1</sub> plants. Both indices showed a high significant correlation ( $r = 0.84$ ,  $P < 0.001$ ) and revealed a high level of canephora genome introgression in most post-F<sub>1</sub> plants (Figure 15). The mean values of SSR- and AFLP-based introgression indices were 0.73 and 0.68, respectively (data not shown). In addition, 87% of post-F<sub>1</sub> plants had an introgression index higher than 0.50, whatever the introgression index considered. Also, the few plants (e.g. L8, D87, S10, S20) showing the lowest introgression rate originated from three different sites.

Table 5. List of identified *C. canephora* alleles for all analyzed SSR loci. For each allele, the presence in the different populations (Dogny, D; Couli, C; Leonard, L; Szemmelveisz, S; Friquet, F) and mean frequencies are reported.

Locus	Alleles	Populations					Mean	Locus	Alleles	Populations					Mean	Locus	Alleles	Populations					Mean	
		D	C	L	S	F				D	C	L	S	F				D	C	L	S	F		
SAT32	Ca	+	+	+	+	+	0,31	SAT171	Ca	+					0,05	SAT235	Ca	+	+				0,04	
	Cb		+			+	0,09		Cb	+	+				0,05		Cb	+	+		+	+	0,36	
	Cc	+					0,03		Cc	+	+		+	+	0,33		Cc					+	0,04	
	Cd	+	+			+	0,15		Cd	+					0,03		Cd	+	+	+	+	+	0,22	
	Ce	+		+	+		0,18		Ce				+	+	0,08		Ce	+	+			+	0,22	
	Cf		+				0,03		Cf		+	+		+	0,29		Cf	+					0,03	
SAT41	Ca	+	+				0,07	Cg			+			0,02	SAT251	Ca		+				0,03		
	Cb	+	+				0,08	SAT177	Ca	+	+		+	0,14		Cb		+		+		0,07		
	Cc	+	+				0,06		Cb	+	+			0,04		Cc	+	+			+	0,12		
	Cd	+	+		+		0,08		Cc	+	+			0,10		Cd				+	+	0,12		
	Ce	+	+				0,06		Cd	+	+	+		0,31		Ce	+	+	+		+	0,31		
	Cf	+			+		0,15		Ce		+			0,08		SAT252	Ca		+				0,02	
Cg		+	+		+	0,23	SAT189		Ca				+	0,04	Cb			+				0,06		
SAT47	Ca	+	+	+				0,20	Cb			+		0,03	Cc		+	+		+	+	0,33		
	Cb	+	+		+	+		0,30	SAT193	Ca	+	+	+	+	+		0,81	Cd		+	+		+	0,29
	Cc	+	+	+		+		0,29		SAT225	Ca		+		+		0,13	Ce	+	+			0,07	
	Cd	+	+					0,06			Cb		+				0,06	SAT253	Ca		+	+		0,09
	Ce	+	+					0,04			Cc	+	+	+		0,36	Cb			+			0,01	
	Cf		+				0,01	Cd			+	+	+		0,22	Cc	+		+	+	+	0,20		
SAT65	Ca	+	+	+	+	+	0,93	Ce			+				0,03	Cd	+		+		+	+	0,28	
	SAT109	Ca	+	+	+		+	0,51	SAT227		Ca				+	0,04	Ce		+	+	+	+	0,30	
		Cb	+	+		+	+	0,21		Cb	+				0,03	SAT254	Ca			+	+		0,08	
		Cc	+	+				0,05		Cc				+	0,04		Cb		+			0,04		
		Cd		+				0,01		Cd				+	0,04		Cc	+	+			0,04		
		SAT157	Ca	+	+	+		+		0,36	Ce	+	+		+		0,13	Cd	+	+		+	0,22	
Cb			+	+	+	+	+	0,67		Cf		+			0,01		Ce	+			+	0,16		
SAT160	Ca		+	+		+	+	0,43	Cg		+		+	0,05	Cf		+				0,03			
	Cb		+	+	+			0,07	Ch		+		+	0,09	Cg				+	0,04				
	Cc		+	+	+			0,16	Ci	+	+	+		0,16	Ch				+	0,01				
	Cd		+					0,03	Cj	+	+	+		0,19	Ci		+			0,01				
	SAT166	Ca	+	+	+		+	0,24	Ck	+	+			0,06	SAT262	Ca					+	0,04		
		SAT170	Ca	+	+	+	+	+	0,77	SAT228	Ca	+					0,03	Cb	+	+		+	0,09	
Cb				+				0,02	Cb		+	+	+	+		+	0,50	Cc	+	+	+	+	0,17	
SAT280			Ca															Ca					+	0,04
			Cb															Cb	+		+	+		0,18
			Cc															Cc	+	+				0,06
	Cd															Cd	+			+		0,09		
	Ce													Ce	+	+		+		0,19				



The frequency of introgression for the 16 diagnostic SSR loci among the 30 post-F<sub>1</sub> hybrid plants were analyzed (Figure 16). For each individual post-F<sub>1</sub> plant, the locus was considered introgressed when exhibiting at least one *C. canephora* allele. The frequency of introgression observed among the different analyzed loci in the post-F<sub>1</sub> plants varied from 0.4 to 0.9 with an average value of 0.7 reflecting a rather homogeneous behaviour for introgression of the chromosomal region linked with the different SSR loci.

#### 6.4.3 Genetic variation of hybrid plants

The genetic variability of hybrid plants from the 5 collected sites was evaluated. The *arabica* alleles were invariably present among the hybrid plants at all loci except the locus M177, which showed a polymorphism similar to the variation observed among the *C. arabica* accessions. In contrast, a considerable variation for the *canephora* alleles was observed. The different *canephora* alleles identified for each of the 24 SSR loci analyzed are listed in Table 5. The total number of *canephora* alleles per locus varied from 1 to 11, with a mean of 4.6. Most alleles were shared by plants originating from different collecting sites. However, few alleles appeared restricted to one single population. Differences between sites were observed. While the plants collected at the sites Digny (D), Couli (C) and Friquet (F) showed marked polymorphism, the plants originating from the Leonard (L) and Szemmelveisz (S) sites appeared more homogeneous. Moreover, for several loci, more than 3 *canephora* alleles per loci were observed among the plants collected at a given site (i.e. D, C and F), suggesting the involvement of several *canephora* genotypes in the hybridization process. Furthermore, of the 42 hybrid plants, we identified 10 plants exhibiting at least one SSR locus presenting two *C. canephora* alleles; three pertain to the F<sub>1</sub> group of hybrids and the others to the post-F<sub>1</sub>.

### 6.5 Discussion

*C. arabica* was first introduced in New Caledonia by Marist Brothers in 1856. After several severe attacks by the pathogen *Hemileia vastatrix* (leaf rust), resistant diploid *Coffea* seeds were imported between 1911 and 1915 (Dussert 1911; Charles 1932), creating sympatric occurrence of both species in several areas. The occurrence of hybridization events between the coffee species, in fields that were forsaken in the 1940s, is clearly established in the present study. Examples of hybridization between pairs of closely related species are numerous (Arnold et al. 1987; Rieseberg et al. 1990; Martinsen

et al. 2001; Sweigart and Willis 2003; Payseur and Nachman 2005). However, few studies have involved polyploid species and their diploid progenitors such as *C. arabica* and *C. canephora*. The present sympatric zone is therefore of particular interest. In addition, the historical context of coffee introduction in New Caledonia provides information about the timescale of hybridization and rules out the possibility of ancient and recurring introgression between the species studied.

At all of the five prospected sites around le Col d'Amieu, hybrid plants were identified. The molecular characterization of hybrid plants indicated the occurrence of numerous and different hybridizations between *C. arabica* and *C. canephora* in the environmental conditions of the New Caledonian central mountains. The repeated hybrid formation suggests a significant overlap between the flowering periods of both species, although differences in flowering phenology between the two species have been reported (Baranski 1996). In particular, *C. arabica* has been reported to flower slightly later than *C. canephora*. The high humidity and rather cold temperature characterizing the New Caledonian central mountains is most probably very propitious for the development of hybrid plants. Moreover, that environment is extremely favourable to the expansion of coffee leaf rust due to *Hemileia vastatrix* (Bettencourt and Rodrigues Jr. 1988) and the leaf rust resistance of hybrid plants inherited from the *C. canephora* parents (Rodrigues Jr. et al. 1975) could represent an important selective advantage for the hybrid plants in comparison with the plants of *C. arabica*.

A substantial proportion of collected hybrids (i.e. 12 of 42) corresponded to first generation hybrids (i.e. F<sub>1</sub>), supporting the view that the hybridization between *C. arabica* and *C. canephora* is a recent and ongoing process in the New Caledonian central mountains. A noteworthy result emerged from the identification of various types of F<sub>1</sub> plants. Triploid as well as tetraploid F<sub>1</sub> plants were detected. Triploid F<sub>1</sub> plants are usually obtained when crossing *C. arabica* and *C. canephora* (Carvalho and Monaco 1968; Le Pierrès 1995). On the other hand, the identified tetraploid F<sub>1</sub> plants are likely to result from the involvement of unreduced gametes of *C. canephora*. Such gametes, which contain the full somatic chromosome number, are commonly considered as the consequence of "gametic non-reduction" or "meiotic nuclear restitution" during micro- and megasporogenesis. These unreduced gametes resulted either from an incomplete first meiotic (= first division restitution or FDR) division, or from an incomplete second meiotic

division (= second division restitution or SDR). The FDR 2n-gametes comprise non-sister chromatids of each homologous pair of chromosomes whereas in SDR 2n-gametes the sister chromatids are included (Hermesen 1984). In the present study, both processes seem to be concerned since of the 8 identified tetraploid F<sub>1</sub> plants, 3 seemed to involve FDR gametes (i.e. presence of loci exhibiting two canephora alleles) while 5 appeared to result from SDR gametes. Furthermore, both micro- and megasporogenesis appeared to be affected. Based on the characterization of hybrid plant cpDNA, of the 8 identified tetraploid F<sub>1</sub> plants, 6 result from crosses with *C. arabica* as female parent (i.e. Arabica type cpDNA) while 2 plants derive from crosses with *C. canephora* as female parent (i.e. Canephora type cpDNA). Several researchers have found that 2n pollen production is stimulated by environmental factors such as temperature, herbivory, wounding, water and nutrient stress (Ramsey and Schemske 1998). Temperature and especially variation in temperature have particularly large effects (Lewis 1943; McHale 1983). In Coffea, Lanaud (1979) reported the production of such unreduced gametes in *C. canephora* and *C. liberica*, two diploid species, under cold treatment. The New Caledonian mid-mountain conditions could therefore favour the production of 2n gametes and tetraploid hybrid plants.

The area of collection appeared as a hybrid swarm not only composed of parental species, and triploid or tetraploid F<sub>1</sub> plants, but also of post-F<sub>1</sub> hybrid plants. Previous studies of artificial hybrids in field conditions revealed that the fertility of hybrid trees is highly dependent on their ploidy. While tetraploid hybrids (i.e. resulting from the hybridization between *C. arabica* and auto-tetraploidized *C. canephora* diploid parents) appear reasonably fertile (Berthaud 1978; Le Pierrès 1995), the triploid hybrids are rather sterile. Nevertheless, viable gametes occur occasionally and triploid hybrids have been successfully backcrossed to *C. arabica* (Orozco 1976). Post-F<sub>1</sub> individuals were detected in high proportion (i.e. 71%) and at the 5 collection sites. All the 30 identified post-F<sub>1</sub> hybrid plants invariably showed the *C. arabica* DNA markers and diverged only for the number and type of canephora SSR-alleles or AFLP-bands suggesting an asymmetrical gene flow (Ostberg et al. 2004). Furthermore, all post-F<sub>1</sub> individuals exhibited DNA content close to that of *C. arabica*. Although allowing only a crude estimation, this observation indicated that most post-F<sub>1</sub> plants were nearly tetraploid, with about 44 chromosomes. Thus, the collected post-F<sub>1</sub>s in the present study are likely to result from either backcross events to *C. arabica* or from intercrosses between hybrid plants. The presence in *C. canephora* of a gametophytic self-incompatibility system (Berthaud 1986,

Lashermes et al. 1996b) could restrain the possibility of backcross to *C. canephora*. Furthermore, as observed in artificial hybrids (Herrera et al. 2002b, 2004), an evolution toward the tetraploid condition also seemed strongly favoured by a preferential transmission from hybrid plants of gametes carrying a diploid number (i.e. 22) of chromosomes. This process is most likely the consequence of both gametic selection and post-zygotic barriers, as reported elsewhere (Herrera et al. 2002b; Ehlenfeldt and Ortiz 1995, Carputo et al. 1999).

The ability to distinguish among backcross generations is rather limited because of the hybrid heterogeneity detected at the collecting sites in terms of ploidy level, the possibility of marker distortion, and the fact that the expected marker frequency distributions for different backcross generations can overlap (Boecklen and Howard 1997). Nevertheless, the amount of introgressed *C. canephora* markers in the collected post-F<sub>1</sub> appeared very high with 87% of plants exhibiting canephora alleles at a majority of analyzed loci. Only one plant (i.e. L8) displayed a low degree of introgression. It is therefore suggested that most if not all collected post-F<sub>1</sub> derived from only one or two post-F<sub>1</sub> generations. This hypothesis is consistent with the first occurrence of hybrid plants in the 1940s, when the coffee fields of the New Caledonian central mountains were forsaken, and the long time (i.e. 10-20 years) required before the first flowering of the coffee-tree in natural conditions (Berthaud 1986).

Gene flow from *C. canephora* to *C. arabica* therefore appears potentially very important in the New Caledonian central mountain conditions. In addition, the rather homogeneous frequency of introgression observed among the different analyzed loci in the post-F<sub>1</sub> plants suggests the absence of strong selection either in favour or against the introgressed canephora alleles. In relation to its outcrossing nature, *C. canephora* is considered to carry a high level of deleterious recessive alleles (Lashermes et al. 1994). The behaviour observed in the present study is likely to be a consequence of the polyploid condition. The buffering capacity of polyploids against the effects of deleterious recessive alleles has been already demonstrated in several plant species such as alfalfa and oilseed rape (Sharpe et al. 1995; Brouwer and Osborn 1999). On the other hand, one could expect a favourable selection of introgressed markers due a linkage disequilibrium (LD) between those markers and loci affecting the resistance to leaf rust. However, further studies on this issue are needed to draw final conclusions. In particular, a large number of plants and markers

should be considered (Comai et al. 2004) and it should be established whether the observed resistance to leaf rust is due to one or several loci.

A widely held view is that diploids and their polyploid derivatives are isolated by strong, post-zygotic genetic barriers (Futuyma 1998; Schluter 2001). However, this assumption is not entirely supported by the literature on wild species of mixed ploidy, and is probably an oversimplification (Husband 2004). From the present study, it appears that *C. arabica* and *C. canephora* could be considered as biological species, i.e. groups of interbreeding populations that are “genetically isolated” rather than “reproductively isolated” (Rieseberg and Carney 1998). The contribution of bridges via interspecific triploid or tetraploid hybrids on the establishment of the allotetraploid *C. arabica* species remains to be evaluated. In particular, a phase of stable coexistence between *C. arabica* and its diploid progenitor species could have occurred, supporting the suggestion (Husband 2000) of an important role for gradual, population processes governing the evolution of polyploidy in natural populations.

Altogether, the New Caledonian central mountains appear to be a highly favourable environment for hybrid formation and gene introgression between *C. canephora* and *C. arabica*. The New-Caledonian sympatric zone could be considered as a centre of neo-diversity for *C. arabica*. The detected interspecific hybrid plants represent a considerable genetic diversity that would justify preservation measures and appropriate conservation strategies. In addition, further analysis in times to come of those introgressed populations appears particularly interesting and offers an opportunity to explore important aspects of polyploidy evolution and speciation, such as the roles of ecological factors (e.g. leaf rust) and the consequences of interploidy introgression.

Chapitre rédigé sous la forme d'un article soumis à PLANT BREEDING

A new source of resistance against coffee leaf rust from  
New-Caledonian natural interspecific hybrids between  
*Coffea arabica* and *C. canephora*

L. MAHÉ, V.M.P. VÁRZEA, D. LE PIERRÈS and P. LASHERMES

## **7 A new source of resistance against coffee leaf rust from New-Caledonian natural interspecific hybrids between Coffea arabica and C. canephora**

### **7.1 Abstract**

The development of cultivars resistant to coffee leaf rust caused by the fungal pathogen *Hemileia vastatrix* is a priority in coffee breeding. However, only very few descendants of interspecific hybrids between *Coffea arabica* and diploid relative species have been used as resistance source. Identification of new sources of resistance appeared therefore particularly worthwhile. Hybrid plants derived from interspecific hybridization between *C. arabica* and *C. canephora* and found in neo-natural coffee tree populations of New Caledonia was therefore investigated. AFLP and microsatellites amplification were used to evaluate the genetic diversity of fourteen hybrid plants, and rust resistance was evaluated by inoculation with a panel of rust races representing a large variability in virulence. An important genetic diversity was characterized in hybrid plants originating from introgressions into *C. arabica* from various *C. canephora* progenitors. On the fourteen plants tested for leaf rust resistance, eight appeared resistant to all experimented races. Such plant material should represent a highly valuable resource for *C. arabica* breeding against coffee leaf rust.

### **7.2 Introduction**

Commercial coffee production relies mainly on two closely related species: *Coffea arabica* and *Coffea canephora*, accounting for approximately 70 and 30% of worldwide coffee production, respectively (International Coffee Organization, <http://www.ico.org>). *C. arabica* ( $2n = 4x = 44$ ), the only polyploid species in the *Coffea* genus, is characterized by very low genetic diversity (Anthony et al. 2002).

Coffee leaf rust caused by the pathogen fungal *Hemileia vastatrix* is regarded as one of the most serious diseases of the coffee-tree (Kushalapa and Eskes 1989, Van der Vossen 2005). Coffee genotypes are classified in resistance groups according to their interaction with physiologic races of the rust pathogen (Bettencourt and Rodrigues 1988, Várzea and Marques 2005). Plants, characterized by resistance to all the known rust races are classified

in group A while the traditional *Typica* and *Bourbon* cultivars that showed susceptibility to most races, are included in group E.

The search for natural coffee-trees showing resistance to leaf rust goes up from 1875. In particular, a tetraploid genotype known as the Timor Hybrid was found in a *C. arabica* field on the island of Timor in 1927 (Bettencourt 1973). Molecular analysis has established that the Timor Hybrid originated from one spontaneous interspecific cross between *C. canephora* and *C. arabica* (Lashermes et al. 2000). Progenies of the Timor Hybrid have been largely distributed worldwide and intensively used in coffee breeding programs as the main source of resistance to coffee leaf rust (Van der Vossen 2001). However, in the last few years, some improved commercial varieties derived from the Timor Hybrid have been reported to gradually lose their resistance to leaf rust in some countries due to the appearance of new virulent races (Várzea and Marques 2005). Identification and use of new sources of resistance to leaf rust would be therefore particularly worthwhile.

Recently, new hybrid plants deriving from interspecific crosses between *C. arabica* and *C. canephora* were described in several areas of New-Caledonian central mountains (Mahé et al. 2007). In the current study, we intend to evaluate the potential of thus plants, denominated hybrid from New Caledonia (HNC), for *C. arabica* improvement. The genetic diversity of fourteen representative HNC plants was established and compared to Timor Hybrid-derived genotypes. In addition their potential of resistance in regard to coffee leaf rust was determined.

## **7.3 Materials and methods**

### **7.3.1 Plant material**

Cuttings of fourteen HNC plants were collected at five sites around le Col d'Amieu in New Caledonia and were transferred to greenhouses in France and Portugal for further evaluation. Those plants included both first generation hybrids (i.e. F<sub>1</sub>) as well as advanced post-F<sub>1</sub> hybrid individuals corresponding to *Canephora*-introgressed arabica genotypes (Mahé et al. 2007). Three Timor hybrid-derived genotypes consisted of three accessions (T18130, T5296, T17931) that derived from the three main Timor hybrid progenies (i.e. 832-1, 832-2, and 1343) used in breeding programs were also included in the study for comparison. Eight *C. canephora* accessions from the IRD collection and two *C. arabica*



varieties namely *Bourbon* and *Typica*, cultivated in New Caledonia, were added as references.

### 7.3.2 Molecular marker analysis

Amplified Fragment Length Polymorphism (AFLP) and microsatellites (SSR) were applied for genetic diversity analysis. The AFLP protocol described by Lashermes et al. (2000) was followed and a total of twelve *EcoRI*+3/*MseI*+3 primer pairs were used in this analysis. In addition, fourteen SSR that revealed constantly distinct alleles between *C. arabica* and the cultivated accessions of *C. canephora* were selected for this analysis (Herrera et al. 2002). The PCR assay of specific SSR primer pairs was carried out as reported by Combes et al. (2000). Genetic distances between genotypes were estimated according to DICE coefficient. Cluster analysis by the unweighted pair group method using arithmetic averages (UPGMA) was performed with the TREECON (version 1.1) software (Van de Peer and De Wachter 1994). An introgression index reflecting the proportion of *C. canephora* genome introgressed was calculated for all HNC plants as reported by Mahé et al. (2007). This index corresponds to the number of observed AFLP *Canephora* marker alleles in a given hybrid plant divided by the estimated number of AFLP *Canephora* marker alleles detected in a theoretical interspecific hybrid *C. arabica* x *C. canephora* using the same set of primer combinations.

### 7.3.3 Rust inoculation test

Evaluation for leaf rust resistance of fourteen plants issued from interspecific crosses between *C. arabica* and *C. canephora* was carried out at Coffee Rusts Research Center (CIFC) in Oeiras, Portugal. The inoculation tests and the assessment of reaction types were carried out by a method developed by D'Oliveira, (1954-57). Each inoculation test was repeated three times. Except for the plant L1, four different rust races from the worldwide collection maintained at CIFC were used: isolates CIFC 1427 (Race II), CIFC 2191 (Race XXXVII), CIFC 5a (incompletely characterised Race), and CIFC 1321 (Race XXIX). Those isolates were selected for representing a large variability in virulence (i.e. virulence genes v5-to-v9) and included the new virulent rust isolates (i.e. isolates 5a, 2191 and 1321) that overcome the resistance of Timor Hybrid-derived genotypes.

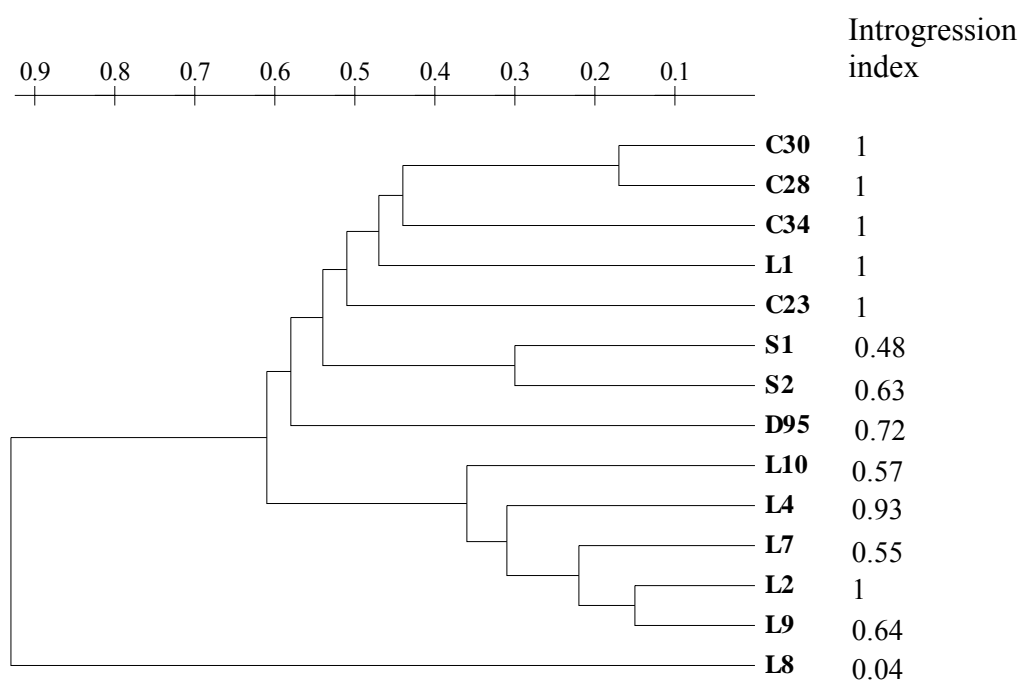


Figure 17. Dendrogram based on AFLP marker (Dice coefficient) constructed for the fourteen hybrids plants sampled in New Caledonia. The canephora AFLP introgression index is indicated on the right side of the dendrogram.

## **7.4 Results and discussion**

Twelve AFLP primer combinations were used. A total of 250 AFLP fragments were scored in the group of HNC plants. All the AFLP fragments (i.e. 121) observed in *C. arabica* accessions were identified within the group of HNC plants. On the other hand, 129 AFLP fragments not detected in any of the accessions of *C. arabica* were observed. Those markers were therefore considered further as *Canephora* markers. Although the number of *C. canephora* clones included in this study was limited, 115 of these 129 marker bands were also observed in at least one of the *Canephora* samples analysed, confirming their *Canephora* origin.

To examine the genetic diversity represented by the HNCs, a dendrogram tree was drawn which computed all the AFLP polymorphic bands detected in HNC plants, which were almost exclusively *Canephora* markers (Figure 17). The dendrogram was widely spread revealing an important diversity. Thus establishing that the HNC constitute a considerable source of genetic diversity.

Furthermore, the total number of *Canephora* markers varied from 2 to 60 among the HNC plants. Thus, the introgressed fraction of *C. canephora* genome as estimated through the AFLP introgression index appeared highly variable (Figure 17). While 6 plants are likely to be F<sub>1</sub> hybrids (i.e. introgression index of 1), the plant L8 displayed a low introgression rate (i.e. 4%). This is of special importance when using this material for *C. arabica* breeding in order to reduce the number of backcross generations required to restore the genetic background of the recipient cultivar and there by ensure good quality of the improved variety. For comparison, using similar approach, the introgressed markers identified in a set of Timor Hybrid-derived genotypes were estimated to represent from 9% to 26% of the *C. canephora* genome (Lashermes et al. 2000).

To compare further the HNCs to the Timor Hybrid-derived genotypes, the allelic diversity at fourteen SSR loci was analysed. While all the SSR loci exhibited *Canephora* introgressed alleles (i.e. alleles never observed in *C. arabica* accessions and exhibited by one or several *C. canephora* accessions) among the HNCs, only five SSR loci appeared introgressed in Timor Hybrid-derived genotypes. For those, the overall allelic diversity related to *C. canephora* introgression observed in the Timor hybrid-derived genotypes appeared to be very low (6 alleles) compared to that detected within the HNC plants (20

Table 6. Reaction type (R = resistance and S = Susceptible) of HNC plants inoculated with the race II and different rust isolates that overcome resistance factors introgressed in *C. arabica* from *C. canephora* through the Timor Hybrid.

Plants	Rust isolates				Putative coffee physiological Group *
	1427 Race II (v5)	2191 Race XXXVII (v2, 5, 6, 7, 9)	5a Undefined Race (v2, 5, 7, 9, ?)	1321 Race XXIX (v5, 6, 7, 8, 9)	
L 1	R	nd	nd	R	A
L 2	R	R	R	R	A
L 4	R	R	S	R	Not yet defined
L 7	R	R	R	R	A
L 8	R	R	R	R	A
L 9	R	R	R	R	A
L 10	R	R	R	R	A
C 23	R	R	R	R	A
C 28	R	S	R	R	Not yet defined
C 30	S	S	R	S	R
C 34	R	R	R	R	A
D 95	R	S	R	R	Not yet defined
S 1	R	S	S	S	E
S 2	R	R	R	R	A

nd: not determined

\* As defined in Várzea and Marques 2005

alleles). Only two *Canephora* alleles were in common between Timor Hybrid-derived genotypes and HNC plants. This result is consistent with the difference in origin of the two types of genetic material: the Timor Hybrid-derived plants originated from a unique interspecific hybrid (Lashermes et al. 2000) while the HNCs derived from numerous crosses involving contrasted *C. canephora* parental genotypes.

With the exception of plant C 30, all HNC plants appeared resistant to the rust race II that is the most frequent worldwide (Kushalapa and Eskes 1989). Interestingly, it is possible to find HNC plants like, L2, L7, L8, L9, L10, C23, C34 and S2 that showed resistance to all experimented races including the new virulent races overcoming the resistance of Timor Hybrid-derived genotypes. Those plants most probably belong to the physiological group A (Table 6). Furthermore, HNC plants (L4, C28, C30 and D95) exhibited heterogeneous responses suggesting the involvement of several resistance factors inherited from *C. canephora*. Although requiring further studies, these plants are likely to correspond to new physiological groups (Várzea and Marques 2005).

In conclusion both AFLP and SSR were very effective at detecting genetic variation. The molecular data points toward elevated level of genetic diversity of *C. canephora* progenitors at the origin of HNCs. Comparison with Timor Hybrid-derived genotypes illustrates the large and new diversity present in HNC plants. Finally rust inoculation test clearly established the resistance potential of those HNCs. Although only a representative set of HNCs was studied, HNC plants appeared very promising as new source of resistance against leaf rust and would justified extended collection and preservation measures. Individuals showing a low level of introgression could be directly used in breeding programs and significantly contribute to the development of improved Arabica cultivars exhibiting a valuable resistance to rust.

## **CHAPITRE III**

### **Cartographie génétique du locus de résistance SH3**

Chapitre rédigé sous la forme d'un article à soumettre

Development of sequence characterized DNA markers linked to leaf rust (*Hemileia vastatrix*) resistance in coffee (*Coffea arabica* L.)

## **8 Development of sequence characterized DNA markers linked to leaf rust (*Hemileia vastatrix*) resistance in coffee (*Coffea arabica* L.)**

### **8.1 Abstract**

Coffee leaf rust due to *Hemileia vastatrix* is one of the most serious diseases in Arabica coffee (*Coffea arabica*). A resistance gene (i.e.  $S_{H3}$ ) has been transferred from *C. liberica* into *C. arabica*. The present work aimed at developing sequence characterised genetic markers for leaf rust resistance. Linkage between markers and leaf rust resistance was tested by analysing two segregating progenies, one  $F_2$  population of 101 individuals and one backcross ( $BC_2$ ) population of 43 individuals, derived from a cross between a susceptible and a  $S_{H3}$ -introgressed resistant genotype. A total of ten sequence characterised genetic markers closely associated with the  $S_{H3}$  leaf rust resistance gene were generated. These included simple sequence repeats (SSR) markers, sequence-characterised amplified regions (SCAR) markers resulting from the conversion of amplified fragment length polymorphism (AFLP) markers previously identified and SCAR markers derived from end-sequences of bacterial artificial chromosome (BAC) clones. Those BAC clones were identified by screening of *C. arabica* genomic BAC library using a cloned AFLP-marker as probe. The markers we developed are easy and inexpensive to run, requiring one PCR step followed by gel separation. While three markers were linked in repulsion with the  $S_{H3}$  gene, seven markers were clustered around the  $S_{H3}$  gene. Notably, two markers appeared to co-segregate perfectly with the  $S_{H3}$  gene in the two plant populations analysed. These markers are suitable for marker-assisted selection for leaf rust resistance and to facilitate pyramiding of the  $S_{H3}$  gene with other leaf rust resistance genes.

### **8.2 Introduction**

Coffee leaf rust (CLR) due to the biotrophic fungus *Hemileia vastatrix* is one of the most serious diseases in Arabica coffee (*Coffea arabica*) and causes annual economic damage estimated at more than US\$ 1 billion (van der Vossen 2005). Although after the first reported epidemics on cultivated arabica coffee in Sri Lanka, India and Indonesia, CLR was confined to Africa and Asia for more than a century, it is now found in all coffee producing countries in the world except Hawaii and Australia. Timely application of



fungicides can provide adequate control, but is usually beyond the financial means of small farmers who are the main coffee producers in most countries (Kushalappa and Eskes 1989). Given the economic context, minimizing the use of chemical inputs for management of the disease, and developing and cultivating durable host-resistant cultivars appears to be the most effective and viable option.

*C. arabica* L. ( $2n = 4x = 44$ ), the only polyploid species in the *Coffea* genus, is an allotetraploid containing two diploid subgenomes, namely  $C^a$  and  $E^a$ , which originated from two different diploid species ( $2n = 2x = 22$ ), *C. canephora* and *C. eugenioides*, respectively (Lashermes *et al.* 1999). In spite of the low divergence between the two constitutive subgenomes, *C. arabica* displays a diploid-like meiotic behaviour and conforms to disomic inheritance (Lashermes *et al.* 2000a). Furthermore, while *C. arabica* is characterized by low genetic diversity, considerable variability has been reported among diploid coffee species (Berthaud and Charrier 1988). The transfer of desired characters, and in particular disease resistance from related diploid species into cultivars of *C. arabica*, is thus an ongoing priority in coffee breeding (van der Vossen 2001). Host resistance to CLR is conditioned primarily by a number of dominant resistance factors, and coffee genotypes are classified in groups based on their interaction with physiologic races of the rust pathogen. Resistance to coffee leaf rust appears to be determined by at least nine resistance genes, referred to as  $S_{H1}$  to  $S_{H9}$ , either singly or in combination; the corresponding virulence's are referred to as  $V_1$  to  $V_9$  (Bettencourt and Rodrigues 1988). Of these resistance factors,  $S_{H1}$ ,  $S_{H2}$ ,  $S_{H4}$  and  $S_{H5}$  have been identified in *C. arabica*. The other genes,  $S_{H6}$ ,  $S_{H7}$ ,  $S_{H8}$  and  $S_{H9}$ , have been introgressed from the diploid species *C. canephora*, and  $S_{H3}$  from another diploid species, *C. liberica* (Wagner and Bettencourt 1965; Bettencourt and Rodrigues 1988; Prakash *et al.* 2004).

Efforts to obtain durable resistance to coffee leaf rust had a long history of initial successes followed by disappointments because of the appearance of new virulent races of rust fungus (Bettencourt and Rodrigues 1988; Prakash *et al.* 2005). To date as many as 45 physiological races are known to infect different coffee genotypes in different coffee growing countries (Várzea and Marques 2005). The resistance genes identified in *C. arabica*, used either singly or in combination, have not provided durable resistance to most races of rust fungus. In contrast, the  $S_{H3}$  gene from *C. liberica* and certain genes from *C. canephora* have provided long-lived protection under field conditions (Eskes *et al.* 1989;

Srinivasan and Narasimhaswamy 1975). Consequently, the combined use of resistance genes introgressed from diploid coffee species -either by accumulating them in one genotype (pyramid varieties) or using them in composite varieties- is expected to provide durable resistance (Kushalappa and Eskes 1989; Moreno and Alvarado 2000; Eskes 2005). However, combined introgression of various resistance genes into outstanding cultivars using traditional plant breeding methods can be cumbersome as the presence of new genes introduced into the common background is masked by genes that are already present. Moreover, there are clear limitations due to the length of time needed for coffee tree generation, the high cost of field trials, and the lack of accuracy in the current strategy used to restore the genetic background of the recipient cultivar and thereby ensure the quality of the improved variety. Availability of molecular markers linked to S<sub>H</sub> genes and other genes that influence resistance to coffee leaf rust would therefore be extremely useful (Lashermes et al. 2000b). In a previous paper (Prakash et al. 2004), we reported the detection of molecular markers closely linked to the S<sub>H3</sub> leaf rust resistance gene introgressed into *C. arabica* from *C. liberica*. We identified a total of 21 amplified fragment length polymorphism (AFLP) markers distributed within a distance of 6.3 cM on the *C. liberica* introgressed fragment carrying the S<sub>H3</sub> gene. Nevertheless, we considered the usefulness of these AFLP markers in molecular-assisted selection (MAS) to be rather limited.

The objectives of the present study were (1) to convert the identified AFLP markers into sequence-characterized amplified region (SCAR) markers suitable for MAS, (2) to develop new markers linked to the S<sub>H3</sub> gene using a genomic BAC (bacterial artificial chromosome) library of *C. arabica* (Noir et al. 2004) and marker resources from a genetic linkage map of *C. canephora* involving microsatellite (i.e. simple sequence repeats, SSR) markers (Lashermes et al. 2001), and 3) to test the robustness and behaviour of the S<sub>H3</sub>-linked markers in two segregating plant populations.

## **8.3 Materials and methods**

### **8.3.1 Plant material**

The plant material analysed in the present study consisted of two *C. arabica* segregating populations involving the leaf rust susceptible line Matari (resistance group F) and the S.288 line (resistance group G): one F<sub>2</sub> population (Matari x S.288) of 101 individuals and one backcross (BC<sub>2</sub>) population of 43 individuals, the Matari line being used as recurrent

parent. The S.288 line is derived from a selfed offspring of S.26, a natural hybrid between *C. arabica* and *C. liberica* (Vishveshwara 1974), and carries the CLR S<sub>H3</sub> resistance gene that has been introgressed from *C. liberica* (Prakash et al. 2002, 2004). In addition to the above populations, the two parental lines were also included in the analysis.

### 8.3.2 Evaluation of rust resistance

Evaluation of the F<sub>2</sub> population for leaf rust resistance has been already reported (Prakash et al. 2004). Evaluation of the backcross population was carried out at CIFC (Oeiras, Portugal) in the same way. As detailed by Bettencourt and Rodrigues (1988), rust resistance patterns were assessed based on artificial inoculations in greenhouse/controlled conditions using the rust races II and VIII that enable segregation analysis of the S<sub>H3</sub> resistance gene. The rust uredospores (1 mg/pair of leaves) were inoculated on the lower side of the fully expanded tender leaves and were then gently spread with a camel-hair brush. The inoculated leaf surfaces were sprayed with distilled water and the plants were kept in moist chambers (18-24°C) or wrapped in wet plastic bags for 48 h under dark/dim light conditions. The reaction type was evaluated 25-35 days after inoculation using a descriptive scale (D'Oliveira 1957; Eskes and Toma-Braghini 1982).

### 8.3.3 Development of SCARs from AFLP markers

AFLP markers whose estimated size was over 150 bp and which were well separated from other visible bands were selected for cloning. By careful matching of the gels and the films the bands were excised, the DNA diffused into water and 2 µl of the eluted DNA was used for PCR analysis with matching unlabelled selective AFLP primers. Amplification was performed in a 25 µl mixture containing 2 µl of the DNA solution, 1X buffer, 2.0 mM of MgCl<sub>2</sub>, 100 µM of dNTPs, 0.4 µM of each of the primers and 0.5 U of *Taq* DNA polymerase. The PCR programme consisted of an initial denaturation step at 95°C for 5 minutes, followed by 35 denaturation cycles at 94°C for 45 sec, primer annealing at 50°C for 45 sec, elongation at 72°C for 45 sec, and a final extension step at 72°C for 10 min. Amplification products were viewed in 2% agarose gel. Fresh PCR products consisting of single bands were cloned using the TOPO TA Cloning kit with pCR 2.1-TOPO vector and chemically competent cells (Invitrogen, Life Technologies) following the manufacturer's instructions. Four positive clones with the right sized inserts were selected per cloning

reaction and plasmid DNA extracted using the QIAprep spin kit (QIAGEN Sciences, Maryland, USA) according to the manufacturer's instructions. The plasmid DNA was sequenced by Genome Express (France) and replicate sequences of the same band were aligned using the CLUSTAL W 1.82 program (European Bioinformatics Institute). Sequence specific primers were designed using the Primer3 program (Whitehead Institute, USA) targeting sizes between 18 and 22 bp and an optimum annealing temperature of 55°C, so that they could all be analysed under the same conditions.

#### 8.3.4 BAC library analysis

A BAC library constructed (Noir *et al.*, 2004) from the genomic DNA of the arabica cultivar IAPAR 59 was used. High-density colony filters were prepared using a Genetix Q-Bot. BAC clones were double-spotted onto Hybond-N<sup>+</sup> filters (Amersham-Pharmacia) in a 4 x 4 pattern as described by Tomkins *et al.* (1999). In total, 27 648 clones from the BAC library were spotted onto a 22.5 x 22.5 cm filter representing approximately 7 basic coffee genome equivalents. The labelling of probes and filter hybridization conditions were as described in Noir *et al.* 2004. The detected BAC clones were grown for 48 h at 37°C in 1.5 ml 2YT medium containing chloramphenicol (12.5 µg/ml). The BAC DNA was extracted using an alkaline lysis procedure (Sambrook *et al.*, 1989) and extremities sequenced from one or the two insert sides using primers pIBf (5' GGATGTGCTGCAAGGCGATTAAGTTGG 3') and pIBr (5' CTCGTATGTTGTGTGGAATTGTGAGC 3'). Sequence specific primers were designed as indicated above.

#### 8.3.5 Analysis of SCAR and SSR markers

Genomic DNA was isolated from lyophilized leaves by means of a nuclei isolation step as described by Agwanda *et al.* (1997). Amplification conditions using specific primer pairs (either SCAR or SSR), radioactive labeling and polyacrylamide gel electrophoresis were as reported elsewhere (Combes *et al.* 2000). The markers were revealed either on Kodak Biomax X-ray films or on Amersham storage Phosphor screen and screened after 24 hr with a Typhoon scanner (9700 series, Amersham Biosciences) to obtain digital images. Single-stranded DNA conformation polymorphism (SSCP) analysis was also performed. Briefly, 8 µl of PCR products were added to 2 µl of Urea (5M), 10 µl of a solution containing 95% formamide, 10 mM NaOH, 0.05% of xylene cyanol, and 0.05% of

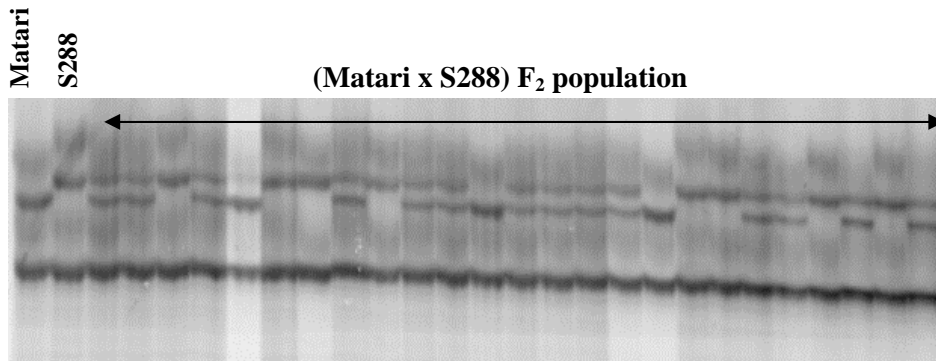


Figure 18. Example of marker (i.e. Sat244) associated with the  $S_{H3}$  leaf rust resistance gene segregating within the  $F_2$  (Matari x S.288) population.

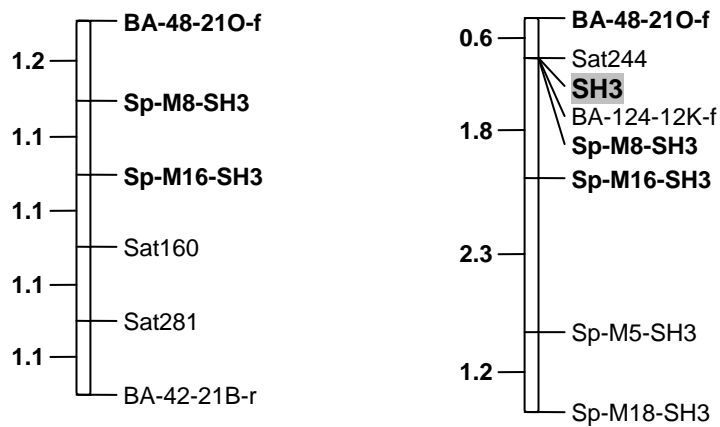


Figure 19. Linkage map based on the  $F_2$  (Matari x S.288) population analysis of the chromosome segments either in coupling (right) or repulsion (left) surrounding the  $S_{H3}$  gene (highlighted). Molecular marker names are on the right, while the estimated map distances in cM are shown on the left. SCAR markers showing codominant polymorphism are in bold.

bromophenol blue and 5 µl of TBE 0.5x. The samples were heat-denatured for 6 min in 94 °C and quickly cooled on ice for 2 min. Then 15 µl of the sample was loaded onto a 0.75 mm 16 cm x 18 cm non-denaturing acrylamide gel (8 ml acrylamide-bisacrylamide (40% w/v) 37.5:1, 4 ml Glycerol 50%, 2 ml TBE 10X, 26 ml of water, 150 µl ammonium persulphate 10% and 15µl TEMED. Electrophoresis was run in 0.5x TBE buffer at cool temperatures (10 degrees) and 450V/cm. After the gel was removed from the apparatus, and stained with Ethidium Bromide for 20 minutes.

Linkage analysis and map construction were performed using the computer program MAPMAKER version 3.0b (Lander et al 1987). The most-likely order of the markers was obtained using exhaustive multipoint analyses. Recombination frequencies were converted into map distances or centiMorgan (cM) values using the Kosambi function (Kosambi 1944).

## **8.4 Results**

### **8.4.1 Conversion of AFLP markers into SCARs**

Attempts were made to clone 15 of the 21 previously identified S<sub>H3</sub>-linked AFLP markers, sequences and specific primer pairs were identified for ten of these AFLP markers. However, only four revealed polymorphism between the two parental lines and within the F<sub>2</sub> (Matari x S.288) population and could consequently be validated. These SCAR, named Sp-M5-SH3, Sp-M8-SH3, Sp-M16-SH3, Sp-M18-SH3, are described in Table 7 and correspond to the AFLP markers M5, M8, M16 and M18, respectively. Depending on the SCAR considered, 2 to 3 different amplified products were observed but in all cases a marker band was found to be present in the resistant parental line (S.288) and absent in the susceptible line (Matari) (Figure 18). Segregation of these markers was detected among the F<sub>2</sub> population as the original AFLP markers. In addition for two SCAR, Sp-M8-SH3 and Sp-M16-SH3, an additional marker band, absent in the resistant parental line (S.288) and present in the susceptible line (Matari), and segregating among the F<sub>2</sub> population was also observed.

Table 7. Sequence-characterised DNA markers linked, either in coupling or in repulsion, to the  $S_H3$  leaf rust resistance gene

Marker	Primer sequences (5'>3')	Linkage (Polymorphism)	PCR Anneal. temp. (°C)	Electrophoresis conditions *
<b>Sp-M5-SH3</b>	For: TTCACGATCCAAGAAGCA Rev: AGCATGCATTGTAGAAAAA	Coupling (Dominant)	50	D
<b>Sp-M8-SH3</b>	For: GAATTCAGCGACGATTG Rev: GATTTGGTGGAAAGGGAGC	Coupling (Co-Dominant)	55	N
<b>Sp-M16-SH3</b>	For: TTAAGTGGAAACTTGGCTTG Rev: ATCTAGCTTTGGAACATCGT	Coupling (Co-Dominant)	50	D
<b>Sp-M18-SH3</b>	For: CTATTTGGTGTGGGAAGTAAC Rev: CTACATCCACGGAGAGAAAC	Coupling (Dominant)	55	D
<b>Sat160</b>	For: TGCTTAGGCACTTGATATAGGA Rev: CACGTGCAAGTCACATACTTTA	Repulsion (Dominant)	60	D
<b>Sat281</b>	For: TCTTCGTCTTTGCTATTGGT Rev: TATTAACGTCCATCCACACA	Repulsion (Dominant)	56	D
<b>Sat244</b>	For: GCATGTGCTTTTTGATGTCGT Rev: GCATACTAAGGAAATTATCTGACTGCT	Coupling (Dominant)	56	D
<b>BA-42-21B-r</b>	For: CACACACAGCCTAAGCATCAA Rev: GGATTGACTCGACTCACCAA	Repulsion (Dominant)	60	D
<b>BA-48-210-f</b>	For: ACAGTGAATTCCCCAAGCAC Rev: ACTTGGCAGGCGTAATTGAA	Coupling (Co-Dominant)	60	D
<b>BA-124-12K-f</b>	For: TGATTTCGCTTGTTGTCGAG Rev: TGCAGATTGATGGCACGTTA	Coupling (Dominant)	60	D

\* Denaturing vs. non-denaturing acrylamide gel electrophoresis

#### 8.4.2 Development of BAC-derived markers

A genomic *C. arabica* BAC library was screened by high-density filter hybridisation using the AFLP-marker clones M5, M8, and M16 as probes. While the M5 and M16 clones hybridised a high number of BAC clones and contained most likely repetitive sequences, only two BAC clones were positive for M8. Subsequently, BAC ends from these two M8-positive BAC clones were used as probes to develop and extend a BAC contig on both sides. A total of 35 BAC clones were isolated and specific primer pairs were designed for 18 different BAC-end sequences. The ability of these sequence-specific primers to reveal polymorphism was tested. Three BAC-end derived SCAR markers showing clear and simple size polymorphism between the two parental lines were identified and are listed in Table 7. The markers were named according to the BAC clones for which they were derived, BA-48-21O-f, BA-124-12K-f and BA-42-21B-r, respectively.

#### 8.4.3 Identification of SSR markers

Preliminary analysis (results not shown) indicated that the markers Sp-M5-SH3, Sp-M8-SH3, and Sp-M18-SH3, are located on linkage group 1 (i.e. chromosome 1) of the *C. canephora* genetic map derived from the clone IF200 (Lashermes et al. 2001). SSR markers mapped in this region were therefore tested for polymorphism between the two parental lines, Matari and S.288, and within the  $F_2$  (Matari x S.288) population. Three SSR (Table 7) revealed polymorphism due to the presence of an additional allele either in Matari (Sat160, Sat281) or in S.288 (Sat244).

#### 8.4.4 Genetic Mapping of the $S_{H3}$ region

Linkage analysis was performed using the  $F_2$  (Matari x S.288) population (Figure 19). The ten sequence characterized DNA markers (i.e. SCAR and SSR) were tested on 101 individuals. Seven markers fitted the 3:1 inheritance mode expected for a dominant marker while three markers appeared to be co-dominant (1:2:1 ratio). The analysis resulted in two small linkage groups (i.e. 5.6 and 5.9 cM) comprising markers tightly linked to each other (LOD score > 10). The markers Sat244, BA-124-12K-f and Sp-M8-SH3 co-segregated perfectly with the  $S_{H3}$  leaf rust resistance gene. The two linkage groups grouped all alleles/markers inherited from either S.288 or Matari and appeared to be connected by the



Table 8. Survey of sequence-characterised DNA markers linked to the  $S_{H3}$  gene in 43 individuals (BC2) derived from a cross between the line S.288 carrying the  $S_{H3}$  rust resistance gene (group G) and the susceptible line Matari (group F) used as recurrent parent. + indicates the presence of the marker bands; – indicates the absence.

Plant code	Resistance Group	BA-48-21O-f	Sat244	BA-124-12K-f	Sp-M16- $S_{H3}$	Sp-M5- $S_{H3}$	Sp-M18- $S_{H3}$
2	F	-	-	-	-	-	-
3	F	-	-	-	-	-	-
6	F	-	-	-	-	-	-
12	F	-	-	-	-	-	-
13	F	-	-	-	-	-	-
14	F	-	-	-	-	-	-
19	F	-	-	-	-	-	-
21	F	+	-	-	-	-	-
24	F	-	-	-	-	-	-
34	F	-	-	-	-	-	-
35	F	-	-	-	-	-	-
36	F	-	-	-	-	-	-
37	F	-	-	-	-	-	-
46	F	-	-	-	-	-	-
47	F	-	-	-	-	-	-
54	F	-	-	-	-	-	-
60	F	-	-	-	-	-	-
63	F	-	-	-	-	-	-
64	F	-	-	-	-	-	-
66	F	-	-	-	-	-	-
4	G	+	+	+	+	-	-
7	G	+	+	+	+	-	-
8	G	+	+	+	-	-	-
10	G	+	+	+	+	-	-
11	G	+	+	+	+	-	-
18	G	+	+	+	+	-	-
22	G	+	+	+	+	-	-
25	G	+	+	+	+	-	-
27	G	+	+	+	+	-	-
32	G	+	+	+	+	-	-
38	G	+	+	+	+	-	-
40	G	+	+	+	+	-	-
41	G	+	+	+	+	-	-
42	G	+	+	+	+	-	-
43	G	+	+	+	+	-	-
45	G	+	+	+	+	-	-
48	G	+	+	+	+	-	-
51	G	+	+	+	+	-	-
52	G	+	+	+	+	-	-
55	G	+	+	+	+	-	-
58	G	+	+	+	+	-	-
67	G	+	+	+	+	-	-
68	G	+	+	+	+	-	-

three co-dominant markers (BA-48-21O-f, Sp-M8-SH3, Sp-M16-SH3). The two linkage groups thus appeared to be related to the S<sub>H3</sub> region, one being the introgressed fragment carrying the S<sub>H3</sub> gene (i.e. markers in coupling) while the other represented the corresponding fragment inherited from the non-introgressed line (i.e. markers in repulsion).

#### 8.4.5 Confirmation of the leaf rust-linked markers in an advanced segregating population

To evaluate their potential usefulness as genetic markers associated with the S<sub>H3</sub> gene, nine of the ten sequence characterized DNA markers (marker Sp-M8-SH3 was not tested) were further evaluated in a backcross population (BC<sub>2</sub>) between S.288 and Matari (recurrent parent) comprising 43 individuals (Table 8). All the markers/alleles in repulsion were always present and thus did not show any segregation. Furthermore, the markers in coupling, Sp-M5-SH3 and Sp-M18-SH3, were constantly absent. Only the markers closest to the S<sub>H3</sub> gene, BA-48-21O-f, Sat244, BA-124-12K-f and Sp-M16-SH3 showed the expected segregation ratios. Moreover, the S<sub>H3</sub> gene appeared to be present in 23 individuals and absent in 20 individuals, a segregation pattern that fits the 1:1 ratio expected for a single dominant gene. The markers Sat244 and BA-124-12K-f co-segregated perfectly with the S<sub>H3</sub> leaf rust resistance gene while the markers BA-48-21O and Sp-M16-SH3 showed only one discrepancy each. All together, the observed segregation patterns were consistent with the marker order and the S<sub>H3</sub> localisation deduced from the analysis of the F<sub>2</sub> (Matari x S.288) population.

### 8.5 Discussion

In the present study, a total of ten sequence characterised genetic markers associated with the S<sub>H3</sub> leaf rust resistance gene were generated. These included SSR markers as well as SCAR markers. In particular, AFLP markers were successfully converted to SCAR markers. The conversion procedure we used thus appears to be quite reliable even though not all AFLP markers were successfully cloned and validated. Afterwards, the cloned sequence corresponding to a SCAR marker that appeared in the vicinity of the S<sub>H3</sub> gene was used to screen a *C. arabica* BAC library and develop a BAC contig encompassing the S<sub>H3</sub> locus. This BAC contig was not only useful to generate additional SCAR markers but

also represents a significant step towards a map-based cloning approach (Peters et al. 2003).

The markers we developed are inexpensive to run, requiring one PCR step followed by gel separation, and they are all distributed within the S<sub>H3</sub> region. While three markers were linked in repulsion with the S<sub>H3</sub> gene, seven markers were clustered around the S<sub>H3</sub> gene. Notably, two markers appeared to co-segregate perfectly with the S<sub>H3</sub> gene in the two plant populations analysed and should be particularly appropriate for coffee MAS (Lashermes et al. 2000b). As anticipated, the markers in repulsion were absent in the advanced backcross population. The two more distant markers on one side of the S<sub>H3</sub> gene were also absent in that population suggesting that crossover events occurred within the introgressed fragment carrying the S<sub>H3</sub> gene during development of the backcross population.

Both co-dominant and dominant markers were identified. The interest of both marker types is rather similar. Co-dominant markers enabled the heterozygous to be distinguished from the homozygous plants for the S<sub>H3</sub> gene. However, adequate combinations of dominant markers in coupling and in repulsion could provide the same information. Furthermore, both marker types produce bands in both resistance and susceptible genotypes and failed-PCR reactions could be easy to identify.

The leaf rust resistance S<sub>H3</sub> gene was targeted for marker development because of its simple genetic inheritance (Prakash et al 2004) and its usefulness for coffee breeding programmes. Even though races of *H. vastatrix* that break down the host resistance conferred by the S<sub>H3</sub> gene originated in India, they nevertheless displayed reduced aggressiveness, and the cultivar harbouring the S<sub>H3</sub> gene has demonstrated agronomically acceptable durable resistance in field conditions for the last 20 years (Prakash et al 2005). In field conditions, a fitness penalty (Leach et al. 2001) appears to be applied to the pathogen to overcome this particular resistance gene (Eskes 1989; Prakash et al 2005). Furthermore, similarity may exist between S<sub>H</sub> resistance genes and the molecular information generated concerning S<sub>H3</sub> provides the insight needed to study other S<sub>H</sub> genes (Michelmore 2000).

Marker-assisted selection offers many advantages in plant breeding, especially when trying to breed for durable resistance against plant pathogens by pyramiding resistance genes into a common background (Melchinger 1990; Youg 1999). Subsequent progeny testing to

confirm the presence of desirable genes can be a laborious process especially in the absence of discriminatory pathotypes of the pathogen concerned. The introgression of genes can easily be monitored using molecular markers without screening with the pathogen and irrespective of environmental conditions.

In conclusion, the sequence characterized DNA markers developed in this study appear to be very closely linked to the S<sub>H3</sub> leaf rust resistance gene. As such, these markers are not only useful for MAS, but will also be extremely useful for fine-mapping studies and ultimately for cloning the S<sub>H3</sub> gene (s).

## **CHAPITRE IV**

### **Génomique comparative**

Chapitre rédigé sous la forme d'un article soumis à PLANT MOLECULAR BIOLOGY

Comparison between a coffee single copy chromosomal region and *Arabidopsis* duplicated counterparts evidenced high level synteny between the coffee genome and the ancestral *Arabidopsis* genome

LAETITIA MAHE, MARIE-CHRISTINE COMBES AND PHILIPPE LASHERMES

## **9 Comparison between a coffee single copy chromosomal region and Arabidopsis duplicated counterparts evidenced high level synteny between the coffee genome and the ancestral Arabidopsis genome**

### **9.1 Abstract**

The *Arabidopsis thaliana* genome sequence provides a catalogue of reference genes that can be used for comparative analysis of other species thereby facilitating map-based cloning in economically important crops. We made use of a coffee bacterial artificial chromosome (BAC) contig linked to the S<sub>H3</sub> leaf rust resistance gene to assess microsynteny between coffee (*Coffea arabica* L.) and *Arabidopsis*. Microsynteny was revealed and the matching counterparts to *C. arabica* contigs were seen to be scattered throughout four different syntenic segments of *Arabidopsis* on chromosomes (Ath) I, III, IV and V. Coffee BAC filter hybridizations were performed using coffee putative conserved orthologous sequences (COS) to *Arabidopsis* predicted genes located on the different *Arabidopsis* syntenic regions. The coffee BAC contig related to the S<sub>H3</sub> region was successfully consolidated and later on validated by fingerprinting. Furthermore, the anchoring markers appeared in same order on the coffee BAC contigs and in all *Arabidopsis* segments with the exception of a single inversion on AtIII and AtIV *Arabidopsis* segments. However, the S<sub>H3</sub> coffee region appears to be closer to the ancestral genome segment (before the divergence of *Arabidopsis* and coffee) than any of the duplicated counterparts in the present-day *Arabidopsis* genome. The genome duplication events at the origin of its *Arabidopsis* counterparts occurred most probably after the separation (i.e. 94 million years ago) of Euasterid (Coffee) and Eurosida (*Arabidopsis*).

### **9.2 Introduction**

During the last decade, comparative genomic approaches have been developed in many crop species, notably in cereals, crucifers and solanaceae (Devos 2000; Lagercrantz and Lydiate 1996; Moore et al. 2005). The comparison of genomic sequences across species enables better understanding of the processes that drive the evolution of genomes. To make such comparisons, homologous regions need to be identified within or between genomes to enable a closer look at the different types of rearrangements (e.g., inversions,

translocations and transpositions), gene duplications and gene loss. Furthermore, given the high degree of genome colinearity at both the broader genetic level and the gene level, comparative genome mapping can be an efficient tool for transferring information and resources from well-studied genomes, such as those of *Arabidopsis* and rice, to related plants. It can also facilitate high-resolution linkage mapping and map-based cloning experiments in many plant species. To date, most comparative genomic and genetic analyses of dicotyledonous plants have used information from the genome sequence of the model plant *Arabidopsis thaliana* (125-Mb genome with ~26,000 annotated genes, The *Arabidopsis* genome initiative) (Grant et al. 2000; Paterson et al. 2000; Ku et al. 2000; Salse et al. 2002; Bowers et al. 2003).

Complete sequencing of *Arabidopsis* revealed a large number of gene duplications, and today, it is accepted that at least three duplication events have occurred in the *Arabidopsis* genome. The oldest duplication event related to angiosperm-gymnosperm divergence is thought to have occurred between 300-400 million years ago (Mya) (Savard et al. 1994; Maere et al. 2005), although more data is needed to confirm this. The second set of duplicated blocks was formed after the monocotyledonous/dicotyledonous divergence, estimated at between 112 and 200 Mya (Wolfe et al. 1989; Ku et al. 2000; Simillion et al. 2002; Bowers et al. 2003). More data is available on the most recent event, which was shown to result from a single polyploidy event that encompassed the whole genome (Simillion et al. 2002; Chapman et al. 2004). This occurred between the *Arabidopsis-Brassica* split and the *Arabidopsis-Cotton* split, and definitively after the *Arabidopsis-Medicago* split (Gandolfo 1998; Grant 2000). Time estimations range from 24 to 75 Mya (Lynch and Conery 2000; Simillion et al. 2002; Vision et al. 2000; Bowers et al. 2003; Ermolaeva et al. 2003; Blanc et al. 2003). Remarkably, despite a large number of duplications, conservation of gene order within small chromosomal regions has been found between *Arabidopsis* and distantly related plants such as rice, sorghum, tomato, *Capsella*, soybean and *Lotus japonicus* (Paterson 1996; Dodeweerd et al. 1999; Grant et al. 2000; Ku et al. 2000; Rossberg et al. 2001; Strake et al. 2004).

Comparative genetic experiments can be based on orthologous genes between different species. However, one difficulty is the fact that many genes are members of gene families, and it is consequently often difficult to determine if a gene mapped in the second species is orthologous or paralogous to that of the first species. The recent development of sequence



datasets for many species raises the possibility of using computational methods to assist in identifying sets of putative orthologs across multiple species (Fulton et al. 2002). Recently, a set of conserved single copy, orthologous genes (COS) was developed for comparative mapping across wide taxonomic distances (Lin et al. 2005; Wu et al. 2006). These sequences correspond to single or low copy genes; they have remained relatively stable in sequence since radiation of dicotyledonous plants (Fulton et al. 2002; Lin et al. 2005). Thus COS markers could help improve our understanding of the evolution of genomes and chromosomes, as well as facilitate the transfer of genome sequence data from fully sequenced genomes to species that lack such information.

Coffee is an important international commodity, ranking among the five most valuable agricultural exports from developing countries (Food and Agriculture Organization, <http://apps.fao.org>). Moreover, production and processing of coffee employs more than 25 million people worldwide. Commercial coffee production relies mainly on two closely related species: *Coffea arabica* and *Coffea canephora*, which account for respectively about 70 and 30% of coffee production worldwide (International Coffee Organization, <http://www.ico.org>). *C. arabica* ( $2n=4x=44$ ) is a recent amphidiploid species native to Africa containing two subgenomes that originated from two related diploid species ( $2n=2x=22$ ), *C. eugenioides* and *C. canephora* (Lashermes et al. 1999). Low divergence has been reported between the two constitutive subgenomes of *C. arabica* (Lashermes et al. 1999; Herrera et al. 2002) and, more generally, among the genomes of the different species belonging to the *Coffea* genus (Cros et al. 1998), supporting the view that all coffee-tree species derived from a basic coffee genome of eleven chromosomes.

Coffee leaf rust caused by the fungus *Hemileia vastatrix* is one of the most serious diseases affecting coffee cultivation in almost all coffee growing countries around the world. One resistance gene from *C. liberica* (i.e.  $S_{H3}$  resistance factor) has been introgressed into susceptible commercial *C. arabica* cultivars (Prakash et al. 2002). Amplified fragment length polymorphism (AFLP) markers linked to the  $S_{H3}$  resistance gene have been identified (Prakash et al. 2004). Recently (unpublished result), AFLP markers co-segregating with the  $S_{H3}$  gene were successfully converted into sequence-characterized amplified region (SCAR) markers and used to screen a *C. arabica* BAC (Bacterial artificial chromosome) library as a further step in the study of coffee leaf rust resistance gene(s) using a positional cloning strategy. A total of 22 positive clones were identified and

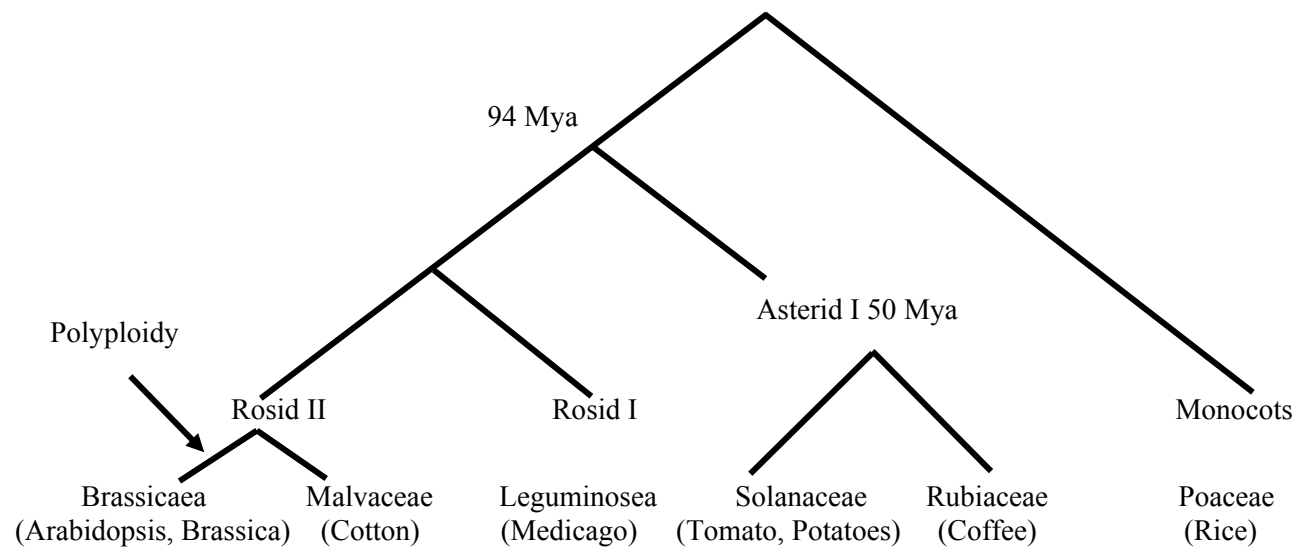


Figure 20. A dendrogram depicting the phylogenetic relationship of Coffee, *Arabidopsis* and other crop species. The arrow marks the probable occurrence of the polyploidy event identified by Blanc et al. 2003.

assembled in a BAC contig corresponding to one region of the coffee homeologous chromosome group 1 (i.e. linkage group 1 as defined in Lashermes et al. 2001).

The objectives of the present study were to assess microsynteny between coffee and *Arabidopsis* and to evaluate the possibility of transferring *Arabidopsis* genome sequence information to the related S<sub>H</sub>3 region in coffee. By combining a search of *Arabidopsis* sequences homologous to coffee BAC-end sequences belonging to the related S<sub>H</sub>3 BAC contig and use of COS markers, we demonstrated microsynteny between single copy segments in coffee and *Arabidopsis* duplicated counterparts. The implications for the evolution of the coffee genome and for genomic comparisons between *Arabidopsis* and coffee based on synteny are discussed.

### 9.3 Materials and methods

#### 9.3.1 BAC resources

The BAC library was constructed from the genomic DNA of the arabica cultivar variety IAPAR 59 (Noir et al. 2004). High-density colony filters were prepared using the Genetix Q-Bot. BAC clones were double spotted onto Hybond-N<sup>+</sup> filters (Amersham-Pharmacia) using a 4 x 4 pattern as described by Tomkins et al. (1999). In total, 27 648 clones of the BAC library were spotted onto a 22.5 x 22.5 cm filter representing approximately seven times basic coffee genome equivalents. Hit BAC clones were grown for 48 h at 37°C in 1.5 ml 2YT medium containing chloramphenicol (12.5 µg/ml). BAC DNA was extracted using an alkaline lysis procedure (Sambrook et al. 1989) and extremities sequenced from one or both insert sides using primers pIBf and pIBr (pIBf: 5' GGATGTGCTGCAAGGCGATTAAGTTGG 3'; pIBr: 5' CTCGTATGTTGTGTGGAATTGTGAGC 3'). All coffee BAC-end sequences that have been used in further comparative experiments between coffee and *Arabidopsis* were submitted to GenBank (accession numbers from ED98446 to ED989460).

#### 9.3.2 *Arabidopsis* homologous sequences

Coffee BAC-end sequences (BEs) were analysed against the *Arabidopsis* genome database using the Translated BLASTX algorithm in TAIR (The *Arabidopsis* information resource, <http://www.Arabidopsis.org/Blast>). The TBLASTX algorithm was chosen because

Table 9. *Arabidopsis thaliana* homologous sequences of coffee BAC-end sequences from the SH3-contig. The identified *Arabidopsis* predicted genes showing co-localisation in close proximity are indicated in bold.

Sequence codes	Coffee BAC end references	GenBank acc. numbers	Homologous genes of <i>A. t.</i> (e value)*	Identification of closest match
BES2	131_11D_f	ED989456	<b>AT3G23160 (2e-51)</b> AT5G04550 (1e-26) AT5G51670 (1e-24) AT1G34320 (8e-04)	expressed protein
BES3	64_1L_r	ED989452	<b>AT5G43650 (0.001)</b>	DNA binding / transcription factor
BES4	34_20L_f	ED989446	<b>AT5G43700 (1e-27)</b> <b>AT1G04240 (4e-25)</b> <b>AT4G14560 (5e-21)</b> <b>AT3G23030 (6e-19)</b> <b>AT4G14550 (4e-16)</b> <b>AT1G04250 (1e-15)</b> <b>AT3G23050 (7e-15)</b> AT4G29080 (2e-14) AT5G65670 (5e-13) AT3G04730 (5e-13) AT3G15540 (1e-12) AT2G22670 (2e-12)	auxin-responsive AUX/IAA family protein
BES6	124_12K_r	ED989455	<b>AT3G23000 (3e-67)</b> <b>AT4G14580 (4e-63)</b> AT5G01810 (2e-53) AT4G30960 (2e-53) AT5G10930 (2e-52) AT1G30270 (7e-52) AT5G25110 (1e-51) AT2G30360 (4e-51) AT5G45820 (6e-51) AT5G21326 (6e-50) AT1G01140 (8e-50) AT2G38490 (1e-49)	CBL-interacting protein kinase
BES7	41_23C_r	ED989447	<b>AT5G43760 (4e-81)</b> AT2G26640 (6e-80) <b>AT1G04220 (4e-76)</b> AT1G19440 (8e-66) AT2G16280 (4e-63) AT2G26250 (3e-55) AT1G68530 (5e-55) AT1G25450 (3e-51) AT1G01120 (8e-48) AT4G34510 (5e-47) AT4G34250 (4e-37) AT4G34520 (3e-35)	Putative Beta-ketoacyl-CoA synthase

\*Best twelve matches (if available) for SH3 Bac-end sequences are indicated in the table for *A. thaliana* (*A. t.*).

homologous amino acid sequences were detected more easily than nucleotidic homologies over the large evolutionary distance separating *Arabidopsis* and coffee (Figure 20).

### 9.3.3 *Arabidopsis* predicted genes present on syntenic *Arabidopsis* fragments.

*Arabidopsis* analogous predicted genes in *Arabidopsis* putative paralogous segments were identified by screening Segmental Genome Duplications in *Arabidopsis thaliana* databases ([http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis\\_genome\\_duplication.shtml](http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis_genome_duplication.shtml)). The *Arabidopsis* BAC names concerned were MWF10, K9D7, MQO24, MQD19, F6B6, MRH10, MLN1, F19P19, F20D22, MWI23, F5N5, MXC7, K13C10, K14B15, F28F24, MLM24, FCAALL.

### 9.3.4 Coffee orthologs

We selected predicted genes with either a single copy in the *Arabidopsis* genome or duplicated on at least three of the four *Arabidopsis* paralogous fragments. Special attention was paid to these paralogous *Arabidopsis* predicted genes as we assumed that they derived by duplications from genes present on the *Arabidopsis* ancestral genome fragment. In order to identify coffee orthologous sequences, the targeted *Arabidopsis* predicted genes were compared against extracted coffee (or by default tomato) ESTs from Solanaceae Genomics Network (SGN) databases (<http://www.sgn.cornell.edu/content/coffee.pl>) using the translated BLASTX algorithm.

### 9.3.5 Hybridization experiment

Primer combinations specific to coffee or tomato orthologs to *Arabidopsis* and primer combinations specific to coffee BEs were designed using Primer3 programme (Whitehead Institute, USA). Probes were labeled and filter hybridization conditions were as described in Noir et al. 2004.

### 9.3.6 BAC fingerprinting and contig assembly

Agarose gel-based BAC DNA fingerprinting was performed as described by Marra et al. (1997) with some modifications. In brief, approximately 250 ng DNA was digested with

Table 10. COS markers corresponding to predicted genes located on the targeted *Arabidopsis* segments identified. Putative orthologous coffee and tomato sequences and derived primers are indicated.

Coffee sequences	Targeted <i>Arabidopsis</i> ORFs	Identified orthologous unigenes		TBLASTX		Primers (5' to 3')		Probe (bp)	
		Coffee	Tomato	Scores	E-values	Forward	Reverse	Coffee	Tomato
<b>COS markers developed from single <i>Arabidopsis</i> ORFs</b>									
COS-18	AT1G04080	SGN-U301490		96	2e-27	ACCTCATGGGAGTTCACCAG	GTGCATAGCTTTGCTGACCA	385	
COS-13	AT1G04190	SGN-U310991		186	7e-48	TCTCTCGAGCAGCTTGCTTA	CCCAAGACGCCTAATCCAA	1800	
COS-8	AT3G22960	SGN-U300680		673	0	ATCATTCTAGCGTCCGATGG	GAGAAGCTCAAGCGAAATGG	1900	
COS-1	AT3G23400	SGN-U310636		158	1e-43	GCATCAAAAAGAGGCTGCATT	TCAGTTCGACAGGTGGAAGA	500	
COS-19	AT5G43680	SGN-U306277		124	1e-50	ACCGGAGTTTGTTCAGTGG	ACTCCTTCCTTGCTGGATGA	530	
COS-5	AT5G43750	SGN-U306927		112	5e-26	GCCCTCAGTTTTCTCCATCA	GCAAAAACAGCTGCTCCAAT	1200	
COS-11	AT5G43850		SGN-U317544	374	1e-104	ACCGGAGTTTGTTCAGTGG	ACTCCTTCCTTGCTGGATGA*	-----	1700
COS-16	AT5G43920	SGN-U304045		233	3e-99	TGCTCAAACCTCTGGGATGTG	TTTACCCACTTCCCTGCAAC	415	
COS-17	AT5G43940	SGN-U299947		857	0	http://www.sgn.cornell.edu/		350 and 500	
<b>COS markers developed from duplicated <i>Arabidopsis</i> ORFs</b>									
COSd-14	AT1G04160, AT5G43900, AT4G28710	SGN-U306823		264	1e-140	TGTGTATCCAGGCACCAAG	ATATGTTTGAGTTCATCCCA	650	
COSd-9	AT3G22890, AT4G14680, AT5G43780	SGN-U304277		540	1e-154	TAGCAAGAACATGGGGAACC	GGAGCCATGCTCAGTACCTC	700	

\*Primers developed from the tomato unigene

either the restriction enzyme *HindIII* or *EcoRI*. DNA fragments were separated into a 1% agarose gel by electrophoresis at 80 V in circulating 1x TAE buffer at 4°C for 16 h. Mixture of marker size DNA (Raoul, MPbiomedicals) was loaded into every five lane. The gel was stained in 1 l of a 1: 20000 dilution of Vistra Green (Molecular Probe, USA) in 1x TAE for 30 min. Gel images were captured using a Typhoon scanner (9400 series, Amersham Biosciences) and exported as tagged image files. Fingerprints were edited using the Image 3.10 module of the finger printed contig (FPC) package (Sanger centre, USA). BAC contigs were assembled using FPC, version 7.2 program. Sulston scores of  $1 \times 10^{-10}$  and  $1 \times 10^{-13}$  were tested using the stringent conditions of tolerance 5.

### 9.3.7 Genetic mapping

For genetic mapping, the plant material consisted of a *C. arabica* F<sub>2</sub> population of 101 individuals derived from a cross between two inbred lines, Matari and S.288 (Prakash et al. 2004). Genomic DNA was isolated from lyophilized leaves by means of a nuclei isolation step as described by Agwanda et al. (1997). Amplification conditions using specific primer pairs (either Coffee BAC-end or COS), radioactive labeling and polyacrylamide gel electrophoresis were as reported elsewhere (Combes et al. 2000). The markers were revealed on Amersham storage Phosphor screen and screened after 24 hr with a Typhoon scanner (9700 series, Amersham Biosciences) to obtain digital images. Linkage analysis and map construction were performed using the computer program MAPMAKER version 3.0b (Lander et al 1987). Recombination frequencies were converted into map distances or centiMorgan (cM) values using the Kosambi function.

## 9.4 Results

### 9.4.1 Anchoring to the Arabidopsis genome

Fourteen BEs from twenty coffee clones belonging to the S<sub>H</sub>3-linked contig (Contig A) were analysed against the *Arabidopsis* genome database using the TBLASTX algorithm in TAIR (The *Arabidopsis* information resource, <http://www.Arabidopsis.org/Blast>). TBLASTX matches with an arbitrary threshold limit of the E-value of less than  $10^{-3}$  were further investigated. *Arabidopsis* sequences homologous to coffee were found for five coffee BEs (Table 9). Among the *Arabidopsis* predicted genes identified, only those showing co-localisation in close proximity (i.e. less than 100kb) on an *Arabidopsis*

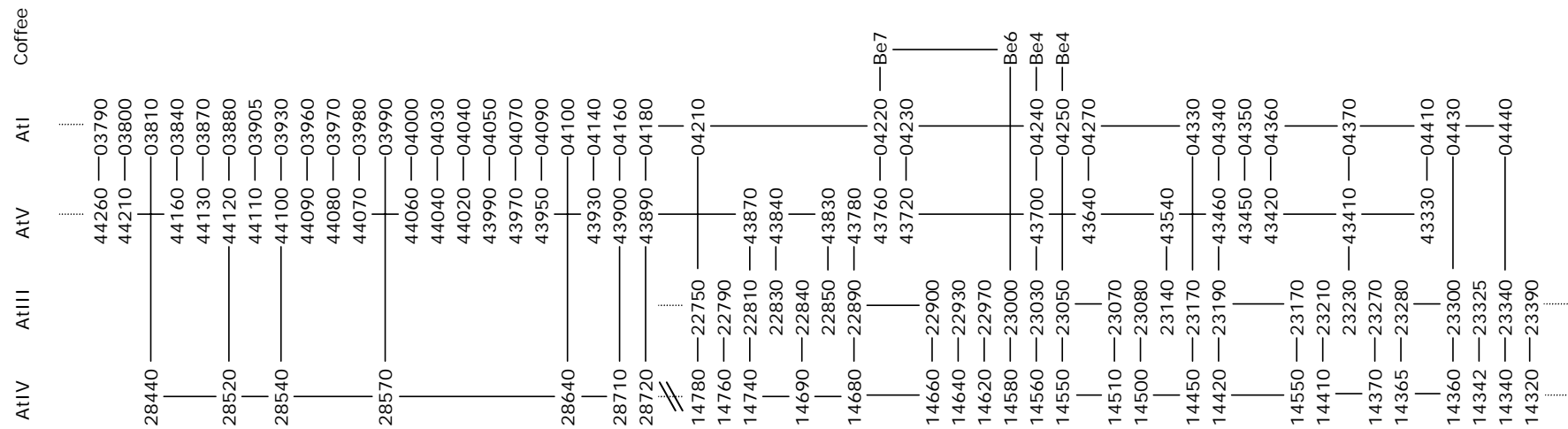


Figure 21. Graphic representation of the putative paralogous predicted genes on the duplicated *Arabidopsis* segments (chromosomes 1, 3, 4, 5) pointed by coffee BAC-end sequence, based on Segmental Genome Duplications of *Arabidopsis* in TIGR database ([http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis\\_genome\\_duplication.shtml](http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis_genome_duplication.shtml)).



chromosomal segment were subsequently considered. In total, 13 *Arabidopsis* sequences homologous to coffee BEs were retained (Table 9) and pointed to four different regions of the *Arabidopsis* genome located on chromosomes I (AtI), III (AtIII), IV (AtIV) and V (AtV).

#### 9.4.2 Analysis of targeted *Arabidopsis* regions

Fine analysis of the targeted *Arabidopsis* segments in TIGR databases of Segmental Genome Duplications of *Arabidopsis* ([http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis\\_genome\\_duplication.shtml](http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis_genome_duplication.shtml)) revealed several predicted genes in duplicate, triplicate or quadruplicate between the four segments of *Arabidopsis*, forming a network of matching paralogous predicted genes (Figure 21). The link between the four regions ran from *At1G03800* to *At1G04440* (306kb), *AT5G44210* to *AT5G43330* (481kb), *AT3G22750* to *AT3G23340* (320kb). There was a break in the syntenic network with the *Arabidopsis* segment IV; paralogous predicted gene matches ended at *At4G28720* and the next started at *At4G14780*. Hence the estimated size of segment AtIV is 84 kb for *AT4G28440* to *AT4G28720* and 147 kb for *At4G14780* to *At4G14340*. The homologous *Arabidopsis* sequences of three coffee BEs anchoring the coffee contig to *Arabidopsis* segments were in duplicate (BEs6 and BEs7) or quadruplicate (BEs4) on different *Arabidopsis* segments. Anchoring on segments AtIII and AtV was further confirmed by one homolog each, i.e. BEs2 and BEs3.

#### 9.4.3 Development of COS markers

Coffee putative COS markers corresponding to 11 predicted genes located on the different targeted *Arabidopsis* segments were developed (Table 10). Nine COS corresponded to single copy *Arabidopsis* predicted genes while two COS corresponded to *Arabidopsis* paralogous genes (COSd). The primers were designed based on the closest orthologous sequence identified in coffee except for the COS-11, for which a tomato sequence was used. The primer pair for the orthologous sequence of *AT5G43850* identified in coffee was already available in the SGN database (<http://www.sgn.cornell.edu/index.pl>).

Strongly hybridised positive BAC clones detected on BAC filters were identified for the 11 COS. To consolidate and extend the synteny between *Arabidopsis* and coffee, nucleotide

Table 11. *Arabidopsis thaliana* homologous sequences from BAC clones hybridized by COS markers presenting significant homologies to *Arabidopsis thaliana* sequences (in bold) located on the targeted *Arabidopsis* segments.

Sequence codes	Coffee BAC end references	GenBank acc. numbers	Homologous genes of <i>A. t.</i> (e value)*	Identification of closest match
BES12	143_8C_r	ED989459	<b>AT5G43870 (2e-15)</b> <b>AT4G14740 (5e-15)</b> <b>AT3G22810 (8e-13)</b> AT3G63300 (4e-12)	Expressed protein
BES20	143_8C_f	ED989458	AT1G05460 (2e-56) AT5G47010 (2e-10) AT2G03270 (4e-10) AT2G19120 (6e-08) AT4G30100 (1e-07) AT1G65810 (2e-07) AT1G08840 (3e-07) AT1G65780 (4e-07) AT1G16800 (5e-07) AT5G35970 (1e-06) AT4G05540 (8e-06) AT5G37150 (1e-05)	RNA Helicase
BES2'	71_19M_r	ED989453	<b>AT4G23160 (1e-48)</b>	Protein kinase family protein
BES10	34_11L_r	ED989460	<b>AT1G04210 (1e-33)</b> AT4G31170 (7e-06) AT2G24360 (4e-05) AT1G50230 (1e-04) AT5G58950 (2e-04) AT5G50180 (3e-04) AT3G46930 (3e-04) <b>AT4G14780 (4e-04)</b> AT1G51660 (5e-04) AT3G27560 (7e-04) AT2G23450 (7e-04) <b>AT3G22750 (0.001)</b>	leucine-rich repeat family protein
BES15	46_24D_f	ED989449	<b>AT1G04150 (3e-60)</b> AT1G51570 (2e-58) AT3G57880 (1e-56) AT5G12970 (6e-56) AT5G48060 (8e-50) AT5G06850 (7e-48) AT4G11610 (8e-47) AT4G00700 (4e-45) AT4G20080 (1e-41) AT1G22610 (7e-41) AT3G61300 (4e-37) AT3G03680 (2e-35)	C2 domain-containing protein
BES16	45_4E_f	ED989448	<b>AT4G14750 (3e-28)</b> AT4G23060 (1e-09) AT3G16490 (5e-09) AT1G01110 (8e-09) AT5G62070 (9e-08) AT5G07240 (4e-07) AT4G29150 (8e-07) AT1G51960 (2e-06) AT4G00820 (6e-06) AT5G13460 (0.002) AT2G43680 (0.002) AT2G43680 (0.002)	Calmodulin-binding family protein
BES17	136_1N_f	ED989457	<b>AT3G22910 (1e-36)</b> AT3G63380 (9e-35) AT3G21180 (3e-32) AT4G29900 (5e-27) AT5G57110 (2e-25) AT5G57110 (2e-25) AT5G53010 (1e-21) AT3G57330 (4e-21) AT2G41560 (4e-21) AT2G22950 (1e-19) AT4G37640 (2e-19) AT1G27770 (3e-18)	Calcium-transporting ATPase

\*Best twelve matches (if available) for S<sub>H</sub>3 Bac-end sequences are indicated in the table for *A. thaliana* (*A. t.*)

sequences of BEs from coffee BAC clones were determined. Sequences were analysed against the *Arabidopsis* genome database with TBLASTX in TAIR (The *Arabidopsis* information resource, <http://www.Arabidopsis.org/Blast>). Significant homologies to *Arabidopsis* predicted genes were revealed for seven BEs, four of which were detected in the segments identified in *Arabidopsis* (Table 11). Further hybridization was conducted with four BEs, one of which had no sequence homology (BEsnh) to specific *Arabidopsis* segments.

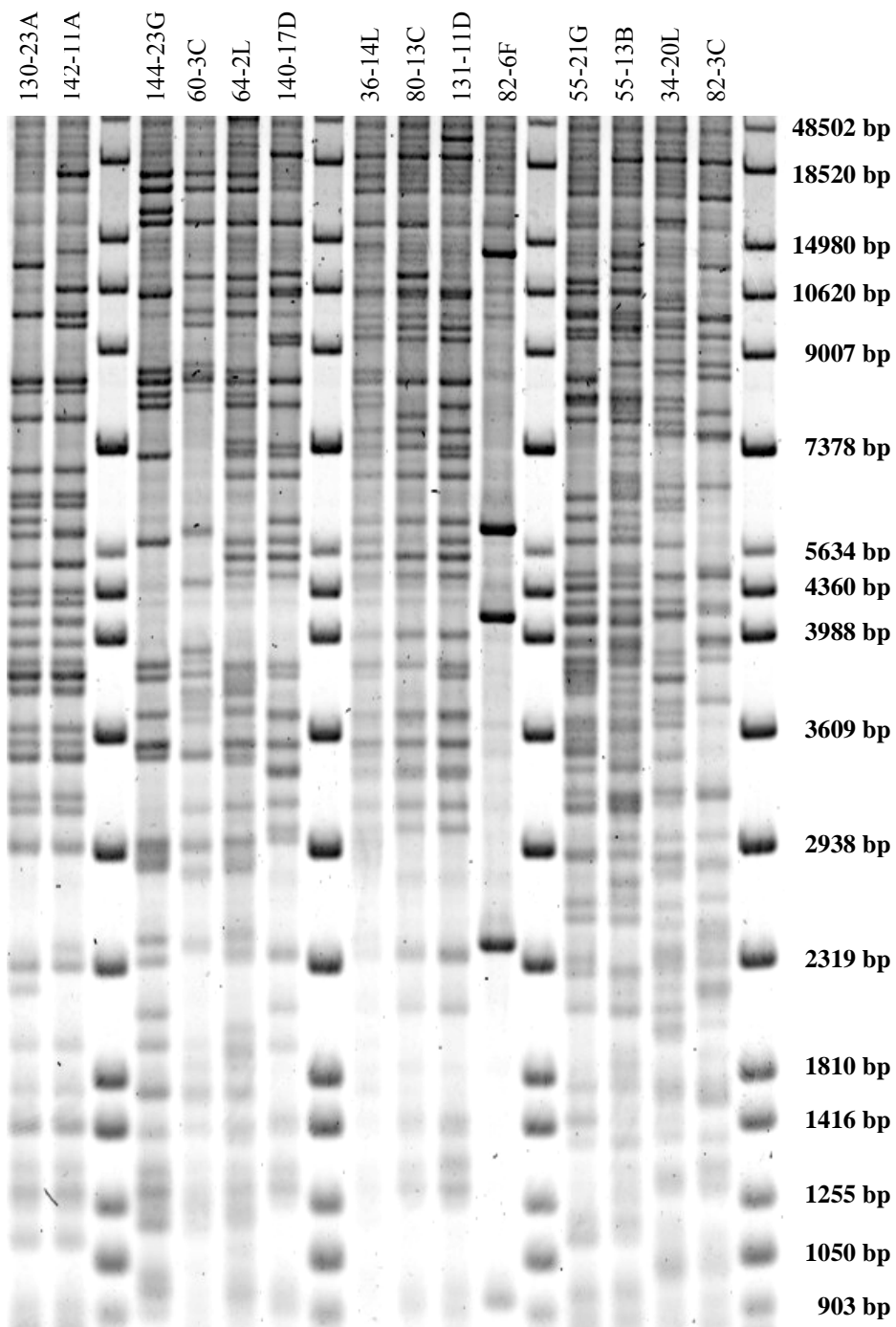
#### 9.4.4 Contig assembly

With the exception of COS-1 probe, all probes detected BAC clones that hybridised with at least one other probe. Contigs were therefore assembled from BAC clones that hybridized to two or more single copy probes. A total of forty-two BACs were co-hybridised with COS and BEs marker probes and they could be assembled in two contigs (Figure 22). Contig A ( $S_{H3}$  related contig) was the original region of interest and was made of 33 BAC clones, with five COS (COS-5, COS-8, COS-9, COS-11 and COS-13) from *Arabidopsis* segments AtI, AtIII, AtIV and AtV and thirteen BEs, four showing strong homologies to *Arabidopsis* predicted genes (BEs2, 3, 4, 6, 7, 12, 16, 17, and 19) and three that do not (BEsnh1 to 3). Co-hybridisation with COS, COSd and BEs enabled the assembly of another BAC contig (i.e. contig B). This contig B consisted of 9 overlapping BAC clones, co-hybridized by COS-14 to COS-18 and BEs15. The two constructed contigs beared markers from different duplicated *Arabidopsis* segments (Figure 23).

#### 9.4.5 Validation of contig assembly

The thirty-three BAC clones belonging to contig A were fingerprinted using restriction enzymes. On average, 30 restriction fragments per clone and per enzyme, ranging from 50 kb to 1000 bp nucleotides, were obtained and included in the FPC analysis. At a Sulston score of  $1 \times 10^{-13}$ , a total of 5 contigs comprising 2 to 7 BAC clones and 1 singleton (i.e. BAC 82-6F) were obtained. This singleton appeared to be associated with poor fingerprinting (Figure 24) and the BAC clone 82-6F was therefore excluded from the further analysis. Increasing the Sulston scores to the less stringent condition of  $1 \times 10^{-10}$  resulted in a unique contig of 32 clones confirming the previous contig assembly based on BAC filter hybridisations.

Figure 24. Restriction fragment comparison (i.e. EcoRI-digest) of Coffee BAC clones belonging to the contig A previously defined based on BAC filter hybridisations. Molecular marker was loaded into every five lane.



Attempts were made to genetically map the different coffee COS and BEs markers using the *C. arabica* mapping F<sub>2</sub> (Matari x S.288) population. Three sequence-specific primers corresponding to BAC-end sequences of clones belonging to contig A (i.e. 48\_21O\_f, 124\_12K\_f and 45\_4E\_r) revealed length polymorphism by polyacrylamide gel electrophoresis between the two parental lines, Matari and S.288. None of COS makers exhibited polymorphism. For the three markers showing polymorphism, the 101 individuals of the F<sub>2</sub> (Matari x S.288) population were genotyped and a linkage analysis was performed. A perfect co-segregation was observed between 124\_12K\_f and 45\_4E\_r while the marker 48\_21O\_f was located at only 0.8 cM.

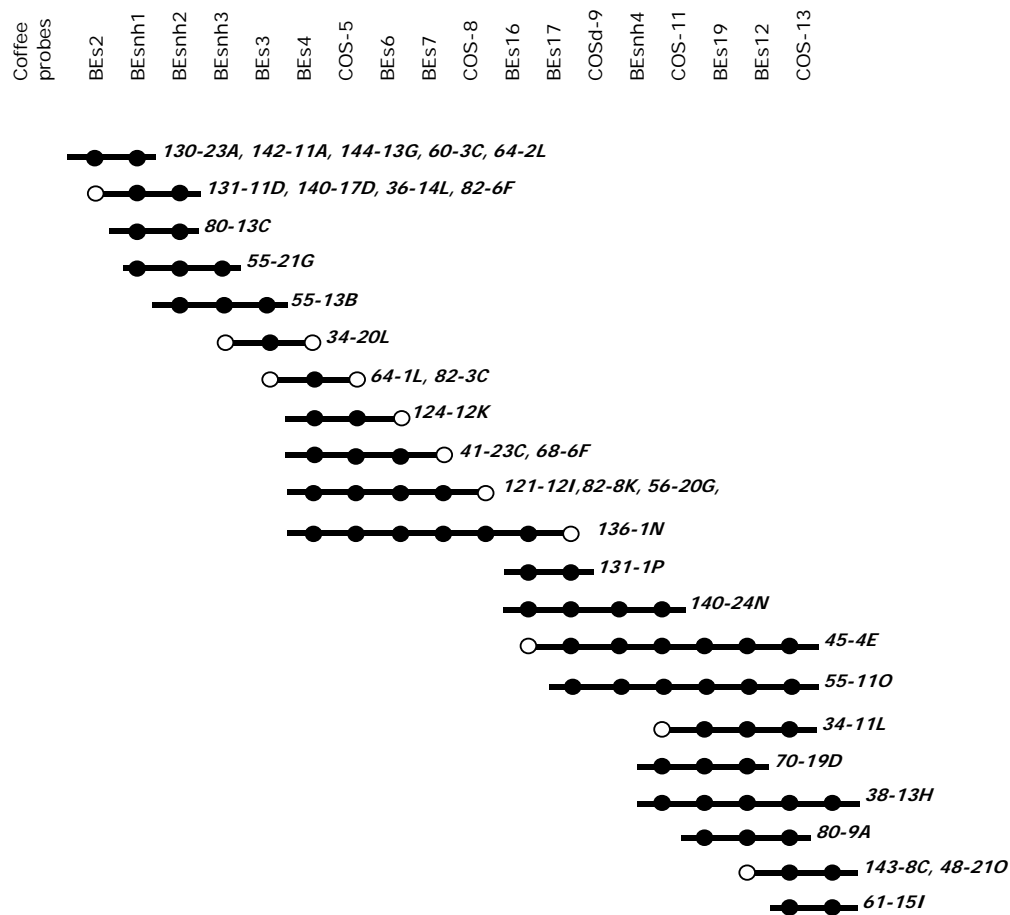
#### 9.4.6 Synteny between coffee and syntenic segments of *Arabidopsis*

The number of anchoring orthologous markers between Coffee and individual *Arabidopsis* regions was seven matches for AtI, ten for AtIII, eight for AtIV and eleven for AtV (Figure 23). Although Coffee can be anchored (by hybridisation or sequence homology) with all four *Arabidopsis* regions, one coffee BE, 143\_8C\_f was significantly matched to another segment of *Arabidopsis*, and two BEs displayed the same homologous match (BEs2 and BEs2'). Despite these limits, close inspection of the regions detected by TBLASTX analysis and hybridisation against coffee BAC libraries using COS and COSd probes provided evidence for microsynteny between coffee and the *Arabidopsis* duplicated segments. Comparative mapping of different orthologous *A. thaliana* and coffee anchored markers suggested perfect conservation of gene order across the coffee contigs and *Arabidopsis* when AtV and AtI segments were assayed (Figure 23). Good conservation of gene order was also observed between segment AtIII and segment AtIV except for markers BEs10 and BEs12 which were inverted compared to the coffee genome. Overall, good microsynteny was observed between three coffee contigs and the different duplicated segments of *Arabidopsis*. Interestingly, microsynteny was based on orthologs located on different *A. thaliana* chromosomes.

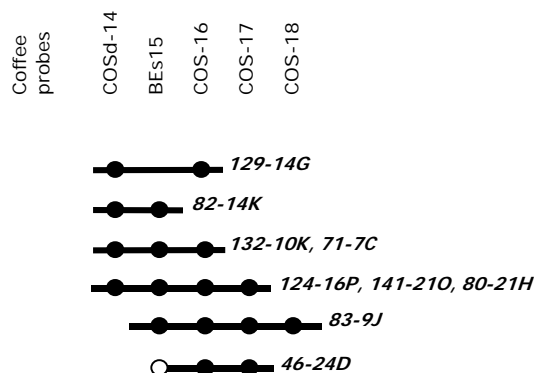
## 9.5 Discussion

Significant microsynteny was revealed between coffee genome segments related to the S<sub>H</sub>3 region and the *Arabidopsis thaliana* genome. Using information from duplicated *Arabidopsis* counterparts, new molecular markers were generated and the Coffee BAC contig related to the S<sub>H</sub>3 region was extended and consolidated. The contig assembly was

Figure 22. Coffee BAC clone contigs assembled using coffee BAC-end sequences (BEs), BAC-ends with non-homologous *Arabidopsis* gene sequences (BEsnh) and COS derived probes from segments homoeologous to *Arabidopsis* duplicated chromosome 1, 3, 4 and 5. Circle: hybridization with Coffee probes. Open circle: sequenced BAC ends.



b) Contig B



confirmed by both BAC fingerprinting analysis and genetic mapping of coffee BAC-end sequences. *Arabidopsis* thus appears to be a useful resource for marker saturation in a coffee specific region.

An original two-way strategy was developed. The first step involved *in silico* localisation of the putative *Arabidopsis* counterparts to coffee BAC-end sequences belonging to an established BAC contig. In the second step, coffee EST sequences putatively orthologous to duplicated or single *Arabidopsis* predicted genes from the candidate syntenic *Arabidopsis* segments were identified and used to derive putative COS markers that were then used in hybridisation screening of a coffee BAC library. BAC contig assembly confirmed the orthology between the identified coffee sequences and the targeted *Arabidopsis* predicted genes. Thus, a network of microsynteny between BAC contigs of the S<sub>H</sub>3 coffee region and four duplicated segments of *Arabidopsis* chromosomes (Ath) I, III, IV and V was successfully established. As long as EST resources are available to enable comparison of the species with *Arabidopsis*, this approach should be quite straightforward and will be useful for many species for which large genome sequence data are not yet available. In particular, the present results highlights the interest of COS markers in comparative genome mapping.

Although the S<sub>H</sub>3 coffee region can be anchored with all four *Arabidopsis* regions via conserved sequence markers, individually, none of the *Arabidopsis* segments contains the full set of anchoring markers. The anchoring markers are scattered throughout the four different syntenic segments of *Arabidopsis*. A similar complex network of synteny has been already observed during comparisons of *Arabidopsis* and other species like *Prunus* (Jung et al. 2006), *Lycopersicon esculentum* (Ku et al. 2000), *Glycine max* (Foster-Hartnett et al. 2002; Yan et al. 2004), *Medicago truncatula* (Yan et al. 2003, 2004; Kevei et al. 2005) or *Oryza sativa* (Mayer et al. 2001). This common configuration is believed to be due to multiple rounds of duplication in the *Arabidopsis* lineage and cannot be explained by simple chromosome rearrangements. As reported by Blanc et al. (2003), the segments of chromosome 3 and 4 on the one hand, and the segments of chromosome 1 and 5 on the other hand, derived from duplications that occurred during the most recent whole-genome duplication event. Based on the present data, it could be assumed that these two sister duplications (i.e. 3-4 and 1-5) derived from an ancient duplication. The four matching segments of *Arabidopsis* would then be the vestiges of these two successive duplication

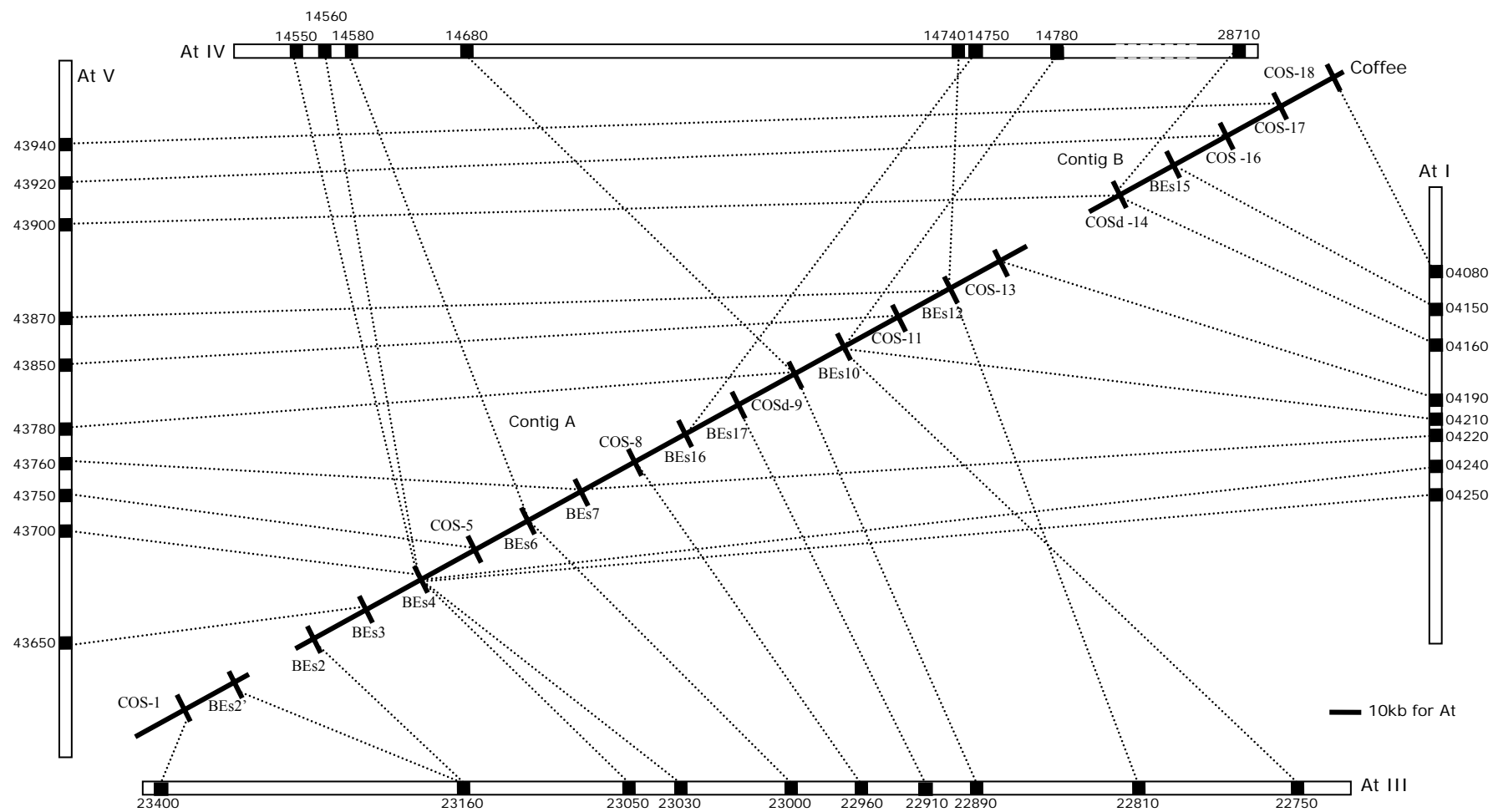


Figure 23. Comparative maps between the four duplicated *Arabidopsis* segments, using the probes (COS and BEs) as ordered on coffee contigs. *Arabidopsis* predicted genes and coffee orthologous counterpart are linked by dashed line.



events and the S<sub>H3</sub> region would be the orthologous counterpart to these in the coffee genome lineage. In this case, the complex network of synteny observed would reflect independent and progressive gene loss in the different *Arabidopsis* segments subsequent to the large-scale duplication events. Furthermore, the anchoring markers appeared in same order on the coffee BAC contigs and in all *Arabidopsis* segments with the exception of a single inversion on AtIII and AtIV *Arabidopsis* segments. This result suggests that all *Arabidopsis* and coffee segments diverged from a common template and that deletion of individual genes must have occurred more frequently than rearrangement (i.e. inversion and translocations) during the evolution of these regions.

Recently, Wu et al. (2006) reported that no polyploidization event occurred immediately prior to or after radiation of the family Rubiaceae that includes the genus *Coffea*. Consistently, the S<sub>H3</sub> region was not duplicated in the coffee genome suggesting that the genome duplication events at the origin of its *Arabidopsis* counterparts occurred after the separation of Euasterid (Coffee) and Eurosoid (*Arabidopsis*) which diverged more than 94 million years ago (Crepet et al. 2004).

In the present study, the complex duplication history of *Arabidopsis* did not prevent the use of *Arabidopsis* as a genetic and genomic model for coffee species. Anchoring markers of the S<sub>H3</sub> coffee region from the different *Arabidopsis* paralogous segments were successfully integrated. On the contrary, the duplicated configurations could be an advantage when selecting and using COS makers in comparative genome mapping. Assuming that the *Arabidopsis* predicted genes that are shared by all or some duplicated fragments are more likely to be present on the targeted ancestral *Arabidopsis* fragment, COS markers corresponding to these *Arabidopsis* predicted genes would have more chance of being present in the orthologous counterpart of the species concerned. Although the data from this study is only preliminary, they tend to support this hypothesis.

Although the importance of large-scale duplications in the organisation of the *Arabidopsis* genome is well documented (The *Arabidopsis* genome initiative, 2000; Blanc et al. 2000, 2003; Simillion et al. 2002, Maere et al. 2005), the number, extent, overlapping and evolution of these duplications are far from being fully established. Reconstruction of the *Arabidopsis* ancestral genome has been attempted by Blanc et al. (2003) using protein sequence similarity to define gene paralogy. Duplications associated with the more recent polyploidy event were clearly identified and the approximate gene order that existed before

the recent polyploidy was partially determined. This pseudo-ancestral *Arabidopsis* genome has been advantageously utilized in comparative genome analysis of *Arabidopsis* with other plant genomes such as *Prunus* (Jung et al. 2006). However, old duplication events appear to be relatively difficult to detect, even with a complete genome sequence in hand, and a more comprehensive history of the *Arabidopsis* genome evolution is still required. In the present study, it is noteworthy that, in terms of gene content, the S<sub>H</sub>3 coffee region appears to be closer to the ancestral genome segment (before the divergence of *Arabidopsis* and coffee) than any of the duplicated counterparts in the present-day *Arabidopsis* genome. Thus, using a large set of COS markers (Wu et al. 2006) and by comparing *Arabidopsis* to plants such as coffee that apparently did not undergo important changes after divergence from the ancestral dicot, we may be able to depict the *Arabidopsis* genome evolution in more detail.

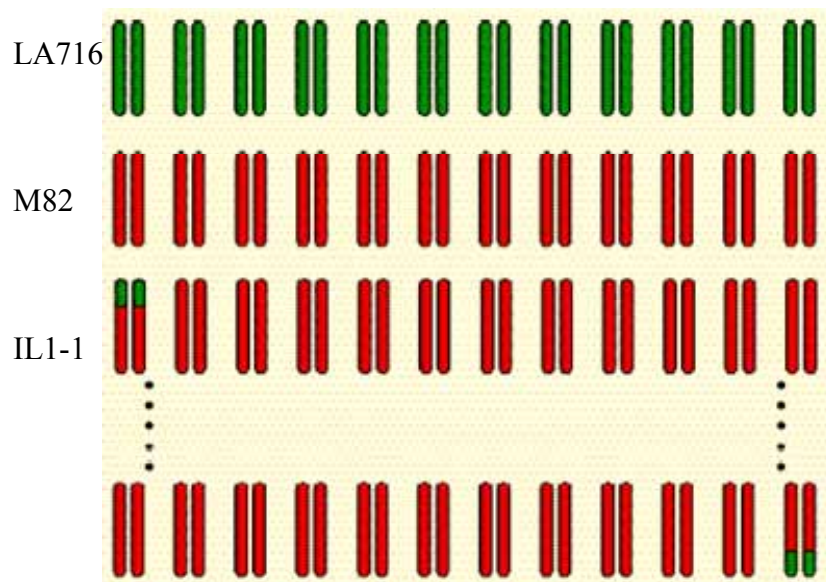


Figure 25. Lignées introgressées quasi-isogéniques issus du croisement entre LA716 et M82 rétro-croisé sur M82.

## **10 Cartographie génétique de la région chromosomique $S_{H3}$ sur *Solanum lycopersicum***

Les ressources moléculaires de la tomate se sont fortement développées. Ils existent des cartes très denses des chromosomes, plusieurs banques cDNA, YAC, BAC et d'autres bibliothèques de données génomiques (Martin *et al.* 1992; Budiman *et al.* 2000; Van der Hoeven *et al.* 2002; <http://sgn.cornell.edu>; <http://www.tigr.org>). Le développement de ressources importantes chez la tomate et le séquençage imminent de son génome en font une plante modèle pour des espèces proches.

La famille des Rubiacées à laquelle appartient le caféier et celle des Solanacées comprenant la tomate, font partie du même groupe Asterid I. Ces deux espèces ont divergé il y a environ 50 millions d'années (Gandolfo *et al.* 1998; Crepet *et al.* 2004). En raison de la relative proximité phylogénétique entre le caféier et la tomate, nous avons entrepris de cartographier sur le génome de la tomate la région chromosomique portant le facteur de résistance  $S_{H3}$ . Cette démarche a été réalisée dans le but d'améliorer la cartographie physique de cette région du caféier par le transfert de marqueur tomate et de développer une approche gène-candidat.

### **10.1 Une population de lignées introgressées pour la cartographie génétique chez la tomate**

Pour faciliter la cartographie génétique, une population polymorphe est nécessaire pour la cartographie génétique de la région chromosomique du caféier sur la tomate. La population utilisée pour la cartographie est composée d'une série de plantes contenant la séquence complète du génome de *L. pennellii* introgressé dans le génome de *L. esculentum* (Eshed et Zamir 1995; Liu et Zamir 1999). Cette population comprend 75 lignées quasi-isogéniques, issues du croisement interspécifique entre *L. pennellii* (LA716) et *L. esculentum* (M82) (Eshed et Zamir 1995) (Figure 25) Les chromosomes sont divisés en 107 bins ou segments qui représentent des positions relatives d'intervalles chromosomiques introgressés et chevauchant entre lignées (Tanksley *et al.* 1992). La carte génétique a été construite par l'analyse en Southern blot des 75 lignées, digérées par différentes enzymes de restriction, en utilisant des sondes tomates RFLP standards (Eshed et Zamir 1995).

Tableau 12. Marqueurs orthologues putatifs de *Lycopersicon* aux séquences des extrémités de BAC ou de gènes du caféier présent sur le contig portant le caractère de résistance S<sub>H3</sub>.

Séquences du caféier	Séquences de <i>Lycopersicon</i>	de Amorces (5' à 3')		Sondes (bp)
		Sens	Antisens	
<b>Marqueurs COS markers développés à partir de séquences codantes du caféier</b>				
SGN-127047	SGN-U218377	GCCCTCAGTTTTCTCCATCA	GCAAAAACAGCTGCTCCAAT	1100
SGN-121379	SGN-U222681	ATCATTCTAGCGTCCGATGG	GAGAAGCTCAAGCGAAATGG	2000
<b>Marqueurs COS markers développés à partir de fin du caféier</b>				
69-17J-r + 140-18D-r	SGN-U215128	ACTTCTGGGTTGGATGTCG	GGCGCTCATAATCACACTGA	1700
41-23C-r	SGN-U223808	TCCTCGCCCTGTATACCTTG	CCACGGAGCTTGTAATGGTT	350
34-20L-f	SGN-U221342	ATCATGTCAAAAACCCACCA	ATCCCAAGCCTCTTGCTTCT	900
131-11D-f	SGN-U238933	TTGTGTGTCTGATGCTGCAA	AAAGGTGCAAGCCATTTGAG	200
124-12K -f	SGN-U227406	GTGCACCCAGTTGGTTTCTT	ATGGAATGTTTCATCCCGAAA	1500

Tableau 13. Résultats d'hybridation sur des lignées introgressées avec les sondes de *Lycopersicon*. Les enzymes de restriction testés et les positions déterminées sur les ILs sont listés.

Sondes tomate	Enzymes	Positions sur les ILs
SGN-U223808	EcoRI	8.2; 8.2.1
	EcoRV	<b>2.1.1</b>
SGN-U238933	XbaI	<b>2.1.1</b> ; 9.3; 9.3.2
SGN-U227406	EcoRV	<b>2.1.1</b>
SGN-U215128	EcoRI	<b>2.1.1</b>
SGN-U221342	XbaI	<b>2.1.1</b> ; 5.3
SGN-U218377	HindIII	Pas de polymorphisme
SGN-U222681	HindIII	2.6.5

## 10.2 Détermination des orthologues tomates

La cartographie génétique sur la tomate est effectuée par la méthode RFLP à l'aide de sonde. Afin d'optimiser l'hybridation des sondes sur les membranes de population de lignées tomate, nous avons développé des sondes provenant de séquences tomate. Ces sondes correspondent aux séquences orthologues putatives des extrémités des BACs du contig S<sub>H</sub>3 du caféier ou de gènes présents sur ce contig. Pour 11 extrémités de BAC du contig S<sub>H</sub>3, cinq séquences orthologues putatives de *Solanum lycopersicum* ont été identifiées et deux gènes sélectionnés grâce à la comparaison des génomes *Arabidopsis*/caféier (voir le chapitre précédent). Les amorces ont été développées pour ces sept séquences tomate (Tableau 12).

## 10.3 Hybridation sur des lignées introgressées quasi-isogéniques

La cartographie génétique sur la tomate a été effectuée sur des lignées introgressées (ILs). La cartographie génétique sur les ILs a été réalisée par RFLP. Dans le but de révéler du polymorphisme l'hybridation des sondes tomate orthologues s'est effectuée dans un premier temps sur des membranes comprenant les parents LA716 et M82 digérés par cinq enzymes de restriction. Les membranes représentant l'ensemble des introgressions de *S. pennellii* digéré par l'enzyme adéquate ont ensuite été hybridées (Figure 26). Les régions montrant une disparition de la bande du parent M82 et/ou apparition de la bande LA716 correspondant au polymorphisme détecté chez les parents sont retenues.

Une sonde ne présentait pas de polymorphisme de séquence entre M62 et LA716, cinq ont indiqué la même région chromosomique et la dernière une région chromosomique différente. Cette dernière sonde tomate ne représentait probablement pas la séquence orthologue du caféier. Sur les cinq ayant indiqué la même région chromosomique, trois détectent aussi des segments de chromosomes différents, ce qui est très probablement lié à la nature multi-copie des sondes utilisées. Le polymorphisme de plusieurs sondes converge sur la région IL2-1.1 ou bin 2.1.1 (Tableau 13). Cette région correspond à l'extrémité du chromosome II de la tomate. D'après la carte génétique de la population de lignées d'introgressions le bin 2.1.1 est entièrement chevauchant avec le bin 2.1 (Tanksley *et al.* 1992), un polymorphisme aurait aussi dû être détecté sur les lignées d'introgression IL-2.1. Il est possible que l'introgression IL-2.1.1 couvre une région plus grande à l'extrémité du

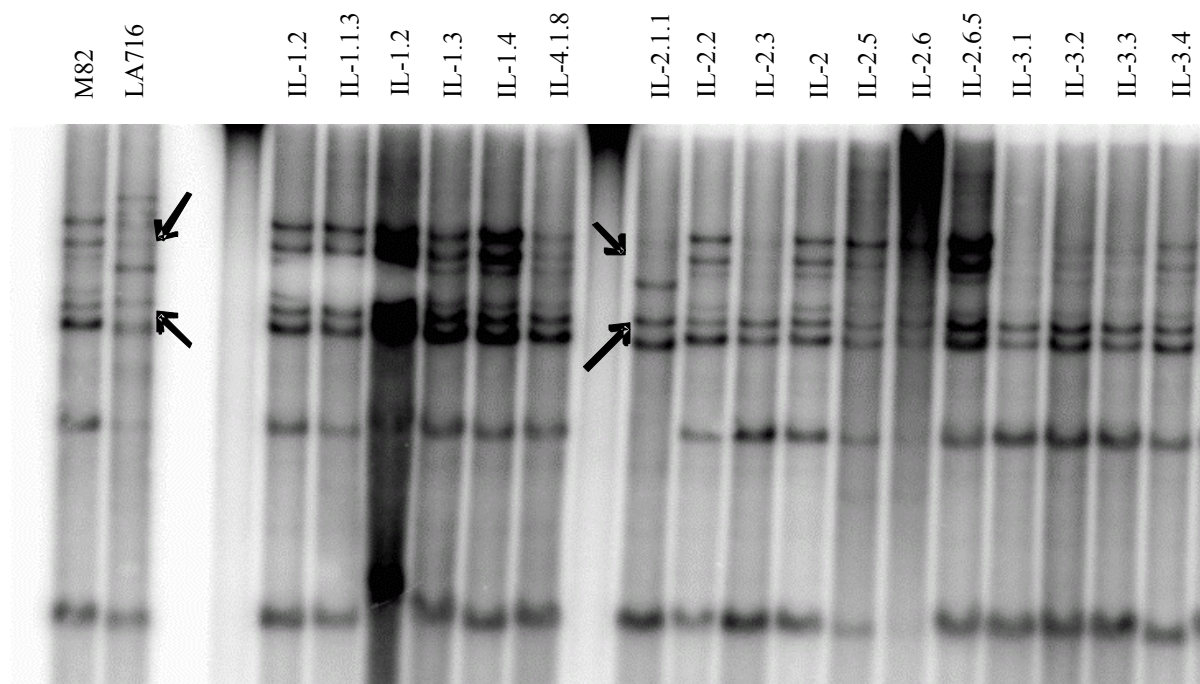


Figure 26. Résultat d'hybridation de la sonde SGN-U215128. Les deux parents sont à gauche et le polymorphisme d'intérêt est présenté avec des flèches.

chromosome II que l'introgression IL-2.1. En conséquence la région homéologue du contig S<sub>H3</sub> est localisée à l'extrémité du chromosome T2 de la tomate.

Nous avons ainsi démontré qu'il était possible de réaliser la cartographie génétique du contig S<sub>H3</sub> en utilisant des sondes orthologues sur la population de lignées *S. lycopersicum* introgressé par *S. pennellii*.

#### **10.4 Transfert des marqueurs tomate chez le caféier**

Dans le but de confirmer la position sur l'extrémité du chromosome II de *S. lycopersicum*, nous avons entrepris de cartographier génétiquement des marqueurs de cette région de *S. lycopersicum* sur le caféier. Pour ce faire des séquences orthologues du caféier ont été recherchées à partir des séquences de marqueurs de *S. lycopersicum*. L'absence de polymorphisme lors de la cartographie sur le caféier ne nous a pas permis pour l'instant de valider la cartographie génétique chez *S. lycopersicum*, s'ajoute à cela le nombre limité des marqueurs de la tomate dans cette région.

Nous avons démontré dans cette étude qu'il est possible de localiser une région caféier chez la tomate à partir de sonde orthologues. Plusieurs gènes de résistance sont identifiés dans la région tomate 2.1.1 du chromosome II (Pan *et al.* 2000), nous permettant d'envisager une approche gène-candidat sur le caféier.



## CHAPITRE V: Discussion générale

## 11 SYNTHÈSE DES PRINCIPAUX RESULTATS

Le développement des connaissances sur les ressources génétiques et sur les résistances est indispensable dans l'élaboration de stratégies d'amélioration d'une plante face à une maladie. L'interaction caféier/*Hemileia vastatrix* n'échappe pas à cette règle. Nous avons souhaité développer plusieurs approches dans le but de répondre à plusieurs besoins concernant le couple caféier/*H. vastatrix*. Les ressources génétiques connues pour l'amélioration de *Coffea arabica* à la rouille orangée sont peu diverses et montrent aujourd'hui leurs limites. Nous nous sommes intéressés dans un premier temps à des zones de sympatrie entre *Coffea arabica* et *Coffea canephora* dans le but d'y découvrir des hybrides interspécifiques ayant un bon potentiel de résistance à la rouille. La première étude a consisté en la caractérisation de ces hybrides et de leur niveau de résistance à la rouille.

Parallèlement à l'étude du niveau de résistance des hybrides interspécifiques de Nouvelle-Calédonie pour l'amélioration de l'arabica pour sa résistance à la rouille, le développement de connaissances relatives à un locus de résistance à la rouille a été pris en compte. Pas moins de neuf facteurs de résistance sont identifiés pour la rouille du caféier. Dans le but à terme de caractériser la nature et la diversité des gènes de résistance et de développer des marqueurs pour une sélection assistée par marqueur, la caractérisation de l'un d'eux,  $S_H3$  a été entreprise. Des approches de génétique comparative avec *Arabidopsis thaliana* et *Solanum lycopersicum* ont été développées pour accélérer la cartographie physique de cette région chromosomique.

### 11.1 Etude de la genèse des hybrides et des flux de gène entre *Coffea arabica* et *Coffea canephora* à l'intérieur de zones sympatriques en Nouvelle Calédonie

Des missions de prospection en Nouvelle-Calédonie réalisées dans le but de découvrir du matériel génétique original, ont indiqué plusieurs secteurs de sympatrie entre *C. canephora* et *C. arabica* (Charmetant et Le Pierrès 1991; Le Pierrès 1999). Rappelons que la première espèce est diploïde ( $2n=2x=22$ ) et que la seconde, allotétraploïde ( $2n=4x=44$ ), est formée de deux génomes ancestraux, provenant de *C. eugenioides* (génome E) et de *C. canephora* (génome C), (Lashermes *et al.* 1999), le génome de *C. arabica* étant noté EaCa. Des plantes montrant des phénotypes intermédiaires entre *C. canephora* et *C. arabica* sont

aisément observables dans ces différents secteurs de sympatrie de l'île. Nous avons étudié les plantes ayant des phénotypes intermédiaires entre *C. arabica* et *C. canephora* sur cinq sites de prospection. Nos objectifs étaient d'établir les processus régissant la formation des hybrides et l'introggression de matériel génétique en conditions naturelles.

### 11.1.1 Diversité et formation des hybrides

La première étape a concerné la confirmation de la nature hybride du matériel collecté dans cinq sites géographiquement séparés. Grâce aux marqueurs AFLP et SSR en plus de confirmer leur nature hybride, il a été démontré que les parents *C. canephora* impliqués dans les croisements interspécifiques étaient multiples, en raison d'une grande diversité génétique. Ajoutons que ce phénomène a été observé dans les cinq sites étudiés. La création d'hybrides interspécifiques naturels est considérée comme un événement rare, ceci en raison de l'habitat géographiquement distinct entre ces deux espèces et des périodes de floraison différentes (Baranski 1996). Nos résultats acquis au cours de ce travail font état d'une diversité d'hybrides interspécifiques uniques au monde.

L'originalité des hybrides présents ne se limite pas à la nature des parents. L'approche combinée des données moléculaires à partir de l'ADN nucléaire et chloroplastique et de la cytométrie de flux ont révélé des hybrides hétérogènes de première génération (F1) et de génération plus avancée (Post-F1) et la participation de différents types de gamètes. Les plantes possédant un jeu complet de marqueurs de *C. canephora* et *C. arabica* sont désignées comme hybrides de première génération. Les hybrides de génération plus avancés nommés post-F1 n'ont plus 100% des allèles aux loci étudiés. Cette disparition peut être expliquée par la constitution génomique des F1 triploïdes. Pour les plantes tétraploïdes de première génération, une méiose bien régulée permettrait de ne pas observer de disparition dans les générations avancées, toutefois les études menées par Herrera *et al.* (2002a; b) et la présence d'appariements non régulés chez les hybrides de première génération tétraploïde (Lashermes *et al.* 2000) sont en faveur d'une ségrégation des marqueurs.

L'étude s'est portée dans un premier temps sur à la genèse des hybrides de première génération, ils sont aussi bien triploïdes que tétraploïdes provenant de croisements ayant comme parent femelle *C. arabica* ou *C. canephora*. Comme attendu (Le Pierrès 1995; Herrera *et al.* 2002b), les plantes triploïdes F1 sont le résultat du croisement de gamètes à

2n de *C. arabica* et de gamètes à n de *C. canephora*. Les plantes tétraploïdes quant à elles sont susceptibles de résulter de gamètes non réduits provenant de *C. canephora* et de gamètes de *C. arabica*. Ces gamètes non réduits sont le résultat d'une première ou d'une deuxième division méiotique inachevée (FDR ou SDR, respectivement). Les réductions méiotiques inachevées ont pour conséquence une non séparation des cellules pendant la méiose, ce qui produit des gamètes à 2n contenant les deux chromatides sœurs (SDR) ou non-sœurs (FDR) (Hermsen 1984). Nous avons remarqué que la SDR affecte aussi bien la micro- que la mégasporogénèse des parents *C. canephora*. La température fraîche des nuits et plus particulièrement les variations importantes de température entre le jour et la nuit dans les montagnes centrales de Nouvelle-Calédonie sont probablement à l'origine de cet effet. Plusieurs chercheurs ont en effet trouvé que cette production est stimulée par des facteurs environnementaux tels que la température, la blessure, le stress hydrique ou nutritionnel (Ramsey et Schemske 1998; Lanaud 1979).

En ce qui concerne les générations plus avancées, ou post-F1, la capacité de distinguer parmi les différents types de croisements est limitée en raison de l'hétérogénéité des hybrides détectés, en terme de niveau de ploïdie, et du fait que la distribution théorique des fréquences des marqueurs pour les différents croisements possibles en condition naturelle se recouvre. Nous avons pu établir quelques conclusions malgré cette limite inhérente au matériel génétique. Chez les hybrides de génération avancée (après la F1) une disparition des allèles de *C. canephora* et *C. arabica* est attendue pour les différents loci observés. Nous avons observé grâce aux marqueurs moléculaires que toutes les plantes hybrides post-F1 conservent l'ensemble des allèles et marqueurs de *C. arabica*. L'absence apparente de ségrégation peut être expliquée par des événements d'hybridation en retour avec *C. arabica*. Deux raisons nous poussent à favoriser l'hypothèse d'un retour sur *C. arabica*. La première est due à la nature allogame de *C. canephora*. En effet, celui-ci est considéré comme porteur d'un niveau élevé d'allèles récessifs délétères (Lashermes *et al.* 1994). Ainsi, le croisement en retour sur *C. arabica* a plus de chance d'être viable grâce à sa capacité tampon contre les effets des allèles délétères. Cela a été déjà démontré dans plusieurs espèces de plante, telle que la luzerne et les oléagineux (Sharpe *et al.* 1995; Brouwer et Osborn 1999). La seconde raison est liée au système d'incompatibilité gamétophytique présent chez *C. canephora* qui peut être une limite pour les croisements en retour avec *C. canephora*. Nous supposons que ces hybrides post-F1 dérivent d'une ou de deux générations de croisement après l'événement F1. Cette hypothèse est plausible

chronologiquement : la date du premier rapport de présence des hybrides se situe dans les années 40, et le temps assez long (10-20 ans) exigé avant la première floraison du caféier en conditions naturelles (Berthaud 1986).

### 11.1.2 Les flux de gènes

Nous avons montré que des flux de gènes, et plus particulièrement une hybridation introgressive (c-à-d l'incorporation dans le patrimoine génétique d'une population de(s) gène(s) d'espèces différentes), a lieu dans ces zones de sympatrie. Les flux de gène sont informatifs à plus d'un titre. Ces flux indiquent un chevauchement significatif entre les périodes de floraison des deux espèces et des hybrides interspécifiques. La présence d'hybrides interspécifiques entre une espèce diploïde et une espèce polyploïde apparentée indique que les barrières génétiques post-zygotiques ne sont pas infranchissables dans les conditions environnementales de Nouvelle-Calédonie. En conséquence, ces deux espèces (*C. arabica* et *C. canephora*) pourraient être considérées "génétiquement isolées" plutôt que "reproductivement isolées", ce qui est la notion d'espèces biologiques définie par Rieseberg et Carney (1998). Ces données permettent de formuler des hypothèses sur la formation de l'espèce *C. arabica*. Par exemple, l'importance des flux de gènes entre *C. arabica* et son parent *C. canephora* permet d'envisager une phase de coexistence stable entre *C. arabica* et les espèces diploïdes ancestrales, appuyant l'idée d'un processus de spéciation graduel. Ce principe a été avancé pour la formation d'espèces autotétraploïdes en milieu diploïde chez l'espèce *Chamerion angustifolium* par Husband *et al.* (2004).

### 11.1.3 Nouvelle diversité d'hybrides interspécifiques pour l'amélioration de *Coffea arabica* à la rouille.

Les variétés cultivées de *C. arabica* (allotétraploïde) présentent une très faible diversité génétique (Anthony *et al.* 2002a; Steiger *et al.* 2002; Sera *et al.* 2003; Chaparro *et al.* 2004). Cette caractéristique rend la culture de *C. arabica* particulièrement vulnérable aux aléas. L'amélioration génétique de l'espèce *C. arabica* est basée principalement sur l'exploitation des ressources génétiques des caféiers diploïdes. Elle consiste à introduire plus de variabilité génétique au sein des variétés cultivées de l'espèce par introgression d'ADN des espèces diploïdes (par exemple de *C. canephora*). La grande diversité des hybrides interspécifiques collectés en Nouvelle-Calédonie (Mahé *et al.* accepté) représente

donc un fort potentiel pour la recherche de nouvelles sources de résistance pour l'amélioration de *C. arabica*.

Notre étude s'est portée sur l'affection fongique la plus grave de *C. arabica* cultivé, notamment en Nouvelle-Calédonie, causée par l'agent pathogène *H. vastatrix*. La recherche de plants résistants pour l'amélioration de *C. arabica* aux attaques à *H. vastatrix* remonte à 1875. Même si quelques rares hybrides ont été découverts ou créés artificiellement, les descendances d'un seul hybride spontané, l'hybride du Timor, sont aujourd'hui à la base de la plupart des programmes d'amélioration des arabicas pour la rouille. Ces descendances furent décrites comme très résistantes à toutes les races de rouille. Il a été rapporté récemment, que de nouvelles races de rouille très virulentes ont contourné l'hybride du Timor. Aussi, la caractérisation de nouvelles sources de diversité génétique pour l'amélioration de *C. arabica* pour sa résistance à la rouille est une priorité.

Afin d'évaluer le spectre de résistance à *H. vastatrix* des hybrides interspécifiques de Nouvelle-Calédonie, nous avons sélectionné un échantillon de quinze plantes sur les 42 de l'étude initiale sur la base de leur divergence allélique et de leur différence de niveau d'introgression (de 4 à 100% de marqueurs canephora présents). Les inoculations des plantes F1 et post-F1 se sont effectuées en deux temps par le CIFC. Dans un premier temps un panel d'isolats représentatifs de la diversité des facteurs de résistance a été retenu, dans le but de déterminer les groupes physiologiques auxquels appartiennent les caféiers. Ces groupes physiologiques sont représentatifs du ou des gènes de résistance présents chez les caféiers testés. Dans un second temps des isolats contournant les dérivés de l'hybride du Timor ont aussi été utilisés. Ceci dans le but de caractériser le potentiel de résistance de ces hybrides interspécifiques en comparaison avec les descendances de l'hybride du Timor.

A l'exception de quelques individus, les plantes issues des croisements interspécifiques se sont avérées résistantes aux différentes races de rouille choisies, y compris celles contournant les descendants de l'hybride du Timor. Ces hybrides interspécifiques naturels pourraient donc servir de relais pour le transfert de caractères agronomiques par rétro-croisements avec *C. arabica*.

Notre étude a mis en valeur l'originalité des hybrides interspécifiques en Nouvelle-Calédonie à travers la diversité des croisements interspécifiques et l'importance de l'hybridation introgressive naturelle. Un programme de sélection généalogique par

rétro-croisements à partir de certains hybrides a donc été initié. Il est à noter que l'utilisation d'hybrides interspécifiques dont la fertilité est rétablie permettra de réduire considérablement le temps nécessaire pour l'obtention de lignées résistantes chez le caféier, qui aujourd'hui s'élève à approximativement 30 ans.

## **11.2 Etude du facteur de résistance $S_{H3}$ à *Hemileia vastatrix* agent de la rouille.**

Parallèlement à l'étude du potentiel des hybrides interspécifiques de Nouvelle-Calédonie pour l'amélioration de *C. arabica* pour sa résistance à la rouille, nous nous sommes intéressés au développement des connaissances relatives à un locus de résistance à la rouille.

Les études de la transmission de la résistance à la rouille chez le caféier ont montré que la théorie « gène pour gène » est applicable à l'interaction aux réactions de la rouille avec le caféier (Silva *et al.* 2006) et pour lesquelles neuf facteurs de résistance sont identifiés. *Coffea arabica* porte les facteurs  $S_{H1}$ ,  $S_{H2}$ ,  $S_{H4}$  et  $S_{H5}$ , *C. canephora* :  $S_{H6}$ ,  $S_{H7}$ ,  $S_{H8}$  et  $S_{H9}$  et *C. liberica*  $S_{H3}$  (Noronha-Wagner et Bettencourt 1967; Bettencourt et Noronha-Wagner 1971; Bettencourt *et al.* 1980). Afin de cartographier génétiquement les régions du génome portant le (les) caractère(s) de résistance à la rouille et de développer un programme d'amélioration combinant plusieurs locus de résistance, il était nécessaire de créer des populations en ségrégation pour la résistance à la rouille. Afin d'entreprendre ce travail, nous disposons de deux populations correspondant respectivement à :

- une population F2, *C. arabica* (cv. Matari) introgressé par *C. liberica*
- une population de cette dernière en retour sur *C. arabica* ( $BC_2$ )

Ces deux populations ont été testées vis-à-vis de deux races (race II et VIII) pour le  $BC_2$  et quatre races (race II, VIII, XVIII et XIX) de rouille pour la population F2, afin de suivre la ségrégation du facteur de résistance  $S_{H3}$  (Prakash *et al.* 2005).

### 11.2.1 Développement de marqueurs liés au locus $S_{H3}$

#### 11.2.1.1 Marqueurs SCAR dérivés des bandes AFLP

Nous avons, pour point de départ, un groupe de liaison de 6.3 cM composé de 21 marqueurs AFLP, associé à un locus majeur de résistance à la rouille chez *C. liberica*  $S_{H3}$ . Le groupe de liaison a été développé par l'analyse de lignées d'arabica introgressées par *C. liberica* résistantes ou sensibles, et l'analyse de 144 individus F2 en ségrégation pour la résistance (F2 : Matari x S288) (Prakash *et al.* 2005). Nous avons développé quatre marqueurs SCAR à partir des marqueurs AFLP. L'association de ces marqueurs à la résistance à *H. vastatrix* a été confirmée d'une part à partir d'un lot de 101 lignées d'introgression arabica (S et R), et d'autre part par le criblage d'une nouvelle population d'individus BC<sub>2</sub> ségrégeant pour le facteur de résistance  $S_{H3}$  pour un des marqueurs SCAR (Sequence Characterised Amplified Region). Au total, 150 individus ont été génotypés et évalués pour cette résistance.

#### 11.2.1.2 Marqueurs SSR dérivés de la carte physique de *C. canephora*

Trois marqueurs SCAR ont été cartographiés sur le groupe de liaison 1 (chromosome 1) de la carte génétique de *C. canephora* construite à partir du clone IF200 (Lashermes *et al.* 2001). Les microsatellites cartographiés dans cette région chromosomique ont été testés sur la population F2 et BC2. L'utilisation de trois microsatellites a permis de caractériser un polymorphisme de présence, deux sur Matari et un sur S288.

#### 11.2.1.3 Développement de marqueurs des extrémités BAC du contig $S_{H3}$

Dix-huit amorces ont été développées pour des extrémités de BAC du contig  $S_{H3}$ . La possibilité de ces marqueurs à révéler du polymorphisme de présence ou de taille sur les population F2 et BC2 a été testée. Trois extrémités de fin de BAC ont été cartographiés sur le groupe de liaison portant le ou les gènes  $S_{H3}$ .

Dans le contexte de la sélection de variétés de caféiers résistantes à la race III d'*H. vastatrix*, plusieurs marqueurs d'amplification proche de  $S_{H3}$  offrent la possibilité d'une sélection assistée par marqueurs. Par rapport aux marqueurs AFLP, les marqueurs PCR développés ont gagné en facilité d'application.



### 11.2.2 Cartographie génétique et physique d'un locus de résistance à *Hemileia vastatrix*

Nous disposons d'une banque BAC provenant de *C. arabica* pour réaliser la cartographie physique de la région chromosomique homéologue à celle contenant S<sub>H3</sub> chez *C. liberica* et développer une approche de clonage positionnel. Nous avons donc entrepris de cartographier la région chromosomique homéologue *C. arabica* avec l'objectif de développer une approche 'gène-candidat' sur *C. liberica*. Plusieurs éléments tendent en effet à montrer que les gènes de résistance pourraient être conservés sur les régions chromosomiques homéologues. Premièrement, il a été montré que les locus de résistance de maladie des plantes ayant des temps de divergence similaire au caféier (3 à 5 millions d'années) semblent être positionnellement bien conservés (Grube *et al.* 2000; Wikstrom *et al.* 2001). Deuxièmement, une bonne colinéarité des régions - codantes ou non- a été observée entre *C. liberica*, *C. arabica* et *C. canephora*.

Le développement de la carte physique de la région chromosomique S<sub>H3</sub> a été initié par le SCAR le plus proche du locus de résistance. Ce marqueur a permis le criblage de la banque BAC IAPAR 59 (Noir *et al.* 2004). Vingt deux clones BAC assemblés en un contig autour du marqueur SCAR M8 ont été développés. La comparaison de la carte physique S<sub>H3</sub> avec *Arabidopsis* a permis le transfert de plusieurs marqueurs et l'extension de la carte physique à 35 BACs (Cf Figure 23). Nous avons développé les deux bornes encadrant la région chromosomique S<sub>H3</sub> et la cartographie génétique du locus S<sub>H3</sub> permet maintenant le positionnement du (ou des) gène(s) correspondant(s) dans un intervalle de moins de 2 cM. L'état d'avancée de la carte physique de ce même locus permet de donner une estimation maximum de la distance physique entre les marqueurs génétiques les plus proches de ce locus d'environ 2 mégabases. Cette valeur est maximale car elle ne prend pas en compte le chevauchement des BACs. La réalisation du fingerprint des BACs de cette région chromosomique permettra de préciser la taille exacte de cet intervalle.

La construction des contig de BAC représente une étape fondamentale pour le clonage positionnel des gènes de résistance.

### 11.3 Développement de deux stratégies de génétique comparative

L'exploitation la plus fréquente de la génétique comparative est l'identification de marqueurs pour une région génomique donnée, ce qui est particulièrement utile pour des expériences de cartographie fine et de clonage positionnel. Plus les espèces sont proches, meilleur sera le transfert de marqueurs. Au plan évolutif la divergence entre les genres *Coffea* et *Arabidopsis* se situe aux alentours de 94 millions d'années (Crepet *et al.* 2004). Celle entre les Solanacées et Rubiacées (asterid I) est d'approximativement 50 millions d'années (Gandolfo *et al.* 1998; Crepet *et al.* 2004). Chacune de ces deux espèces (*S. lycopersicum* et *A. thaliana*) représentent un avantage dans notre stratégie de transfert de marqueurs pour l'amélioration de la cartographie fine de la région S<sub>H3</sub>. L'utilisation d'*Arabidopsis* est avantageuse car la connaissance de son génome fournit un catalogue de gènes applicables à l'analyse comparative avec d'autres espèces. Par comparaison, le séquençage du génome de la tomate n'est pas encore achevé. En revanche plusieurs familles de gène de résistance ont été localisées et bien documentées chez les Solanacées (Pan *et al.* 2000; Grube *et al.* 2000). Ainsi, en raison de sa plus grande proximité avec le caféier, une approche gène-candidat pourrait être envisagée à partir des ressources génomiques des Solanacées.

#### 11.3.1 Détermination de la microsyténie entre *Arabidopsis* et le caféier.

##### 11.3.1.1 Amélioration de la carte physique

Les séquences des extrémités de BAC du contig associé à S<sub>H3</sub> nous ont permis d'initier notre étude pour laquelle nous avons développé une stratégie originale basée sur deux étapes.

- localisation *in silico* des séquences orthologues présumées d'*Arabidopsis* aux extrémités des séquences de BAC du caféier appartenant au contig S<sub>H3</sub>. Quatre fragments d'*Arabidopsis* ont été identifiés, qui se sont révélés être synténiques entre eux.
- les séquences EST du caféier supposées orthologues aux séquences uniques ou dupliquées, présentes sur les fragments candidats d'*Arabidopsis*, ont été identifiées et

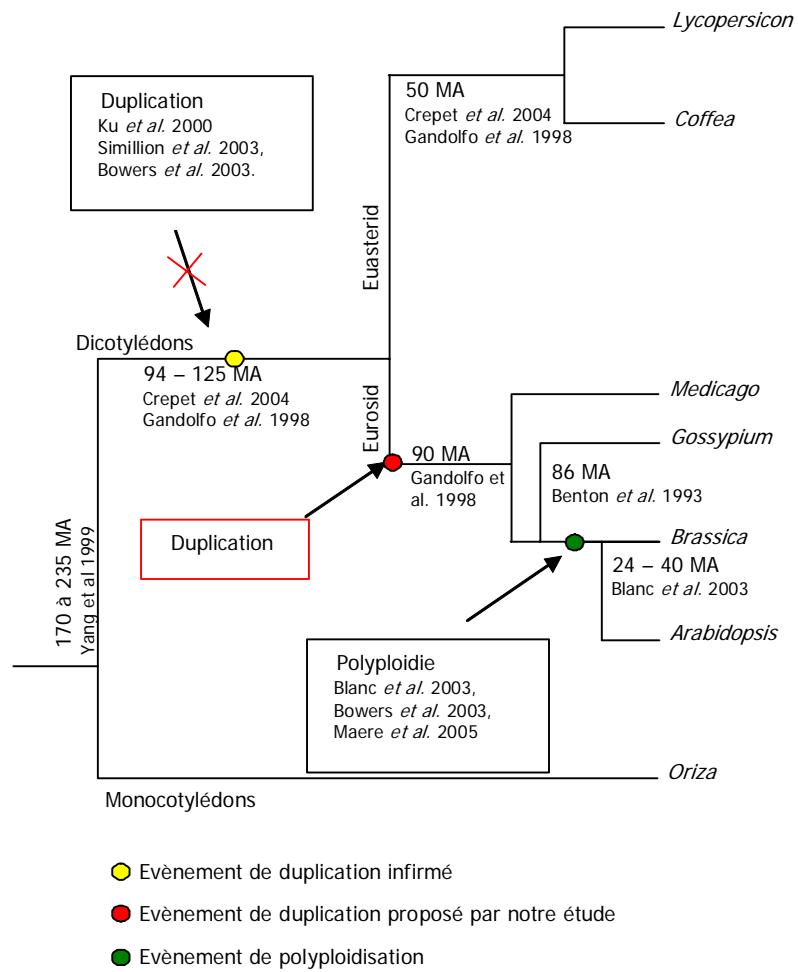


Figure 27. Evénements de duplication entre le groupe Euasterid et Eurosoid.

employées pour générer des marqueurs COS (« Conserved Orthologous Sequence ») utilisés dans le criblage d'une bibliothèque de BAC du caféier.

Cette démarche d'ancrage de séquences orthologues conservées sur le caféier provenant des quatre segments synténiques d'*Arabidopsis* a permis d'établir une microsyténie *Arabidopsis*-caféier. Ce faisant, nous avons consolidé et étendu la cartographie physique de la région chromosomique S<sub>H3</sub>, nous permettant ainsi de générer rapidement des marqueurs cartographiés sur le groupe liaison S<sub>H3</sub> de *C. liberica*, et de démontrer l'intérêt des marqueurs COS. Un nouveau marqueur est primordial pour la cartographie physique car il permet en combinaison avec un SCAR dérivé des marqueurs AFLP l'encadrement physique de la région chromosomique portant le (ou les) gène(s) de résistance. Nous avons ensuite prouvé que les duplications présentent chez *Arabidopsis*, souvent décrites comme une limite à la génétique comparative, peuvent permettre le développement de marqueurs moléculaires et *Arabidopsis* est apparue comme une plante modèle utile pour l'amélioration de la cartographie physique de la région chromosomique caféier liée à S<sub>H3</sub>.

L'approche de génomique comparative avec *Arabidopsis* peut être recommandée aux espèces pour lesquelles les séquences génomiques ne sont pas encore disponibles, mais ayant des ressources d'EST importantes

#### 11.3.1.2 Implication sur l'évolution des espèces

Chez *Arabidopsis*, trois événements de duplication sont décrits, l'un, très ancien qui s'est produit au moment de la séparation des angiospermes et gymnospermes, le deuxième est situé au moment de la séparation entre mono- et dicotylédones, et le troisième intervenant avant la séparation de *Brassica* et *Arabidopsis* (Figure 27). L'étude de Blanc *et al.* (2003) avait permis d'identifier les fragments issus de la dernière duplication. Notre étude a permis d'identifier les fragments issus des deux dernières duplications (4 fragments en tout) (Figure 28).

Les marqueurs développés à partir des quatre fragments d'*Arabidopsis* sont localisés sur une seule région chromosomique de *C. arabica* (Figure 28), ce qui indique l'absence de duplication dans cette région chromosomique et notamment avant la séparation d'*Arabidopsis* et du caféier. Ce résultat suggère que le deuxième événement de duplication ait eu lieu après la séparation du groupe des Euasterid (Coffee) et Eurosid (*Arabidopsis*) (Figure 27) et non avant. Dans le but de confirmer (ou non) la position du deuxième

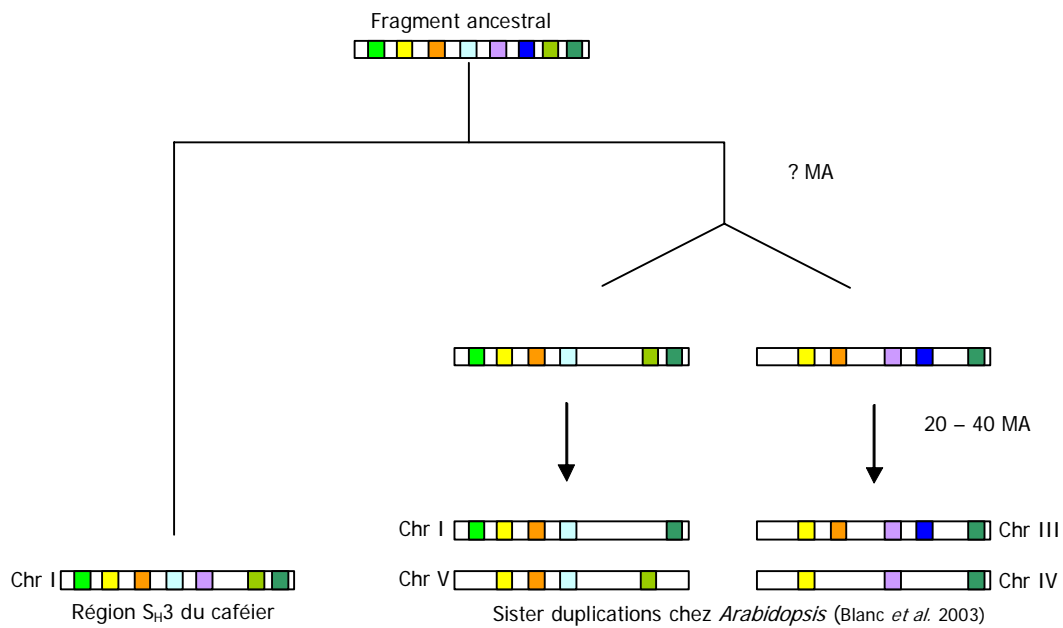


Figure 28. Représentation schématique des gènes présents dans les régions synténiques caféier - *Arabidopsis* et le fragment ancestral avant les événements de duplication. Les gènes présents sur *C. arabica* ont été révélés par hybridation des orthologues caféiers des gènes d'*Arabidopsis*, et ceux présents chez *Arabidopsis* proviennent du site TIGR (*Arabidopsis* duplication [http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis\\_genome\\_duplication.shtml](http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis_genome_duplication.shtml)). Les carrés de même couleur correspondent à des paralogues.

événement d'hybridation, nous nous proposons d'étendre cette étude à d'autres régions chromosomiques.

L'étude de la synténie a aussi montré que la région chromosomique étudiée du caféier était plus proche de la région homéologue ancestrale *Arabidopsis*/caféier qu'aucunes des régions chromosomiques d'*Arabidopsis* (Figure 28). Un résultat similaire a été obtenu en comparant la séquence d'un BAC de *S. lycopersicum* avec *Arabidopsis* (Ku *et al.* 2000), ce qui suggère que cette proximité soit une caractéristique du groupe des Euasterid I.

La comparaison des génomes entre *Arabidopsis* et le caféier nous permet d'envisager de nouveaux champs d'investigation en relation avec l'évolution des espèces.

### 11.3.2 Localisation de la région chromosomique homéologue S<sub>H</sub>3 chez *Solanum lycopersicum*.

Nous avons employé la méthode RFLP sur des lignées introgressées (ILs) pour notre démarche de cartographie génétique de la région chromosomique homéologue S<sub>H</sub>3 sur la tomate. Ces lignées introgressées comprenant 75 lignées quasi-isogéniques, sont issues du croisement interspécifique entre *L. pennellii* (LA716) et *L. esculentum* (M82) (Eshed et Zamir 1995). La population de plantes utilisées pour la cartographie est composée d'une série de plantes qui contiennent la séquence complète du génome de *L. pennellii* introgressée dans le génome de *L. esculentum* (Eshed et Zamir 1995; Liu et Zamir 1999). La carte génétique a été construite par l'analyse en Southern blot de 75 lignées quasi-isogéniques, digérées par différentes enzymes de restriction, en utilisant des sondes tomates RFLP standards. Les chromosomes sont divisés en 107 bins ou segments représentant des positions relatives d'intervalles chromosomiques introgressés et chevauchant entre lignées (Cf Figure 25). Des sondes tomate orthologues aux extrémités de BAC caféier ont été développées et hybridées sur des membranes comprenant les parents LA716 et M82 pour une recherche de polymorphisme dans un premier temps, puis sur les membranes représentant l'ensemble des introgressions de *L. pennellii*. La région chromosomique IL2-1.1 ou bin 2.1.1 a été localisée, cette région chromosomique correspond à l'extrémité du chromosome II de la tomate. Plusieurs gènes de résistance sont localisés dans cette région, dont une famille présente sur la région synténique *Arabidopsis*.

Le développement des ressources génomiques de la tomate nous permettra prochainement d'envisager le transfert de marqueurs de la tomate vers le caféier pour l'amélioration de la cartographie physique et génétique. De plus, la présence de plusieurs gènes de résistance nous permet d'envisager une approche gène-candidat sur la région chromosomique S<sub>H3</sub> de *C. arabica*.

## 12 PERSPECTIVES

Les connaissances relatives à la résistance aux parasites chez le caféier sont une priorité dans les programmes de recherche, étant donné la dimension économique que représente la culture du caféier et l'importance des pertes de production liées à la pression parasitaire. Les travaux réalisés pendant cette thèse ont d'une part accru les ressources génétiques pour les programmes d'amélioration de *C. arabica* et, d'autre part, permis le développement des outils moléculaires associés aux locus de résistance  $S_{H3}$ . De plus, des outils génomiques (carte génétique et contig de BAC pour la région chromosomique homéologue  $S_{H3}$ ) qui caractérisent plus spécifiquement cette région sont maîtrisés. La combinaison des travaux développés pendant la thèse permet d'envisager des perspectives tant appliquées que fondamentales pour une meilleure connaissance des gènes des caféiers, notamment ceux impliqués dans la résistance à la rouille.

### 12.1 Etude des facteurs affectant l'évolution de l'allotétraploïde *C. arabica* dans les régions sympatriques de Nouvelle Calédonie.

Nous avons démontré que les zones de sympatrie de Nouvelle-Calédonie où sont réunis en un même lieu *C. arabica* et un de ses ancêtres diploïdes, *C. canephora*, représentaient des essaims d'hybrides avec de forts flux de gènes et des hybridations introgressives. Plusieurs facteurs peuvent influencer l'évolution de ces régions de sympatrie. Nous souhaiterions approfondir nos connaissances sur l'évolution de ces hybrides dans la relation hôte/parasite avec la rouille.

L'environnement calédonien est extrêmement favorable à l'expansion de la rouille du caféier causé par *H. vastatrix* (Bettencourt et Rodrigues Jr. 1988). L'introgession de *C. canephora* pourrait présenter un avantage sélectif important pour les plantes hybrides. On pourrait prévoir une sélection naturelle favorable de marqueurs introgressés dus à un déséquilibre de liaison (LD) entre ces marqueurs et les locus affectant la résistance à la rouille. Des études sont sur cette question nécessaires, en particulier un grand nombre de plantes et de marqueurs doivent être considérées (Comai *et al.* 2004) et il doit être établi si la résistance observée à la rouille est due à un ou plusieurs locus. En outre, un suivi du devenir des plantes introgressées semble particulièrement intéressant et donnera une



occasion d'explorer des aspects importants de l'évolution et de la spéciation des polyploïdes soumis à une contrainte écologique forte telle que la rouille.

## **12.2 Vers une connaissance plus approfondie du locus $S_{H3}$**

A terme, la caractérisation du gène  $S_{H3}$  constitue une étape incontournable dans la volonté de mieux comprendre les mécanismes de résistance associés à ce gène. Ces données constitueront également une ouverture vers le mode de fonctionnement et l'identification de nouveaux gènes de résistance chez le caféier.

### 12.2.1 Affinement de la cartographie physique

La cartographie génétique du locus  $S_{H3}$  permet le positionnement du (ou des) gène(s) correspondant(s) dans un intervalle de moins de 2 cM. L'état d'avancée de la carte physique de ce même locus permet pour l'instant de donner une estimation de la distance physique entre les marqueurs génétiques les plus proches de ce locus d'environ 2 mégabases. Le fingerprint des BACs de la région chromosomique est nécessaire afin de confirmer cette estimation. Dans l'optique d'améliorer la cartographie physique, de nouvelles extrémités de BAC du contig  $S_{H3}$  seront transférées sur la carte génétique de groupe de liaison de *C. liberica*.

Les stratégies de clonage positionnel de gènes nécessitent la mise en évidence de marqueurs génétiques physiquement proches du gène d'intérêt (Tanksley *et al.* 1995; Patocchi *et al.* 1999). Or, il a été montré, chez différentes espèces végétales, que le niveau de recombinaison génétique est loin d'être homogène sur l'ensemble d'un génome (Werner *et al.* 1992). Ces taux de recombinaison sont fonction, notamment, du niveau de ploïdie de l'espèce considérée, de la structure (par exemple euchromatine ou hétérochromatine) et de la région chromosomique (par exemple centromérique, télomérique) étudiée (Moore, 2000). Plus particulièrement, différents exemples font références à des phénomènes de répression de la recombinaison notamment au sein d'hybrides interspécifiques et de fragments introgressés, impliquant une forte sous estimation du *ratio* distance génétique/distance physique (cM/kb) (Van Daelen *et al.* 1993; Wei *et al.* 1999; Chin *et al.* 2001).

La région chromosomique portant le locus  $S_{H3}$  pourrait être le siège de restriction de recombinaison. Toutefois, l'analyse des relations génétiques entre lignées dérivées de l'hybride de Timor (Lashermes *et al.* 2000; Noir *et al.* 2003), l'estimation des fréquences de recombinaison au sein d'hybrides tétraploïdes (*C. arabica* x *C. canephora* 4x) (Herrera *et al.* 2002a), et l'introgression homogène de *C. canephora* dans les hybrides sauvages de *C. arabica* montrent que les phénomènes d'introgression et les recombinaisons entre chromosomes homéologues ne sont pas limités au sein de ce matériel végétal. Ces résultats sont sans doute à associer à la faible différenciation génomique observée entre les sous-génomes Ca et Ea de *C. arabica*, et entre ses sous-génomes et le génome de *C. canephora*. La région chromosomique portant la résistance  $S_{H3}$  provient de *C. liberica*, il est possible que les phénomènes d'introgression et les recombinaison soient limités. Dans la perspective de cloner le gène  $S_{H3}$  nous devons préparer des effectifs de populations plus importantes pour affiner la cartographie génétique et physique et avoir des marqueurs plus proches du ou des gènes candidats.

#### 12.2.2 Clonage du ou des gènes de résistance au locus homéologue à $S_{H3}$ chez *C. arabica*

Une approche gène-candidat sera développée en parallèle, notamment par l'identification de gène de résistance analogue (RGA). La marche chromosomique combinée à la localisation des BACs portant des motifs RGA sera développée pour le clonage du gène de résistance (Bendahmane *et al.* 1999; Van der Vossen *et al.* 2000; Ballvorra *et al.* 2002).

La région chromosomique d'*Arabidopsis* synténique à la région chromosomique du caféier possède plusieurs gènes de résistance. Nous avons eu l'opportunité de donner un BAC à séquencer; sur la base de cette synténie nous avons choisi le BAC synténique de la région chromosomique *Arabidopsis* portant des gènes de résistance. Ces résultats nous permettront d'approfondir nos connaissances sur la synténie *Arabidopsis*/caféier avec la perspective d'avoir des candidats de gènes de résistance.

A terme, l'étude comparative entre les séquences d'un locus de résistance à *H. vastatrix*, issues de caféiers sensibles et résistants à *H. vastatrix*, permettra de mettre en évidence les variations moléculaires liées aux caractères résistants ou sensibles du génotype considéré. Nous pourrions envisager de déterminer les formes alléliques qui pourraient avoir un coût adaptatif important sur l'agent pathogène et permettre de réduire son agressivité.

### 12.2.3 Etude des locus de résistance à la rouille au sein des caféiers.

Neuf facteurs de résistance ont été identifiés chez les caféiers. *Coffea arabica* porte les facteurs S<sub>H</sub>1, S<sub>H</sub>2, S<sub>H</sub>4 et S<sub>H</sub>5, *C. canephora* : S<sub>H</sub>6, S<sub>H</sub>7, S<sub>H</sub>8 et S<sub>H</sub>9 et *C. liberica* S<sub>H</sub>3 (Noronha-Wagner et Bettencourt 1967; Bettencourt et Noronha-Wagner 1971; Bettencourt *et al.* 1980). Dans le but de tester l'hypothèse d'une diversité allélique au même locus, les marqueurs liés au locus S<sub>H</sub>3 ont été testés sur une population (31 individus) en F4 inoculée avec un panel de races sauvages. Cette population a été sélectionnée car la moitié des plantes étaient sensibles et l'autre moitié résistantes ce qui indiquait une ségrégation simple d'un locus de résistance. Cette population issue de croisements interspécifiques naturels entre *C. arabica* et *C. canephora* de Nouvelle-Calédonie a montré une co-ségrégation entre la résistance à la rouille et un marqueur lié à S<sub>H</sub>3. Cette étude devra se poursuivre sur des effectifs plus importants et être développée sur d'autres croisements interspécifiques. Elle permet d'appuyer l'hypothèse d'une diversité allélique des facteurs de résistance à la rouille pour des régions chromosomiques homéologues entre espèces de caféier.

Des études génétiques et moléculaires ont révélé une organisation génomique particulière des gènes de résistance chez les plantes. La plupart d'entre eux sont génétiquement étroitement associée à d'autres gènes ou séquences homologues, et constituent des locus complexes ou clusters (Pryor et Ellis 1993). Nous envisageons que deux organisations des locus puissent expliquer les spécificités de résistance à la rouille et la colocalisation de régions de résistance chez des espèces différentes: soit un locus simple à multiples spécificités de résistance, soit de multiples copies de séquences analogues organisées en clusters. Le développement des connaissances pour la région chromosomique de résistance à la rouille nous permettra de répondre à cette question.

## 12.3 La génomique comparative avec le caféier

### 12.3.1 Génomique comparative de la région chromosomique S<sub>H</sub>3 avec d'autres espèces

La génomique comparative permet le transfert d'information d'une espèce à une autre dans la mesure où ces génomes ont une synténie suffisamment conservée. Nous avons démontré l'utilité des ressources connues d'*Arabidopsis thaliana* pour l'amélioration de la cartographie physique du caféier. L'importance des duplications présente chez *Arabidopsis*

n'en fait toutefois pas un organisme modèle pour la génétique comparative avec le caféier. Les différents évènements de duplication et l'évolution des séquences génomiques compliquent les alignements des génomes. Récemment, le génome du peuplier a été entièrement séquencé (Tuksan *et al.* 2006), des duplications anciennes et récentes sont rapportées. Toutefois il apparaît, probablement en raison de la nature pérenne du peuplier, que l'évolution de son génome en terme de perte de gène réarrangement et duplication en tandem procède plus lentement que chez *Arabidopsis* (Tuksan *et al.* 2006). Nous nous proposons de comparer la région chromosomique S<sub>H3</sub> à celle du peuplier, ceci dans le but d'évaluer la synténie caféier/peuplier et d'en dériver de nouveaux marqueurs. Le résultat du séquençage du BAC caféier nous permettra en plus de comparer l'ordre et la présence des gènes au sein de ce BAC en comparaison avec *Arabidopsis* et le peuplier. La même démarche sera réalisée lorsque le séquençage de la tomate sera achevé, en particulier le séquençage du chromosome T2.

### 12.3.2 Evolution des génomes

L'analyse des substitutions nucléotidiques (synonymes et ou non synonymes) chez *Arabidopsis* a permis de préciser les temps d'évolution entre les paires de gènes étudiées (Blanc *et al.* 2003). Les régions chromosomiques I et V, d'une part et III et IV d'autre part sont issue d'un même événement de duplication il y a 20 à 40 millions d'années. Pour la première fois nous avons confirmation que les gènes paralogues putatifs situés sur les trois ou quatre fragments synténiques d'*Arabidopsis* sont bien issus d'un ancêtre commun. Nous sommes en mesure d'étudier l'évolution des gènes en fonction du nombre de duplication au sein d'*Arabidopsis*. La détermination de la séquence orthologue chez le caféier permettra l'évaluation de la séparation entre *Arabidopsis* et le caféier.

## Références bibliographiques

**Acarkan A., Rossberg M., Koch M. and Schmidt R.** (2000). Comparative genome analysis reveals extensive conservation of genome organisation for *Arabidopsis thaliana* and *Capsella rubella*. *The Plant Journal*, **23**: 55-62.

**Adams K.L. and Wendel J.F.** (2005). Allele-Specific, Bidirectional Silencing of an Alcohol Dehydrogenase Gene in Different Organs of Interspecific Diploid Cotton Hybrids. *Genetics*, **171**: 2139-2142.

**Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A. and Galle R.F.** (2000). The genome sequence of *Drosophila melanogaster*. *Science*, **287**: 2185-2195.

**Agwanda C., Lashermes P., Trouslot P., Combes M.-C. and Charrier A.** (1997). Identification of RAPD markers for resistance to Coffee Berry Disease, *Colletotrichum kahawae*, in Arabica coffee. *Euphytica*, **97**: 241-248.

**Ahmad J. and Vishveshwara S.** (1980). *Coffea liberica* Bull. ex Hiern: a review. *Indian Coffee*, **44**: 29-36.

**Alves M.J., Coelho M.M., Prospero M.I. and Collares-Pereira M.J.** (1999). Production of fertile unreduced sperm by hybrid males of the *Rutilus alburnoides* complex (*Teleostei, Cyprinidae*). An alternative route to genome tetraploidization in unisexuals. *Genetics*, **151**: 277-83.

**Anthony F.** (1992). "Les ressources génétiques des caféiers." ORSTOM Editions.

**Anthony F. and Lashermes P.** (2005). Origin, evolution and diversity of coffee (*Coffea arabica* L.) genome. In: *Plant Genome, Biodiversity and Evolution*. Ed. by K.A Sharma and A. Sharma. Science Publishers Inc. Plymouth, UK. pp. 208-228.

**Anthony F., Astorga C. and Berthaud J.** (1999). Los recursos genéticos: las bases de una solución genética a los problemas de la caficultura latinoamericana. In "Desafíos de la caficultura centroamericana" (B. Bertrand, et B. Rapidel, Eds.), IICA-CIRAD, San José, Costa Rica, 369-406.

**Anthony F., Combes M.C., Astorga C., Bertrand B., Graziosi G. and Lashermes P.** (2002a). The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. *Theoretical and Applied Genetics*, **104**: 894-900.

**Anthony F., Topart P., Anzueto F., Astorga C. and Bertrand B.** (2002b). La resistencia genética del café (*Coffea* spp.) a los nematodos (*Meloidogyne* spp.): identificación y utilización para la caficultura latinoamericana. *Manejo integrado de plagas y Agroecología - CATIE*.

**Anzueto F., Bertrand B., Sarah J.L., Eskes A.B. and Decazy B.** (2001). Resistance to *Meloidogyne incognita* in Ethiopian *Coffea arabica* accessions. *Euphytica*, **118**: 1-8.

**Arnold M.L.** (1997). *Natural Hybridization and Evolution*. Oxford University Press, New York.

**Arnold M.L.** (2004). Transfer and Origin of Adaptations through Natural Hybridization : Were Anderson and Stebbins Right ? *The Plant Cell*, **16**: 562-570.

**Arnold M.L. and Hodges S.A.** (1995). Are natural hybrids fit or unfit relative to their parents? *Trends Ecology and Evolution*, **10**: 67-70.

**Arnold M.L., Shaw D.D. and Contreras, N.** (1987). Ribosomal RNA-encoding DNA introgression across a narrow hybrid zone between two subspecies of grasshopper. *Proceedings of the National Academy of Sciences (USA)*, **84**: 3946-3950.

**Avelino J. and Savary S.** (2002). Rational and optimized chemical control of coffee leaf rust (*Hemileia vastatrix*). *Integrated Control, in Plantations, recherche et développement*, pp. 141-143.

**Ballvora A., Ercolano M.R., Weis J., Meksem K., Bormann C.A., Oberhagemann P., Salamini F. and Gebhardt C.** (2002). The *R1* gene for potato resistance to late blight (*Phytophthora infestans*) belongs to the leucine zipper/NBS/LRR class of plant resistance genes. *The Plant Journal*, **30**: 361-371.

**Bancroft C.C., Chen Z., Dong G., Sunwoo J.B., Yeh N., Park C. and Van Waes, C.** (2001). Coexpression of proangiogenic factors IL-8 and VEGF by human head and neck squamous cell carcinoma involves coactivation by MEK-MAPK and IKK-NF- $\kappa$ B signal pathways. *Clinical Cancer Research*, **7**: 435-442.

**Baranski O.** (1996). Etude des caractères morphologiques, phénologiques et de la fertilité de 29 taxons de caféiers sauvages africains. *Rapport VSN, ORSTOM. Paris.* pp. 38.

**Barnes S.** (2002). Comparing *Arabidopsis* to other flowering plants. *Current Opinion in Plant Biology*, **5**: 128-134.

**Barth S., Melchinger A.E., Devezi-Savula B. and Lubberstedt T.** (2001). Influence of genetic background and heterozygosity on meiotic recombination in *Arabidopsis thaliana*. *Genome*, **44**: 971-978.

**Barton N. H. and Hewitt G. M.** (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**: 113-148.

**Beaumont M.A., Barratt E.M., Gottelli D., Kitchener A.C., Daniels M.J., Pritchard J.K., and Bruford M.W.** (2001). Genetic diversity and introgression in the Scottish wildcat. *Molecular Ecology*, **10**: 319-336.

**Bendahmane A., Kanyuka K. and Baulcombe D.C.** (1999). The Rx gene from potato controls separate virus resistance and cell responses. *Plant Cell*, **11**: 781-791.

**Bennetzen J.L. and Freeling M.** (1993). Grasses as a single genetic system: Genome composition, collinearity and compatibility. *Trends in Genetics*, **9**: 259-261.

**Bennetzen J.L. and Freeling M.** (1997). The unified grass genome: Synergy in synteny. *Genome Research*, **7**: 301-306.

**Berkeley M.J.** 1869. *Note. Gardens Chronicle*, **45**: 1157

**Berthaud J.** (1978). L'hybridation interspécifique entre *Coffea arabica* L. et *Coffea canephora* Pierre. Obtention et comparaison des hybrides triploïdes, Arabusta et hexaploïdes. *Café Cacao Thé*, **22**: 87-109.

- Berthaud J.** (1980). L'incompatibilité chez *Coffea canephora*: méthode de test et déterminisme génétique. *Café Cacao Thé*, **24**: 267-274.
- Berthaud J. and Charrier A.** (1988). Genetic resources of *Coffea*. In: Clarke RJ and Macrae R (Eds) *Coffee vol 4: Agronomy*. Elsevier Applied Science, London, pp 1-42
- Berthaud J. and Lourd M.** (1982). La résistance à *Hemelia vastatrix* des caféiers de l'espèce *Coffea canephora* de Côte d'Ivoire. Etude de la transmission de ce caractère par croisements contrôlés. *Garcia de Orta, Sér. Est. Agron.* **9**: 89-96.
- Berthaud J.** (1986). Les ressources génétiques pour l'amélioration des caféiers africains diploïdes. Collection "Travaux et Documents" ed ORSTOM, Paris, pp.188.
- Berthou F., Mathieu C. and Vedel F.** (1983). Chloroplast and mitochondrial DNA variation as indicator of phylogenetic relationship in genus *Coffea L.* *Theoretical and Applied Genetics*, **65**: 77-84.
- Bertrand B., Aguilar G., Santacreo R. and Anzueto F.** (1999). El mejoramiento genético en América Central. In "Desafios de la caficultura centroamericana" (Bertrand B., et Rapidel, B., Eds), IICA-CIRAD, San José, Costa Rica, pp. 405-456.
- Bertrand B., Anthony F. and Lashermes P.** (2001). Breeding for resistance to *Meloidogyne exigua* of *Coffea arabica* by introgression of resistance genes of *Coffea canephora*. *Plant Pathology*, **50**: 637-643.
- Bettencourt A.** (1973). Considerações gerais sobre o "Híbrido de Timor", Instituto Agronomico de Campinas, Campinas, Brasil, pp. 256.
- Bettencourt A.J. and Noronha-Wagner M.** (1971). Genetic factors conditioning resistance of *Coffea arabica L.* to *Hemileia vastatrix* Berk et Br. *Agronomia Lusitana*, **31**: 285-92.
- Bettencourt A.J. and Rodrigues C.J.** (1988). Principles and practice of coffee breeding for resistance to rust and other diseases. In: Clarke, R.J. & Macrae, R.(eds) *Coffee*, vol. 4 *Agronomy*. Elsevier Applied Science, London & New York, pp. 199-234.
- Bettencourt A.J., Rodrigues C.J. and Lopes J.** (1965). Study of the physiologic specialization of the coffee rust *H. vastatrix* and selection of coffee clones for the establishment of a standard range of differential hosts for this rust. In: Progress report 1960-65, pp.28-46, Coffee Rusts Research Center, Oeiras, Portugal.
- Bettencourt A.J., Noronha-Wagner M. and Lopes J.** (1980). Fator genético que condiciona a resistência do clone 1343/269 (Híbrido de Timor) à *Hemileia vastatrix* Berk. & Br. *Broteria Série Genética*, **1**: 53-8.
- Blanc G., Barakat A., Guyot R., Cooke R. and Delseny M.** (2000). Extensive duplication and reshuffling in the *Arabidopsis* Genome, *Plant Cell*, **12**: 1093-1101.
- Blanc G., Hokamp K. and Wolfe K.H.** (2003). A Recent Polyploidy Superimposed on Older Large-Scale Duplications in the *Arabidopsis* Genome. *Genome Research*, **13**: 137-144.



- Bock K.R.** (1962). Seasonal periodicity of coffee leaf rust and factors affecting the severity of outbreaks in Kenya colony. *Trans. Br. Mycol. Soc.* **45**: 289-300.
- Boecklen W.J., and Howard D.J.** (1997). Genetic Analysis of Hybrid Zones: Numbers of Markers and Power of Resolution. *Ecology*, **78**: 2611-2616.
- Bonierbale M.W., Plaisted R.L. and Tanksley S.D.** (1988). RFLP maps based on a common set clones reveal modes of chromosomal evolution in potato and tomato. *Genetics*, **120**: 1095-1103.
- Bowers J.E., Chapman B.A., Rong J. and Paterson A.H.** (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, **422**: 433-438.
- Bridson D.M.** (1987). Nomenclatural notes on *Psilanthus*, including *coffea* sect. *Paracoffea* (*Rubiaceae* tribe *Coffeae*). *Kew Bulletin*, **42**: 817-859.
- Bridson D.M. and Verdcourt B.** (1988). "Flora of tropical east Africa - *Rubiaceae*." (Polhill R. M., Ed) Part 2, 727 pp.
- Brouwer D.J. and Osborn T.C.** (1999). A molecular marker linkage map of tetraploid alfalfa *Medicago sativa* L. *Theoretical and Applied Genetics*, **99**: 1194-1200.
- Budiman M.A., Mao L., Wood T.C. and Wing R.A.** (2000). A deep-coverage tomato BAC library and prospectus toward development of an STC framework for genome sequencing. *Genome Research*, **10**: 129-136.
- Carels N. and Bernardi G.** (2000). Two classes of genes in plants. *Genetics*, **154**: 1819-1825.
- Carputo D., Monti L., Werner J.E., and Frusciante L.** (1999). Uses and usefulness of endosperm balance number. *Theoretical and Applied Genetics*, **98**: 478-484.
- Carson H.L., Johnson W.E., Nair P.S. and Sene F.M.** (1975). Genetic variation in Hawaiian *Drosophila*. III. Allozymic and chromosomal similarity in two *Drosophila* species. *Proceedings of the National Academy of Sciences (USA)*, **72**: 4521-4525.
- Carvalho A.** (1988). Principles and practices of coffee plant breeding for productivity and quality factors: *Coffea arabica*. In "Coffee vol 4: Agronomy", (Clarke R. J., et Macrae R., Eds), London, pp. 129-165.
- Carvalho A. and Krug C. A.** (1950), Genética de Coffea XIII -Hereditariedade da característica erecta em Coffea arabica L. *Bragantia*, **10**: 321-328.
- Carvalho A., Eskes A.B., Castillo-Z J., Sreenivasan M.S., Echeverri J.H., Fernandez C.E. and Fazuoli L.C.** (1989). Breeding programmes. In: Kushalappa, A.C. & Eskes, A.B. (eds) *Coffee Rust: Epidemiology, Resistance and Management*. CRC Press Inc., Boca Raton, Florida, pp. 293-335.
- Carvalho A. and Monaco L.C.** (1968). Relaciones geneticas de especies seleccionadas de Coffea. *Cafe*, Lima, **9**: 1-19.

- Cavell A.C., Lydiate D.J., Parkin I.A.P., Dean C. and Trick M.** (1998) Collinearity between a 30-centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. *Genome*, **41**: 62-69.
- Chalfoun S.M. and Silva S.M.** (1979). Avaliação de ferrugem de cafeeiro em relação às variáveis climáticas de algumas localidades do estado de Minas Gerais. *Congr. Bras. Pesq. Cafeeiras* **7**: 184.
- Chao S., Sharp P.J., Worland A.J et al.** (1989). RFLP-based genetic maps of wheat homeologous group 7 chromosomes. *Theoretical Applied Genetics*, **78**: 495-504.
- Chaparro A.P., Cristancho M.A., Cortina H.A. and Gaitan A.L.** (2004). Genetic variability of *Coffea arabica* L. accessions from Ethiopia evaluated with RAPDs. *Genetic Resources and Crop Evolution*, **51**: 291-297.
- Chapman B.A., Bowers J.E., Schulze S.R. and Paterson A.H.** (2004). A comparative phylogenetic approach for dating whole genome duplication events. *Bioinformatics*, **20**: 180-185.
- Charles J.** (1932). Le caféier. *Revue Agricole de la Chambre d'Agriculture de la Nouvelle-Calédonie*. Août-sept. **32**: 794-837.
- Charmetant P. and Le Pierrès D.** (1991). Rapport de mission en Nouvelle-Calédonie : prospection et collecte d'hybrides naturels de caféiers cultivés. IRCC-CIRAD / ORSTOM, pp. 21.
- Charrier A. and Berthaud J.** (1985). Botanical classification of coffee. In: *Coffee: Botany, biochemistry and production of beans and beverage*. Clifford M.N. et Willson K.C. (éd.). Croom Helm Ltd. UK. pp. 13-47.
- Charrier A. and Eskes A.B.** (1997). "Les caféiers" (Charrier A., Jacquot M., Hamon, S. et Nicolas D., Eds) Paris.
- Charrier A. and Vigreux M.** (1972). L'intercompatibilité des clones de caféiers cultivés sur la Côte est malgache. *Café, Cacao, Thé*, **16**: 111-122.
- Cheng S.V., Nadeau J.H., Tanzi R.E., Watkins P.C., Jagadesh J., Taylor B.A., Haines J.L., Sacchi N. and Gusella J.F.** (1988). Comparative mapping of DNA markers from the familial Alzheimer disease and Down syndrome regions of human chromosome 21 to mouse chromosomes 16 and 17. *Proceedings of the National Academy of Sciences (USA)*, **85**: 6032-6036.
- Chin D.B., Arroya-Garcia R., Ochoa O.E., Keselli R.V. and Lavelle D.O. et al.** (2001). Recombination and spontaneous mutation at the major cluster of resistance genes in lettuce (*Lactuca sativa*). *Genetics*, **157**: 831-849.
- Chinnappa C.C.** (1968). Interspecific hybrids of *Coffea canephora* and *Coffea arabica*. *Current Science*, **37**: 676-677
- Chowdhary B.P., Raudsepp T., Frönicke L. and Scherthan H.** (1998). Emerging Patterns of Comparative Genome Organization in Some Mammalian Species as Revealed by Zoo-FISH. *Genome Research*, **8**: 577-589.

**Comai L., Young K., Till B.J., Reynolds S.H., Greene E.A., Codomo C.A., Enns L.C., Johnson J., Burtner C., Oden A.R. and Henikoff S.** (2004). Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. *The Plant Journal*, **37**: 778-786.

**Combes M.-C., Andrzejewski S., Anthony F., Bertrand B., Rovelli P., Grazios G. and Lashermes P.** (2000). Characterisation of microsatellite loci in *Coffea arabica* and related coffee species. *Molecular Ecology*, **9**: 1178-1193.

**Cornu A., Farcy E. and Mousset C.** (1989). A genetic basis for the variation in meiotic recombination in *Petunia hybrida*. *Genome*, **32**: 46-53.

**Couturon E., Lashermes P. and Charrier A.** (1998). First intergeneric hybrids (*Psilanthus ebracteolatus* H. x *Coffea arabica*) in coffee trees. *Canadian Journal of Botany*, **76**: 542-546.

**Craig Venter J., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G., Smith H.O., Yandell M., Evans C.A., Holt R.A., et al.** (2001). The Sequence of the Human Genome. *Science*, **291**: 1304-1351.

**Cramer P.J.S.** (1957). A review of literature of coffee research in Indonesia. SIC Ed., IICA, Turrialba, Costa Rica. 262 p.

**Crepet W.L., Nixon K.C. and Gandolfo M.A.** (2004). Fossil evidence and phylogeny: the age of major angiosperm clades based on mesofossil and macrofossil evidence from Cretaceous deposits. *American Journal of Botany*, **91**: 1666-1682.

**Cros J., Combes M.-C., Chabrillange N., Duperray C., Monnot des angles A. and Hamon S.** (1995). Nuclear DNA content in the sugenus *Coffea* (Rubiaceae): inter- and intra-specific variation in African species. *Canadian Journal of Botany*, **73**: 14-20.

**Cros J., Combes M.-C., Trouslot P., Anthony F., Hamon S., Charrier A. and Lashermes P.** (1998). Phylogenetic analysis of chloroplast DNA variation in *Coffea* L. *Molecular Phylogenetics. Evolution*, **9**: 109-117.

**D'Oliveira B.** (1957). *Revista do Café Português*, **4**(16): 5-15.

**D'Oliveira B. and Rodrigues Jr C.J.** (1961). O problema das ferrugens do cafeeiro. *Review Café Português*, **8**: 5-50.

**Davis A.P., Maurin O., Chester M., Mvungu E.F. and Fay M.F.** (2006). Phylogenetic relationship in *Coffea* (Rubiaceae) inferred from sequence data and morphology. *Proceedings of the 21<sup>st</sup> international conference on coffee Science*, Montpellier, pp. 56.

**Devos K.M. and Gale M.D.** (2000). Genome relationships: the grass model in current research. *The Plant Cell*, **12**: 637-646.

**Devos K.M., Atkinson M.D., Chinoy C.N., Liu C.J. and Gale M.D.** (1993). RFLP-based genetic map of the homoeologous group 2 chromosomes of wheat, rye, and barley. *Theoretical and Applied Genetics*, **85**: 784-792.

- Doganlar S., Frary A., Daunay M.C., Lester R.N. and Tanksley S.D.** (2002). A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the solanaceae. *Genetics*, **161**: 1697-1711.
- Dolezel J., Binarova P. and Lucretti S.** (1989). Analysis of nuclear DNA content in plant cells by flow cytometry. *Biologia Plantarum*, **31** : 113-120.
- D'Oliveira A.B.** (1954-57). As ferrugens do cafeeiro. *Rev. Café Port.* 1(4) 5-13; 2(5):5-12; 2 (6):5-15; 2 (7): 9-17; e 2 (8): 5-22; 4 (16): 5-15.
- Dominguez I., Graziano E., Gebhardt C., Baraka A., Berr S., Arus P., Delseny M. and Barnes S.** (2003). Plant genome archaeology: evidence for conserved ancestral chromosome segments in dicotyledonous plant species. *Plant Biotechnology Journal*, **1**: 91-100.
- Dufour M., Anthony F., Bertrand B. and Eskes A.B.** (1997). Identification de caféiers mâles stériles de *Coffea arabica* au CATIE, Costa Rica. *Plantations, Recherche, Développement*, **4**: 401-407.
- Dunham I., Shimizu N., Roe B.A., Chisoe S., et al.** (1999). The DNA sequence of human chromosome 22. *Nature*, **402**: 489-496.
- Dussert P.** (1911). Semences de caféiers réfractaires à *Hemileia*. *Revue Agricole de la Chambre d'Agriculture de la Nouvelle-Calédonie*, **11**: 17-20.
- Echard G., Broad T.E., Hill D. and Pearce P.** (1994). Present status of ovine gene map (*Ovis aries*); comparison with the bovine map (*Bos Taurus*). *Mamm. Genome*, **5**: 324-332.
- Ehlenfeldt M.K. and Ortiz R.** (1995). Evidence on the nature and origins of endosperm dosage requirements in *Solanum* and other angiosperm genera. *Sexual Plant Reproduction*, **8**: 189-196.
- Ermolaeva M.D., Wu M., Eisen J.A. and Salzberg S.L.** (2003). The age of the *Arabidopsis thaliana* genome duplication. *Plant Molecular Biology*, **51**: 859-866.
- Eshed Y. and Zamir D.** (1995). An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTLs. *Genetics*, **141**: 1147-1162.
- Eskes A.B. and Toma-Braghini M.** (1982). Assessment methods for resistance to coffee leaf rust (*Hemileia vastatrix* Berk. and Br.). *Plant Protection Bulletin, FAO (Rome)*, **29**: 56-66.
- Eskes A.B.** (2005). Phenotypic expression of resistance to coffee leaf rust and its possible relationship with durability. *Durable Resistance to Coffee Leaf Rust, Chap 13*: 305-332.
- Etienne H., Anthony F., Dussert S., Fernandez D., Lashermes P. and Bertrand B.** (2002). Biotechnological applications for the improvement of coffee (*Coffea arabica*). *In Vitro Cellular and Development Biology - Plant*, **38**: 129-138.
- FAO** (1969) FAO coffee mission to Ethiopia 1964-1965. Collection report. Roma, Italy, FAO. 200p

- Feldman M.** (1993). Cytogenetic activity and mode of action of the pairing homoeologous (Ph1) gene of wheat. *Crop Science*, **33**: 894-897.
- Feldman M., Liu B., Segal G., Abbo S., Levy A.A. and Vega J.M.** (1997). Rapid elimination of low-copy DNA sequences in polyploid wheat: A possible mechanism for differentiation of homoeologous chromosomes. *Genetics*, **147**: 1381-1387.
- Fitch W.M.** (2000). Homology a personal view on some of the problems. *Trends in Genetics*, **16**: 227-231.
- Flor H.H.** (1971). Current status of the gene-for-gene concept. *Annual Review Phytopathology*, **9**: 275-296.
- Foote T., Roberts M., Kurata N., Sasaki T. and Moore G.** (1997). Detailed comparative mapping of cereal chromosome regions corresponding to the Ph1 locus in wheat. *Genetics*, **147**: 801-807.
- Foster-Hartnett D., Mudge J., Larsen D., Yan H., Denny R., Penuela S. and Young ND.** (2002). Comparative genomic analysis of sequences sampled from a small region on soybean (*Glycine max*) molecular linkage group. *Genome*, **45**: 634-645
- Fulton T.M., Van der Hoeven R., Eannetta N.T. and Tanksley S.D.** (2002). Identification, Analysis, and Utilization of Conserved Ortholog Set Markers for Comparative Genomics in higher plants. *The Plant Cell Online*, **14**: 1457-1467.
- Futuyma D.J.** (1998). *Evolutionary Biology*, Third Edition, Sinauer Associates.
- Gale M.D. and Devos K.M.** (1998). Comparative genetics in the grasses. *Proceedings of the National Academy of Sciences (USA)*, **95**: 1971-1974.
- Gandolfo MA, Nixon KC and Crepet WL** (1998). New fossil flower from the Turonian of New Jersey: *Dressiantha bicarpellata* gen. et sp. nov. (*Capparales*). *American Journal of Botany*, **85**: 964-974.
- Gautier M., Hayes H. and Eggen A.** (2002). An extensive and comprehensive radiation hybrid map of bovine Chromosome 15 : comparison with human Chromosome 11. *Mammalian Genome*, **13**: 316-319.
- Gebhardt C., Ritter E., Barone A., Debener T., Walkemeier B., Schachtschnabel U., Kaufmann H., Thompson R.D., Bonierbale M.W., Ganai M.W., et al.** (1991). RFLP maps of potato and their alignment with the homoeologous tomato genome. *Theoretical and Applied Genetics*, **83**: 49-57.
- Gebhardt C., Walkemeier B., Henselewski H., Barakat A., Delseny M. and Stuber K.** (2003). Comparative mapping between potato (*Solanum tuberosum*) and *Arabidopsis thaliana* reveals structurally conserved domains and ancient duplications in the potato genome. *The Plant Journal*, **34**: 529-542.
- Goffeau A., Barrell B.G., Bussey H., Davis R. W., Dujon B., Feldmann H., Galibert F., Hoheisel J. D., Jacq C., Johnston M., et al.** (1996). Life with 6000 Genes. *Science*, **274**: 563-567.

**Gonçalves M.M. et Rodrigues M.L.** (1976). Estudos sobre o café de Timor. II. Nota sobre as possibilidades de produção do híbrido de Timor no seu habitat natural. In "Missão de Estudos Agronômicos do Ultramar", Lisboa, Portugal, Comunicações, **86**: 31-72.

**Gopalkrishnan K.S.** (1951). Notes on the morphology of the genus *Hemileia*. Mycologia, **43**: 271-83.

**Grant D., Cregan P. and Shoemaker R.C.** (2000). Genome organization in dicots: Genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. Proceedings of the National Academy of Sciences (USA), **97**: 4168-4173.

**Grant M.R., McDowell J.M., Sharpe A.G., De Torres Zabala M., Lydiate D.J. and Dangl J.L.** (1998). Independent deletions of a pathogen-resistance gene in *Brassica* and *Arabidopsis*. Proceedings of the National Academy of Sciences (USA), **95**: 15843-15848.

**Grant V.** (1981). Plant Speciation (Columbia Univ. Press, New York).

**Grassias M. and Kammacher P.** (1975). Observations sur la conjugaison chromosomique de *Coffea arabica* L. Café Cacao Thé, **19**: 177-190.

**Gregory S.G., Sekhon M., Schein J., Zhao S., Osoegawa K., Scott C.E., Evans R.S., Burridge P.W., Cox T.V., Fox C.A. et al.** (2002). A physical map of the mouse genome. Nature, **418**: 743-750

**Griffiths S., Sharp R., Foote T.N., Bertin I., Wanous M., Reader S., Colas I. and Moore G.** (2006). Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. Nature, **439**: 749-752.

**Grube R.C., Radwanski E.R. and Jahn M.** (2000). Comparative genetics of disease resistance within the solanaceae. Genetics, **155**: 873-87.

**Guillaumet J.L. and 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'Éthiopie. Mult. ORSTOM (Adiopodoumé), Abidjan.

**Guimarães C.T., Sills G.R. and Sobral B.W.S.** (1997). Comparative mapping of Andropogoneae : *Saccharum* L. (sugarcane) and its relation to sorghum and maize. Proceedings of the National Academy of Sciences (USA). **94**: 14261-14266.

**Harlan S.C.** (1936). The genetical conception of the species. Biological Review, **11**: 83-112.

**Hermesen J.G.Th.** (1984). Some fundamental considerations on interspecific hybridization. Iowa State Journal research. **58**: 461-474.

**Herrera J.-C., Combes M.-C., Anthony F., Charrier A. and Lashermes P.** (2002a). Introgression into the allotetraploid coffee (*Coffea arabica* L.): segregation and recombination of the *C. canephora* genome in the tetraploid interspecific hybrid (*C. arabica* x *C. canephora*). Theoretical and Applied Genetics, **104**: 661-668.

**Herrera J.-C., Combes M.-C., Cortina H., Alvarado G. and Lashermes P.** (2002b). Gene introgression into *Coffea arabica* by way of triploid hybrids. Heredity, **89**: 488-494.

**Herrera J.-C., Combes M.-C., Cortina H. and Lashermes P.** (2004). Factors influencing gene introgression into the allotetraploid *Coffea arabica* L. from its diploid relatives. *Genome*, **47**: 1053-1060.

**Huang S., Van Der Vossen E.A.G., Kuang H., Vleeshouwers V.G.A.A., Zhang N., Borm T.J.A., Van Eck H.J., Baker B., Jacobsen E. and Visser R.G.F.** (2005). Comparative genomics enabled the isolation of the R3a late blight resistance gene in potato. *The Plant Journal*, **42**: 251-260.

**Hulbert S.H., Richter T.E., Axtell J.D. and Bennetzen J.L.** (1990). Genetic Mapping and Characterization of Sorghum and Related Crops by Means of Maize DNA Probes. *Proceedings of the National Academy of Sciences of the United States of America*, **87**: 4251-4255.

**Husband B.C. and Sabara H.A.** (2004). Reproductive isolation between autotetraploids and their diploid progenitors in fireweed, *Chamerion angustifolium* *Onagraceae*. *New Phytologist*, **161**: 703-713.

**Husband B.C.** (2000). Constraints on polyploid evolution: a test of the minority cytotype exclusion principle. *Proceedings of the Royal society of London. Série B. Biological sciences*, **267**: 217-223.

**International Coffee Organization**, <http://www.ico.org>

**International Human Genome Sequencing Consortium** (2001). Initial sequencing and analysis of the human genome. *Nature*, **409**: 860-921.

**Jung S., Main D., Staton M., Cho I., Zhebentyayeva T., Arús P. and Abbott A.** (2006). Synteny conservation between the *Prunus* genome and both the present and ancestral *Arabidopsis* genomes. *BMC Genomics*, **7**: 81-97.

**Kammacher P. and Capot, J.** (1972). Sur les relations caryologiques entre *Coffea arabica* et *C. canephora*. *Café Cacao Thé*, **16**: 289-294.

**Kearsey M.J., Ramsay L.D., Jennings D.E., Lydiate D.J., Bohuon E.J.R. and Marshall D.F.** (1996). Higher recombination frequencies in female compared to male meioses in *Brassica oleracea*. *Theoretical and Applied Genetics*, **92**: 363-367.

**Kevei Z., Seres A., Kereszt A., Kaló P., Kiss P., Tóth G., Endre G. and György B.** (2005). Significant microsynteny with new evolutionary highlights is detected between *Arabidopsis* and legume model plants despite the lack of macrosynteny. *Molecular Genetics Genomics*, **274**: 644-657.

**Kilian A., Chen J., Han F., Steffenson B. and Kleinbols A.** (1997). Towards map-based cloning of the barley stem rust resistance genes Rpg1 and Rpg4 using rice as an intergenomic cloning vehicle. *Plant Molecular Biology*, **35**:187-195.

**Kim S.-C. and Rieseberg L.H.** (1999). Genetic architecture of species differences in annual sunflowers : implications for adaptive trait introgression. *Genetics*, **153**: 965-977.

**Kosambi DD.** (1944). The estimation of map distances recombination values. *Annals of Eugenics*, **12**: 172-175.

**Kowalski S.P., Lan T-H., Feldmann K.A. and Paterson A.H.** (1994). Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. *Genetics*, **138**: 499-510.

**Krug C.A. and Mendes A.J.T.** (1940). Cytological observations in *Coffea*. IV. *Journal of Genetics*. **39**: 189-203.

**Ku H-M., Vision T., Liu J. and Tanksley S.D.** (2000). Comparing sequenced segments of the tomato and *Arabidopsis* genomes: Large-scale duplication followed by selective gene loss creates a network of synteny. *Proceedings of the National Academy of Sciences (USA)*, **97**: 9121-912.

**Kurata N., Moore G., Nagamura Y., Foote T., Yano M., Minobe Y. and Gale M.D.** (1994). Conservation of Genome Structure Between Rice and Wheat. *Bio-Technology*, **12**: 276-278.

**Kushalappa A.C. and Eskes A.B.** (1988). *Coffee Rust: epidemiology, Resistance and Management*. CRC Press, Inc. Boca Raton, Florida.

**Kushalappa A.C. and Eskes A.B.** (1989). *Advances In Coffee Rust Research*. *Annual Review of Phytopathology*, **27**: 503-531.

**Lagercrantz U.** (1998). Comparative mapping between *Arabidopsis thaliana* and *brassica nigra* indicates that brassica genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics*, **150**: 1217-1228.

**Lagercrantz U. and Lydiate D.** (1996). Comparative genome mapping in *Brassica*. *Genetics*, **144**: 1903-1910.

**Lanaud C.** (1979). Etude de problèmes génétiques posés chez le caféier par l'introggression de caractères d'une espèce sauvage *C. kianjavatensis*, *Mascarocoffea*. dans l'espèce cultivée *C. canephora Eucoffea*.. *Café Cacao Thé*, **23**: 3-28.

**Lander E.S., Green P., Abrahamson J., Barlow A., Daly M.J., Lincoln S.E. and 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., Cros J., Trouslot P., Anthony F. and Charrier A.** (1995). Origin and genetic diversity of *Coffea arabica* L. based on DNA molecular markers. In : 16ème Colloque Scientifique International sur le Café (Kyoto, Japon, 9-14 Avril 1995), Volume II, pp. 528-536, Association Scientifique Internationale du Café, Paris, France.

**Lashermes P., Couturon E. and Charrier A.** (1994). Combining ability of doubled haploids in *Coffea canephora* P. *Plant breeding*, **112**: 330-337.

**Lashermes P., Cros J., Combes M.-C., Trouslot P., Anthony F., Hamon S. and Charrier A.** (1996a). Inheritance and restriction fragment length polymorphism of chloroplast DNA in the genus *Coffea* L. *Theoretical and Applied Genetics*, **93**: 626-632.



**Lashermes P., Couturon E., Moreau N., Paillard M. and Louarn J.** (1996b). Inheritance and genetic mapping of self-incompatibility in *Coffea canephora* Pierre. *Theoretical and Applied Genetics*, **93**: 458-462.

**Lashermes P., Combes M.-C., Trouslot P. and Charrier A.** (1997). Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. *Theoretical and Applied Genetics*, **94**: 947-955.

**Lashermes P., Combes M.-C., Robert J., Trouslot P., D'Hont A., Anthony F. and Charrier A.** (1999). Molecular characterisation and origin of the *Coffea arabica* L. genome. *Molecular Genomics Genetics*, **261**: 259-266.

**Lashermes P., Combes M.-C., Topart P., Graziosi G., Bertrand B. and Anthony F.** (2000). Molecular breeding in coffee (*Coffea arabica* L.) In: *Coffee Biotechnology*. Pandey A (ed), Academic Press

**Lashermes P., Paczek V., Trouslot P., Combes M. C., Couturon E. and Charrier A.** (2000a). Single-locus inheritance in the allotetraploid *Coffea arabica* L. and interspecific hybrid *C. arabica* x *C. canephora*. *Journal of Heredity*, **91**: 81-85.

**Lashermes P., Andrzejewski S., Bertrand B., Combes M.-C., Dussert S., Graziosi G., Trouslot P., et Anthony F.** (2000b). Molecular analysis of introgressive breeding in coffee (*Coffea arabica* L.). *Theoretical and Applied Genetics*, **100**: 139-146.

**Lashermes P., Combes M.C., Prakash N.S., Trouslot P., Lorieux M. and Charrier A.** (2001). Genetic linkage map of *Coffea canephora*: effect of segregation distortion and analysis of recombination rate in male and female meiosis. *Genome*, **44**: 589-596.

**Lashermes P., Combes M.-C., Mahé L., Noir S., Prakash N.S. and Varzea V.M.P.** (2005). Progress in genetic and physical mapping of a leaf rust resistance locus in Coffee. *Durable Resistance to Coffee Leaf Rust*, Chap **14**: 333-362.

**Le Pierrès D.** (1995). Etude des hybrides interspécifiques tétraploïdes de première génération entre *Coffea arabica* L. et les caféiers diploïdes. PhD thesis, University of Paris-XI.

**Le Pierrès D.** (1999). Etudes des hybrides naturels entre *Coffea arabica* et *Coffea canephora* de Nouvelle-Calédonie. IRD, Centre de Nouméa. pp. 44.

**SOL genomic networks:** <http://www.sgn.cornell.edu/>

**Leach J.E., Vera Cruz C.M., Bai J. and Leung H.** (2001). Pathogen fitness penalty as a predictor of durability of disease resistance genes. *Annual Review of Phytopathology*, **39**: 187-224.

**Leister D., Kurth J., Laurie D.A., Yano M., Sasaki T., Devos K., Graner A. and Schulze-Lefert P.** (1998). Rapid reorganization of resistance gene homologues in cereal genomes. *Proceedings of the National Academy of Sciences (USA)*, **95**: 370-375.

**Leitch I.J. and Bennett M.D.** (1997). Polyploidy in angiosperms. *Trends in Plant Science*, **2**: 470-476.

- Leroy J.F.** (1980). Evolution et taxogenèse chez les caféiers (*Coffea* L., *Psilanthus* Hook. f. et *Nostolachma* Durand). Hypothèse sur leur origine. In "Comptes rendus de l'Académie des Sciences", Paris, pp. 593-596.
- Lewis D.** (1943). The Incompatibility sieve for producing polyploids. *Journal of Genetics*, **45**: 264-264.
- Li G., Gao M., Yang B. and Quiros C.F.** (2003). Gene for gene alignment between the Brassica and *Arabidopsis* genomes by direct transcriptome mapping. *Theoretical and Applied Genetics*, **107**: 168-180.
- Lin C, Mueller L.A., McCarthy J., Crouzillat D., Pétiard V. and Tanksley S.D.** (2005). Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. *Theoretical and Applied Genetics*, **5**: 1-17.
- Lin X., Kaul S., Rounsley S., Shea T.P., Benito M.-I., et al.** (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature*, **402**: 761-772.
- Liu Y.-S. and D. Zamir** (1999). Second generation *L. pennellii* introgression lines and the concept of bin mapping. *Tomato Genetics Cooperative*. **49**:26-30.
- Livingstone K.D., Lackney V.K., Blauth J.R., van Wijk R. and Jahn M.K.** (1999) Genome Mapping in Capsicum and the Evolution of Genome Structure in the Solanaceae. *Genetics*, **152**: 1183-1202.
- Lunn J.E.** (2003). Sucrose-Phosphatase gene families in plants. *Gene*, **303**: 187-196.
- Lynch M., Conery J.S.** (2000). The evolutionary fate and consequences of duplicate genes. *Science*, **290**: 1151-1155.
- Maere S. De Bodt S., Raes J., Casneuf T., Van Montagu M., Kuiper M. and Van de Peer Y.** (2005). Modeling gene and genome duplications in eukaryotes. *Proceedings of the National Academy of Sciences (USA)*, **102**: 5454-5459.
- Mahé L., Le Pierrès D, Combes M.-C. and Lashermes P.** Introgressive hybridization between the allotetraploid *Coffea arabica* and one of its diploid ancestors *C. canephora* in an exceptional sympatric zone in New Caledonia. *Genome*, Accepted.
- Mammadov J.A., Steffenson B.J. and Saghai Maroof M.A.** (2005). High-resolution mapping of the barley leaf rust resistance gene Rph5 using barley expressed sequence. *Theoretical and Applied Genetics*, **111**: 1651-1660.
- Marra M.A., Kucaba T.A., Dietrich N.L., Green E.D., Brownstein B., Wilson R.K., McDonald K.M., Hillier L., McPherson J., Waterston R.** (1997) High Throughput Fingerprint Analysis of Large-Insert Clones. *Genome Research*, **7**: 1072-1084
- Marques D.V. and Bettencourt A.J.** (1979). Resistência à *H. vastatrix* numa população de Icatú. *Garcia de Orta (Série de Estudos Agronômicos)*, **6**: 19-24.

- Martin G.B., Ganai M.W. and Tanksley S.D.** (1992). Construction of a yeast artificial chromosome library of tomato and identification of cloned segments linked to two disease resistance loci. *Molecular Genomics Genetics*, **233**: 25-32.
- Martin N.H., Bouck A.C. and Arnold, M.L.** (2006). Detecting adaptive trait introgression between *Iris fulva* and *I. brevicaulis* in highly selective field conditions. *Genetics*, **172**: 2481-2489.
- Martinez-Perez E., Shaw P. and Moore G.** (2001). The Ph1 locus is needed to ensure specific somatic and meiotic centromere association. *Nature*, **411**: 204-207.
- Martinsen G.D., Whitham T.G., Turek R.J. and Keim, P.** (2001). Hybrid populations selectively filter gene introgression between species. *Evolution*, **55**: 1325-1335.
- Masterson J.** (1994). Stomatal size in fossil plants : evidence for polyploidy in majority of angiosperms. *Science*, **264**: 421-424.
- Mayer K., Murphy G., Tarchini R., Wambutt R., Volckaert G., Pohl T., Düsterhöft A., Stiekema W., Entian K.-D., Terryn N., Lemcke K., Haase D., Hall C.R., Van Dodeweerd A.-M., Tingey S.V., Mewes H.-W., Bevan M.W. and Bancroft I.** (2001). Conservation of microstructure between a sequenced region of the genome of rice and multiple segments of the genome of *Arabidopsis thaliana*. *Genome Research*, **11**: 1167-1174.
- Mayer K., Schüller C., Wambutt R., Murphy G., Volckaert G., et al.** (1999). Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature*, **402**: 769-780.
- Mayne W.W.** (1932). Physiologic specialization of *Hemileia vastatrix* B. et Br. *Nature*, **129**: 510.
- Mazzafera P. and Carvalho A.** (1992). Breeding for low seed caffeine content of coffee (*Coffea L.*) by interspecific hybridization. *Euphytica*, **59**: 55-60.
- McFadden D.E., Kwong L.C., Yam I.Y.L and Langlois S.** (1993). Parental origin of triploidy in human fetuses: evidence for genomic imprinting. *Human Genetics*, **1993**: 465-69.
- McHale N.A.** (1983). Environmental induction of high frequency 2n pollen formation in diploid *Solanum*. *Canadian Journal of genetics and cytology*, **25**: 609-615.
- McIntosh R.A. and Watson I.A.** (1982). The Rust Fungi (eds Scott, K. J. & Chakravorty, A. K.) Academic, London, 121-149.
- Michelmore R.W. and Meyers B.C.** (1998). Clusters of Resistance Genes in Plants Evolve by Divergent Selection and a Birth-and-Death Process. *Genome Research*, **8**: 1113-1130.
- Ming R., Liu S.C., Lin Y.R., da Silva J., Wilson W., Braga D., van Deynze A., Wensla T.F., Wu K.K., Moore P.H., Burnquist W., Sorrells M.E., Irvine J.E. and Paterson A.H.** (1998). Detailed alignment of saccharum and sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes. *Genetics*, **150**: 1663-1682.

- Mitchell-Olds T. and Clauss M.J.** (2002). Plant evolutionary genomics. *Current Opinion in Plant Biology*, **5**: 74-79.
- Monaco L.C.** (1977). Consequences of the introduction of coffee leaf rust into Brazilian Annual NY Academy of Science, **287**: 57-71.
- Moncada P. and McCouch S.** (2004). Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. *Genome*, **47**: 501-509.
- Montagnon C. et Leroy T.** (1993). Réaction à la sécheresse de jeunes caféiers *Coffea canephora* de Côte-d'Ivoire appartenant à différents groupes génétiques. *Café, cacao, thé*, 1993, **37**, 3, pp. 179-190
- Moore G.** (2000). Cereal chromosome structure, evolution and pairing. *Annual Review of Plant Physiology and Plant Molecular Biology*, **51**: 195-222.
- Moore S., Payton P., Wright M., Tanksley S. and Giovannoni J.** (2005). Utilization of tomato microarrays for comparative gene expression analysis in the *Solanaceae*. *Journal of Experimental Botany*, **56**: 2885-2895.
- Moreno G. and Alvarado G.** (2000). La variedad Colombia: veinte anos de adopcion y comportamiento frente a nuevas razas de la roya del cafeto. *Cenicafe Technical Bulletin* (22), pp. 32.
- Mungall A.J., Palmer S.A., Sims S.K., Edwards C.A., Ashurst J.L., et al.** (2003). The DNA sequence and analysis of human chromosome 6. *Nature*, **425**: 805-813.
- Nadeau J.H. and Taylor B.A.** (1984). Lengths of chromosomal segments conserved since divergence of man and mouse. *Proceedings of the National Academy of Sciences (USA)*, **81**: 814-818.
- Noir S., Anthony F., Bertrand B., Combes M-C. and Lashermes P.** (2003). Identification of a major gene (*Mex-1*) from *Coffea canephora* conferring resistance to *Meloidogyne exigua* in *Coffea arabica*. *Plant Pathology*, **52**: 97-103.
- Noir S., Pateyron S., Combes M-C., Lashermes P. and Chalhoub B.** (2004). Construction and characterisation of a BAC library for genome analysis of the allotetraploid coffee species (*Coffea arabica* L). *Theoretical and Applied Genetics*, **109**: 225-30.
- Noronha-Wagner H., and Bettencourt A.J.** (1967). Genetic study of resistance of *Coffea* sp. to leaf rust. I. Identification and behaviour of four factors conditioning disease reaction in *Coffea arabica* to twelve physiologic races of *Hemileia vastatrix*. *Canadian Journal of Botany*, **45**:2021-2031.
- O'Brien S.J., Menotti-Raymond M., Murphy W.J., Nash W.G., Wienberg J., Stanyon R., Copeland N.G., Jenkins N.A., Womack J.E. and Marshall Graves J.A.** (1999). The Promise of Comparative Genomics in Mammals. *Science*, **286**: 458-481.
- O'Brien T.G. and Kinnaird M.F.** (2003). Caffeine and conservation. *Science*, **300**: 587.

- O'Neill C.M. and Bancroft I.** (2000). Comparative physical mapping of segments of the genome of *Brassica oleracea* var. *alboglabra* that are homoeologous to sequenced regions of chromosomes 4 and 5 of *Arabidopsis thaliana*. *The Plant Journal*, **23**: 233-243.
- Oh K.C., Hardeman K., Ivanchenko M.G., Ellard-Ivey M., Nebenführ A., White T.J. and Lomax T.L.** (2002). Fine mapping in tomato using microsynteny with the *Arabidopsis* genome: the Diageotropica (Dgt) locus. *Genome biology*, **3**: 1-11.
- Orozco, F.J.** (1976). Utilizacion del hibrido triploide de *C. arabica* x *C. canephora* en cruzamientos interespecificos. *Cenicafé*, **27**: 143-157.
- Ostberg C.O., Slatton S.L. and Rodriguez R.J.** (2004). Spatial partitioning and asymmetric hybridization among sympatric coastal steelhead trout *Oncorhynchus mykiss*., coastal cutthroat trout *O. clarki clarki*. and interspecific hybrids. *Molecular Ecology*, **13**: 2773-2788.
- Ozkan H., Levy A.A. and Feldman M.** (2001). Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *The Plant Cell*, **13**: 1735-1747.
- Pan Q., Liu Y.-S., Budai-Hadrian O., Sela M., Carmel-Goren L., Zamir D. and Fluhr R.** (2000). Comparative genetics of nucleotide binding site-leucine rich repeat resistance gene homologues in the genomes of two dicotyledons: Tomato and *Arabidopsis*. *Genetics*, **155**: 309-322.
- Parkin I.A.P., Lydiate D.J. and Trick M.** (2002). Assessing the level of collinearity between *Arabidopsis thaliana* and *Brassica napus* for *A. thaliana* chromosome 5. *Genome*, **45**: 356-366.
- Parrella G., Ruffel S., Moretti A., Morel C., Palloix A. and Caranta C.** (2002). Recessive resistance genes against potyviruses are localized in colinear genomic regions of the tomato (*Lycopersicon* spp.) and pepper (*Capsicum* spp.) genomes. *Theoretical and Applied Genetics*, **105**: 855-861.
- Paterson A.H.** (1996). Towards a unified genetic map of higher plants, transcending the monocot-dicot divergence. *Nature Genetics*, **14**: 380-382.
- Paterson A.H., Lin Y.-R., Li Z., Schertz K.F., Doebley J.F., Pinson S.R.M., Liu S.-C., Stansel J.W. and Irvine J.E.** (1995). Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science*, **269**: 1714-1718.
- Paterson A.H., Bowers J.E., Burow M.D., Draye X., Elsik C.G., Jiang C.X., Katsar C.S., Lan T.H., Lin Y.R., Ming R., et al.** (2000). Comparative genomics of plant chromosomes. *Plant Cell*, **12**: 1523-1540.
- Patocchi A., Vinatzer B.A., Gianfranceschi L., Tartarini S., Zhang H.B., Sansavini C. and Gessler C.** (1999). Construction of a 550 kb BAC contig spanning the genomic region containing the apple scab resistance gene *Vf*. *Molecular Genomics Genetics*, **262**: 884-891.

- Payseur B.A. and Nachman M.W.** (2005). The genomics of speciation: investigating the molecular correlates of X chromosome introgression across the hybrid zone between *Mus domesticus* and *Mus musculus*. *Biological Journal of the Linnean Society*, **84**: 523-534.
- Peng J., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M., Flintham J.E., Beales J., Fish L.J., Worland A.J., Pelica F. et al.** (1999). 'Green revolution' genes encode mutant gibberellin response modulators. *Nature*, **400**: 256-261.
- Pereira M.G., Lee M., Bramel-Cox P., Woodman W., Doebley J. and Whitkus R.** (1994). Construction of an RFLP map in sorghum and comparative mapping in maize. *Genome*, **37**: 236-243.
- Prakash N.S., Ganesh D. and Bhat S.S.** (2005) Population dynamics of coffee leaf rust (*Hemileia vastatrix* BERK. et BR.) and recent advances in rust research in India. *Durable Resistance to Coffee Leaf Rust*, Chap **17**: 411-442.
- Prakash N.S., Combes M.-C. , Somanna N. and Lashermes P.** (2002). AFLP analysis of introgression in coffee cultivars (*Coffea arabica* L.) derived from a natural interspecific hybrid. *Euphytica*, **124**: 265-271.
- Prakash N.S., Combes M.-C., Dussert S., Naveen K.S. and Lashermes, P.** (2005). Analysis of genetic diversity in Indian robusta coffee genepool *Coffea canephora*. in comparison with a representative core collection using SSRs and AFLPs. *Genetic Resources and Crop Evolution*, **52**: 333-343.
- Prakash N.S., Marques D.V., Varzea V.M.P., Silva M.C., Combes M.-C. and Lashermes P.** (2004). Introgression molecular analysis of a leaf rust resistance gene from *Coffea liberica* into *C. arabica* L. *Theoretical and Applied Genetics*, **109**: 1311-1317.
- Proudhon D., Wei J., Briat J. and Theil E.C.** (1996). Ferritin gene organization : differences between plants and animals suggest possible kingdom-specific selective constraints. *J Mol Evol*, **42**: 325-336.
- Raes J., Vandepoele K., Simillion C., Saeys Y. and Van de Peer Y.** (2003). Investigating ancient duplication events in the *Arabidopsis* genome. *Journal of Structural and Functional Genomics*, **3**: 117-129.
- Rafalski A.** (2002). Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology*, **5**: 94–100.
- Ramachandran M. and Srinivasan C.S.** (1979). Four generations of selection for resistance to race I of leaf rust in arabica cv. S.288 x 'Kents'. *Indian Coffee* **43**(6): 159 - 161
- Ramsey J. and Schemske D.W.** (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual review of ecology and systematics*, **29**: 467-501.
- Rieseberg L. H. and JWendel. F.** (1993). Introgression and its consequences in plants. In R.G. Harrison, editor. *Hybrid zones and the evolutionary process*. Oxford University Press. pp. 70-109.

**Rieseberg L.H. and Carney S.E.** (1998). Tansley review no. 102: plant hybridization. *New Phytologist*, **140**: 599-624.

**Rieseberg L.H., Beckstrom-Sternberg S. and Doan K.** (1990). *Helianthus annuus* ssp. *Texanus* has chloroplast DNA and nuclear ribosomal RNA genes of *Helianthus debilis* ssp. *cucumerifolius*. *Proceedings of the National Academy of Sciences (USA)*, **87**: 593-597.

**Riley R. and Chapman V.** (1958). Genetic control of the cytologically diploid behavior of hexaploid wheat. *Nature*, **182**: 713-715.

**Rink A., Santschi M.E., Eyer K.M., Roelofs B., Hess M., Godfrey M., Karajusuf E.K., Yerle M., Milan D. and Beatie C.W.** (2002). A first-generation EST RH comparative map of the porcine and human genome. *Mammalian Genome*, **13**: 578-587.

**Robic A., Jeon J.T., Amarger V., Chardon P., Looft C., Andersson L., Gellin J. and Milan D.** (2001). Construction of a high-resolution RH map of the human 2q35 region on TNG panel and comparison with a physical map of the porcine homologous region 15q25. *Mammalian Genome*, **12**: 380-386.

**Robic A., Seroude V., Jeon J.T., Yerle M., Wasungu L., Andersson L., Gellin J. and Milan D.** (1999). A radiation hybrid map of the RN region in pigs demonstrates conserved gene order compared with the human and mouse genomes. *Mammalian Genome*, **10**: 565-568.

**Rodrigues Jr. C.J., Bettencourt A.J. and Rijo L.** (1975). Races of the pathogen and resistance to coffee rust. *Annual Review of Phytopathology*, **13**: 49-70.

**Rodrigues Jr. CJ** (1990). Coffee Rusts: history, taxonomy, morphology, distribution and host resistance. *Fitopatologia brasileira*, **15**:5-9.

**Roger L.** (1951). *Phytopathologie des pays chauds*. Paul le Chavalier Vol11 Paris, pp. 838-851.

**Rossberg M., Theres K., Acarkan A., Herrero R., Schmitt T., Schumacher K., Schmitz G. and Schmidt R.** (2001). Comparative sequence analysis reveals extensive microcolinearity in the lateral suppressor regions of the Tomato, *Arabidopsis*, and *capsella* genomes. *Plant Cell*, **13**: 979-988.

**Ruffel S., Dussault M.H., Palloix A., Moury B., Bendahmane A., Robaglia C. and Caranta C.** (2002). A natural recessive resistance gene against potato virus Y in pepper corresponds to the eukaryotic initiation factor 4E (eIF4E). *The Plant Journal*, **32**: 1067-1075.

**Ruffel S., Gallois J.L., Lesage M.L. and Caranta C.** (2005). The recessive potyvirus resistance gene pot-1 is the tomato orthologue of the pepper pvr2-eIF4E gene. *Molecular Genomics Genetics*, **274**: 346-353.

**Sahrawy M., Hecht V., Lopez-Jaramillo J., Chueca A., Chartier Y. and Meyer Y.** (1996). Intron Position as an Evolutionary Marker of Thioredoxins and Thioredoxin Domains. *Journal of Molecular Evolution*, **42**: 422-431.

**Salanoubat M., Lemcke K., Rieger M., Ansorde W., Unseld M., et al.** (2000). Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. *Nature*, **408**: 820-822.

**Salse J., Piegu B., Cooke R. and Delseny M.** (2002). Synteny between *Arabidopsis thaliana* and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequencing project. *Nucleic Acids Research*, **30**: 2316-2328.

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

**Sankoff D. and Nadeau J.H.** (2003). Chromosome rearrangements in evolution : From gene order to genome sequence and back. *Proceedings of the National Academy of Sciences (USA)*, **100**: 11188-11189.

**Savard L., Li P., Strauss S.H., Chase M.W., Michaud M. and Bousquet J.** (1994). Chloroplast and nuclear gene sequences indicate late Pennsylvanian time for the last common ancestor of extant seed plants. *Proceedings of the National Academy of Sciences (USA)*, **91**: 5163-5167.

**Sawkins M.C., Maass B.L., Pengelly B.C., Newbury H.J., Ford-Lloyd B.V., Maxted N. and Smith R.** (2001). Geographical patterns of genetic variation in two species of *Stylosanthes* Sw. using amplified fragment length polymorphism. *Molecular Ecology*, **10**: 1947-1958.

**Schibler L. Vaiman D., Oustry A., Giraud-Delville C. and Cribiu E.P.** (1998). Comparative gene mapping : a fine-scale survey of chromosome rearrangements between ruminants and humans. *Genome Research*, **8**: 901-915.

**Schluter D.** (2001). Ecology and the origin of species. *Trends Ecology and Evolution*, **16**: 372-380.

**Schmidt T.** (2000). Synteny : recent advances and future prospects. *Current Opinion in Plant Biology*, **3**: 97-102.

**Schwarzbach A.E., Donovan L.A. and Rieseberg L.H.** (2001). Transgressive character expression in a hybrid sunflower species. *American Journal of Botany*, **88**: 270-277.

**Sears E.R.** (1976). Genetic control of chromosome pairing in wheat. *Annual Review of Genetics*, **10**: 31-51.

**Sera T., Ruas P.M., Ruas C.F., Diniz L.E.C., Carvalho V.P., Rampim L., Ruas E.A. and Silveira S.R.** (2003). Genetic polymorphism among 14 elite *Coffea arabica* L. cultivars using RAPD markers associated with restriction digestion. *Genetics and Molecular Biology*, **26**: 59-64.

**Sharpe A.G., Parkin I.A.P., Keith D.J. and Lydiate D.J.** (1995). Frequent nonreciprocal translocations in the amphidiploid genome of oilseed rape *Brassica napus*.. *Genome*, **38**: 1112-1121.

**Silva M.C., Nicole M., Rijo L., Geiger J. P. and Rodrigues C.J.** (1999). Cytochemical Aspects of the Plant-Rust Fungus Interface during the Compatible Interaction *Coffea*



*arabica* (cv. Caturra)-*Hemileia vastatrix* (race III). International Journal of Plant Sciences, **160**: 79-91.

**Silva M.D.S, Várzea V., Guerra-Guimarães L., H.G. Azinheira H.G., Fernandez D., Petitot A.-S., Bertrand B., Lashermes P. and Nicole M.** (2006). Coffee resistance to the main diseases: leaf rust and coffee berry disease. Brazilian Journal of Plant Physiology, **18**: 119-147.

**Simillion C., Vandepoele K., Van Montagu M.C.E., Zabeau M. and Van de Peer Y.** (2002). The hidden duplication past of *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences (USA), **99**: 13627-13632.

**Srinivasan K.H. and Narasimhaswamy R.L.** (1975). A review of coffee breeding work done at the Government coffee experiment station, Balehonnur. Indian coffee, **34**: 311-321.

**Stebbins G.L.** (1958). The inviability, weakness, and sterility of interspecific hybrids. Advanced Genetics, **9**: 147-215.

**Steiger D., Nagai C., Moore P., Morden C., Osgood R. and Ming R.** (2002). AFLP analysis of genetic diversity within and among *Coffea arabica* cultivars. Theoretical and Applied Genetics, **105**: 209-215.

**Stracke S., Sato S., Sandal N., Koyama M., Kaneko T., Tabata S. and Parniske M.** (2004). Exploitation of colinear relationships between the genomes of *Lotus japonicus*, *Pisum sativum* and *Arabidopsis thaliana*, for positional cloning of a legume symbiosis gene. Theoretical and Applied Genetics, **108**: 442-449.

**Sweigart A.L. and J.H. Willis** (2003). Patterns of nucleotide diversity are affected by mating system and asymmetric introgression. Genetics, **13**: 488-507.

**Sybenga J.** (1959). Some sources of error in the determination of chromosome length. Chromosoma, **10**: 355-364.

**Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., et al.** (2000). Sequence and analysis of chromosome 5 of the plant *Arabidopsis thaliana*. Nature, **408**: 823-826.

**Tanksley S.D., Ganai M.W., and Martin G.B.** (1995). Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. Trends in Genetics, **11**: 63-68.

**Tanksley S.D., Ganai M.W., Prince J.P., de Vicente M.C., Bonierbale P.B., Broun P., Fulton T.M., Giovanonni J.J., Grandillo S., Martin G.B., et al.** (1992). High density molecular linkage maps of the tomato and potato genomes. Genetics, **132**: 1141-1160.

**Taylor M.S. and Semple C.A.** (2002). Sushi gets serious : the draft genome sequence of the pufferfish *Fugu rubripes*. Genome Biology, **3**:1025.

**The Arabidopsis Genome Initiative** (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature, **408**: 796-815.

**The *C. elegans* Sequencing Consortium** (1998). Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science*, **282**: 2012-2018.

**The chromosome 21 mapping and sequencing consortium** (2000). *Nature*, **405**: 311-319.

**The Institute for Genomic Research** (<http://www.tigr.org>)

**Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., et al.** (2000). Sequence and analysis of chromosome 1 of the plant *Arabidopsis thaliana*. *Nature*, **408**: 816-819.

**Tomkins J.-P., Yu Y., Miller-Smith H., Frisch D.A., Woo S.S. and Wing R.A.** (1999). A bacterial artificial chromosome library for sugarcane. *Theoretical and Applied Genetics*, **99**: 419-424.

**Tsukaya H., Fukuda T., and Yokoyama J.** (2003). Hybridization and introgression between *Callicarpa japonica* and *C. mollis* *Verbenaceae*. in central Japan. *Molecular Ecology*, **12**: 3003-3011.

**Uchida I.A. and Freeman V.C.** (1985). Triploidy and chromosomes. *American Journal of Obstretician Gynecology*, **151**: 65-69.

**Ungerer M.C., Baird S.J.E., Pan J. and Rieseberg L.H.** (1998). Rapid hybrid speciation in wild sunflowers. *Proceedings of the National Academy of Sciences (USA)*, **95**: 11757-11762.

**Van Daelen R.A.J.J., Gerbens F., Van Ruissen F., Aarts J., Hontelez J. and Zabel P.** (1993). Long-range physical maps of two loci (Aps-1 and GP79) flanking the root-knot nematode resistance gene (Mi) near the centomere of tomato chromosome 6. *Plant Molecular Biology*, **23**: 185-192.

**Van der Hoeven R., Ronning C., Giovannoni J., Martin G. and S. Tanksley** (2002). Deductions about the Number, Organization, and Evolution of Genes in the Tomato Genome Based on Analysis of a Large Expressed Sequence Tag Collection and Selective Genomic Sequencing. *Plant Cell*, **14**(7): 1441-1456.

**Van de Peer Y. and De Wachter R.** (1994). TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Computer Applied Bioscience*, **10**: 569-570.

**Van der Plank J.E.** (1978). *Genetics and molecular basis of plant pathogenesis*. Berlin: Springer-Verlag. pp.163.

**Van der Vossen E.A.G., Rouppe van der Voort J.N.A.M., Kanyuka K., Bendahmane A., Sandbrink J.M., Baulcombe D.C., Bakker J., Stiekema W.J. and Klein-Lankhorst R.M.** (2000). Homologues of a single resistance gene cluster in potato confer resistance to distinct pathogens: a virus and a nematode. *The Plant Journal*, **23**: 567-576.

**Van der Vossen H.A.M.** (1985). Coffee selection breeding. In "Coffee: Botany, biochemistry and production of beans and beverage" (Clifford M. N., et Wilson K. C., Eds), Croom Helm, London, Sydney, pp. 48-97.

- Van der Vossen H.A.M.** (1997). Strategies of variety improvement in cocoa with emphasis on durable disease resistance. International Group for Genetic Improvement of Cocoa. pp. 9-18.
- Van der Vossen H.A.M.** (2001). Coffee breeding practices. In: Clarke RJ, Vitzthum OG (eds) Coffee Recent Developments - Agronomy 1 (eds), Blackwell Science Ltd, London, pp. 184-201.
- Van der Vossen H.A.M.** (2005) State of the art of developing durable resistance to biotrophic pathogens in crop plants, such as coffee leaf rust. Durable Resistance to Coffee Leaf Rust, Chap 1: 1-32.
- Van Dodeweerd A-M., Hall C.R., Bent E.G., Johnson S.J., Bevan M.W. and Bancroft I.** (1999). Identification and analysis of homoeologous segments of the genomes of rice and *Arabidopsis thaliana*. Genome, **42**: 887-892.
- Vandepoele K., Simillion C. and Van de Peer Y.** (2002). Detecting the undetectable : uncovering duplicated segments in *Arabidopsis* by comparison with rice. Trends in Genetics, **18**: 606-608.
- Vardi A. and Zohaky D.** (1967). Introgression in wheat via triploid hybrids. Heredity, **22**: 541.
- Várzea V.M.P., Rodrigues Jr. C.J. and Lewis B.G.** (2002). Distinguishing characteristics and vegetative compatibility of *Colletotrichum kahawe* in comparison with other related species from coffee. Plant Pathology, **51**: 202-207.
- Várzea V.M.P. and D.V. Marques** (2005). Population variability of *Hemileia vastatrix* vs coffee durable resistance. In: L. Zambolim, Zambolim E. and Várzea V.M.P (eds), Universidade Federal de Viçosa, Brasil. Durable resistance to coffee leaf rust. Chap 4: 53-74.
- Ventelon M., Deu M., Garsmeur O., Doligez A., Ghesquière A., Lorieux M., Rami J.F., Glaszmann J.C. and Grivet L.** (2001). A direct comparison between the genetic maps of sorghum and rice. Theoretical and Applied Genetics, **102**: 379-386.
- Vishveshwara S.** (1971). Breeding for quality in coffee. Indian Coffee (Bangalore). **35**: 509-512.
- Vision T.J., Brown D.G. and Tanksley S.D.** (2000). The Origins of Genomic Duplications in *Arabidopsis*. Science, **290**: 2114-2117.
- Wagner M. and Bettencourt A.J.** (1965) Inheritance of reaction to *Hemileia vastatrix* Berk & Br. in *Coffea arabica* L. Progress report 1960-65, Coffee Rusts Research Center, Oeira, Portugal.
- Watkins-Chow D.E., Buckwalter M.S., Newhouse M.M., Lossie A.C., Brinkmeier M.L. and Camper S.A.** (1997). Genetic mapping of 21 genes on mouse chromosome 11 reveals disruptions in linkage conservation with human chromosome 5. Genomics, **40**: 114-122.

- Wei F., Gobelman-Werner K., Morroll S.M. et al.** (1999). The Mla (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within 240-kb DNA interval on chromosome 5S (1HS) of barley. *Genetics*, **153**: 1929-1948.
- Werner J. E., Endo T. R. and Gill B. S.** (1992). Towards a cytogenetically based physical map of the wheat genome. *Proceedings of the National Academy of Sciences (USA)*, **89**: 11307-11311.
- Whittemore A. T., and Schaal B. A.** (1991). Interspecific gene flow in oaks. *Proceedings of the National Academy of Sciences (USA)*, **88**: 2540-2544.
- Wikstrom N., Savolainen V. and Chase M.W.** (2001). Evolution of the angiosperms: calibrating the family tree. *Proceedings of the Royal society of London. Série B. Biological sciences*, **268**: 2211-2220.
- Wolfe K., Gouy M., Yang Y., Sharp P. and Li W.H.** (1989). Date of the monocot dicot divergence estimated from chloroplast DNA sequence data. *Proceedings of the National Academy of Sciences (USA)*, **86**: 5201-5205.
- Wu F., Mueller L.A., Crouzillat D., Pétiard V. and Tanksley S.D.** (in press). Combining Bioinformatics and Phylogenetics to Identify Large Sets of Single Copy, Orthologous Genes (COS) for Comparative, Evolutionary and Systematic Studies: A Test Case in the Euasterid Plant Clade. *Genetics*, in press
- Yan H.H., Mudge J., Kim D.J., Larsen D., Denny R., Shoemaker R.C., Cook D.R. and Young N.D.** (2004). Comparative physical mapping reveals features of microsynteny between *Glycine max*, *Medicago truncatula*, and *Arabidopsis thaliana*. *Genome*, **47**: 141-155.
- Yan H.H., Mudge J., Kim D.J., Larsen D., Shoemaker R.C., Cook D.R. and Young N.D.** (2003). Estimates of conserved microsynteny among the genomes of *Glycine max*, *Medicago truncatula* and *Arabidopsis thaliana*. *Theoretical and Applied Genetics*, **106**: 1256-1265.
- Yang Y.W., Lai K.N., Tai P.Y. and Li W.H.** (1999). Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other lineages. *Journal of Molecular Evolution*, **48**: 597-604.
- Yu G.X., Bush A.L. and Wise R.P.** (1996). Comparative mapping of homoeologous group 1 regions and genes for resistance to obligate biotrophs in *Avena*, *Hordeum*, and *Zea mays*. *Genome*, **39**: 155-164.

## Annexes

## Liste des abréviations

ADN	Acide Désoxyribo-Nucléique
ADNcp	Acide Désoxyribo-Nucléique chloroplastique
AFLP	Amplification Fragment Length Polymorphism
ARN	Acide Ribonucléique
BAC	Bacterial Artificial Chromosome
BE	BAC end
CIFC	Coffee Rust Research Center
cM	centimorgans
cv	cultivar
EST	Expressed Sequence Tag
FAO	Food and Agriculture Organization
HdT	Hybride du Timor
ILs	Introgressed Lines
IRD	Institut de Recherche pour le Développement
ITS	Internal Transcribed Spacers
LRR	Leucine-Rich Repeat
ORF	Open Reading Frame
NBS	Nucleotide Binding Site
pb/kb	paires de bases/kilo-paires de bases
PCR	Polymerase Chain Reaction
pg	picogramme
RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
RGA	Resistance Gene Analogues
SSR	Simple Sequence Repeat
SCAR	Sequence-Characterised Amplified Region
Spp	Espèces
µm	Micromètre

## Liste des figures

Figure 1. Classification phylogénétique de 32 taxons de caféiers, sur la base du polymorphisme de l'ADN nucléaire ribosomal et distribution géographique des principaux groupes de la classification (Lashermes *et al.* 1997).

Figure 2. Origine de l'espèce allotetraploïde *Coffea arabica*

Figure 3. Représentation schématique des principales étapes dans l'histoire de la culture du caféier extrait de Anthony *et al.* 2002.

Figure 4. Développement des urédospores à la face abaxiale d'une feuille de caféier

Figure 5. Qu'est-ce qu'un polyploïde ? d'après Leitch et Bennett (1997)

Figure 6. Evolution des relations entre gènes et terminologie. ([http://www.nature.com/nrd/journal/v2/n8/slideshow/nrd1152\\_bx2.html](http://www.nature.com/nrd/journal/v2/n8/slideshow/nrd1152_bx2.html))

Figure 7. Alignement entre une partie du chromosome 6 de l'homme (*Hsa6*) et du chromosome 4 de la souris (*Mmu4*). (Gregory *et al.* 2002)

Figure 8. Une carte comparative consensus des graminées de Gale et Devos (1998)

Figure 9. Carte moléculaire de liaison de deux chromosome de l'aubergine et comparaison avec les régions homéologues du génome de la tomate (Doganlar *et al.* 2002).

Figure 10. Segment dupliqués d'*Arabidopsis*

[http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis\\_genome\\_duplication.shtml](http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis_genome_duplication.shtml)

Figure 11. Dendrogramme dépeignant les rapports phylogénétiques des taxons des plantes supérieures. Les lignes rouges indiquent le clade contenant la tomate et l'*Arabidopsis*. Figure basée sur le schéma 2 dans la référence (Chase *et al.*, 1993).

Figure 12. *ORF* commun entre la tomate II (dessus) et les *ORFs* des segments des chromosomes d'*Arabidopsis* 2-5.

Figure 13. Organisation de gènes de résistance chez les solanacées. (Grube *et al.* 2000)

Figure 14. Localization of sites in New Caledonia where *C. arabica* (striped stars), *C. canephora* (white star) are reported. The sampled sympatric zone (Col d'Amieu) between *C. arabica* and *C. canephora* is identified by an open circle.

Figure 15. Distribution of post-F1 hybrid plants according to the *C. canephora* introgression index based on SSR data.

Figure 16. Distribution of SSR loci according to their frequency of introgression among the 30 post-F1 hybrid plants. For each individual post-F1 plant, the locus was considered introgressed when exhibiting at least one *C. canephora* allele.

Figure 17. Dendrogram based on AFLP marker (Dice coefficient) constructed for the fourteen hybrids plants sampled in New Caledonia. The *canephora* AFLP introgression index is indicated on the right side of the dendrogram.

Figure 18. Example of marker (i.e. Sat244) associated with the SH3 leaf rust resistance gene segregating within the F2 (Matari x S.288) population.

Figure 19. Linkage map based on the F2 (Matari x S.288) population analysis of the chromosome segments either in coupling (right) or repulsion (left) surrounding the SH3 gene (highlighted). Molecular marker names are on the right, while the estimated map

distances in cM are shown on the left. SCAR markers showing codominant polymorphism are in bold.

Figure 20. A dendrogram depicting the phylogenetic relationship of Coffee, *Arabidopsis* and other crop species. The arrow marks the probable occurrence of the polyploidy event identified by Blanc *et al.* 2003.

Figure 21. Graphic representation of the putative paralogous predicted genes on the duplicated *Arabidopsis* segments (chromosomes 1, 3, 4, 5) pointed by coffee BAC-end sequence, based on Segmental Genome Duplications of *Arabidopsis* in TIGR database ([http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis\\_genome\\_duplication.shtml](http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis_genome_duplication.shtml)).

Figure 22. Coffee BAC clone contigs assembled using coffee BAC-end sequences (BEs), BAC-ends with non-homologous *Arabidopsis* gene sequences (BESnh) and COS derived probes from segments homoeologous to *Arabidopsis* duplicated chromosome 1, 3, 4 and 5. Circle: hybridization with Coffee probes. Open circle: sequenced BAC ends.

Figure 23. Comparative maps between the four duplicated *Arabidopsis* segments, using the probes (COS and BEs) as ordered on coffee contigs. *Arabidopsis* predicted genes and coffee orthologous counterpart are linked by dashed line.

Figure 24. Restriction fragment comparison (i.e. EcoRI-digest) of Coffee BAC clones belonging to the contig A previously defined based on BAC filter hybridisations. Molecular marker was loaded into every five lane.

Figure 25. Lignées introgressées quasi-isogéniques issus du croisement entre LA716 et M82 rétro-croisé sur M82.

Figure 26. Résultat d'hybridation de la sonde SGN-U215128. Les deux parents sont à gauche, le polymorphisme d'intérêt est présenté avec des flèches.

Figure 27. Evénements de duplication entre le groupe Euasterid et Eurosoid.

Figure 28. Représentation schématique des gènes présents dans les régions synténiques caféier - *Arabidopsis* et le fragment ancestral avant les événements de duplication. Les gènes présents sur *C. arabica* ont été révélés par hybridation des orthologues cafés des gènes d'*Arabidopsis*, et ceux présents chez *Arabidopsis* proviennent du site TIGR ([http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis\\_genome\\_duplication.shtml](http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis_genome_duplication.shtml)). Les carrés de même couleur correspondent à des paralogues.



## Liste des tables

Tableau 1. Exemples de caractères d'intérêt agronomique identifiés au sein d'espèces diploïdes de caféiers ou de formes sauvages de *Coffea arabica*.

Tableau 2. Liste des génotypes de caféier pour l'étude de la rouille du caféier (Bettencourt 1981; Bettencourt et Rodriguez 1988)

Tableau 3. Hybrides interspécifiques spontanés ou artificiels avec *Coffea arabica*

Table 4. List and characterization of identified first-generation hybrids.

Table 5. List of identified *C. canephora* alleles for all analyzed SSR loci. For each allele, the presence in the different populations (Dogny, D; Couli, C; Leonard, L; Szemmelveisz, S; Friquet, F) and mean frequencies are reported.

Table 6. Reaction type (R = resistance and S = Susceptible) of HNC plants inoculated with the race II and different rust isolates that overcome resistance factors introgressed in *C. arabica* from *C. canephora* through the Timor Hybrid.

Table 7. Sequence-characterised DNA markers linked, either in coupling or in repulsion, to the  $S_{H3}$  leaf rust resistance gene

Table 8. Survey of sequence-characterised DNA markers linked to the  $S_{H3}$  gene in 43 individuals (BC2) derived from a cross between the line S.288 carrying the  $S_{H3}$  rust resistance gene (group G) and the susceptible line Matari (group F) used as recurrent parent. + indicates the presence of the marker bands; – indicates the absence.

Table 9. *Arabidopsis thaliana* homologous sequences of coffee BAC-end sequences from the SH3-contig. The identified *Arabidopsis* predicted genes showing co-localisation in close proximity are indicated in bold.

Table 10. COS markers corresponding to predicted genes located on the targeted *Arabidopsis* segments identified. Putative orthologous coffee and tomato sequences and derived primers are indicated.

Table 11. *Arabidopsis thaliana* homologous sequences from BAC clones hybridized by COS markers presenting significant homologies to *Arabidopsis thaliana* sequences (in bold) located on the targeted *Arabidopsis* segments.

Tableau 12. Marqueurs orthologues putatifs de *S. lycopersicum* aux séquences des extrémités de BAC ou de gènes du caféier présent sur le contig portant le caractère de résistance  $S_{H3}$ .

Tableau 13. Résultats d'hybridation sur des lignées introgressées avec les sondes de *S. lycopersicum*. Les enzymes de restriction testés et les positions déterminées sur les ILs sont listés.

**Contribution à l'amélioration génétique de la résistance des caféiers (*Coffea arabica* L.) à la rouille (*Hemileia vastatrix*) : De l'étude des hybrides interspécifiques naturels de Nouvelle-Calédonie à la cartographie d'un locus de résistance.**

**Résumé** La rouille due au champignon *Hemileia vastatrix* est une des contraintes phytoparasitaires majeures à la culture du caféier *Coffea arabica*. L'amélioration génétique de l'espèce allotétraploïde *C. arabica* est basée principalement sur l'exploitation des ressources génétiques des caféiers diploïdes tel que *C. canephora* ou *C. liberica*. Elle consiste à introduire au sein des variétés cultivées, des gènes d'intérêt, notamment de résistance aux maladies, à partir des espèces diploïdes. Les objectifs de ce travail sont en relation avec l'amélioration de *C. arabica* pour sa résistance à la rouille et consistent, d'une part, en la caractérisation de nouvelles sources de diversité génétique et d'autre part, au développement des connaissances moléculaires d'un locus de résistance à la rouille.

La découverte en Nouvelle-Calédonie d'individus de phénotype intermédiaire entre *C. arabica* et *C. canephora* a motivé la recherche de nouvelle diversité génétique pour l'amélioration de *C. arabica*. L'étude de la genèse et des flux de gène chez ces hybrides interspécifiques et l'évaluation de leur résistance pour plusieurs races de rouille ont mis en évidence de nouvelles ressources génétiques pour l'amélioration de *C. arabica*. Parallèlement à cette étude, la cartographie génétique du locus de résistance  $S_{H3}$  de *C. liberica* introgressé chez *C. arabica* a été réalisée ouvrant ainsi la voie à une sélection assistée par marqueurs pour l'amélioration de *C. arabica*.

Dans le but de cloner le (ou les) gènes de résistance à la rouille, la cartographie physique sur *C. arabica* de la région chromosomique homéologue à celle de *C. liberica* portant le facteur  $S_{H3}$  a été réalisée. Les marqueurs liés au locus  $S_{H3}$  et l'analyse génomique comparative avec *Arabidopsis thaliana* ont permis de définir une région d'environ deux mégabases contenant le ou les gènes de résistance homéologue à  $S_{H3}$ . L'établissement de la microsyténie entre le caféier et cette petite crucifère a permis le transfert des marqueurs pour l'avancement de la cartographie physique, et apporté des informations sur l'évolution de ces deux espèces. Finalement pour ouvrir la voie à une approche gène candidat et le transfert de nouveaux marqueurs sur la carte physique de *C. arabica*, la cartographie génétique de la région chromosomique liée à  $S_{H3}$  a été réalisée chez *Solanum lycopersicum*. Plusieurs gènes de résistance sont présents dans les régions orthologues chez *S. lycopersicum* et *Arabidopsis*. L'ensemble de ces résultats permet d'envisager le clonage du (ou des) gène(s) de résistance à la rouille chez le caféier.

**Mots clefs:** *Coffea arabica*, *Coffea liberica*, *Hemileia vastatrix*, Résistance à la rouille, Hybrides interspécifiques naturels, introgression, flux de gènes, marqueurs moléculaires, évolution, *Arabidopsis thaliana*, Génétique comparative, Microsyténie.

---

**Contribution to the genetic improvement of coffee resistance (*Coffea arabica* L.) to rust disease (*Hemileia vastatrix*): From the study of the natural interspecific hybrids of New Caledonia to the mapping of resistance locus.**

**Abstract** Coffee Rust disease due to the pathogenic agent *Hemileia vastatrix* is one of the major obstacles in *Coffea arabica* cultivation. The genetic improvement of the allotetraploid *C. arabica* species is mainly based on the exploitation of the genetic resources of diploid species such as *C. canephora* or *C. liberica*. It consists of introducing genes of interest, particularly disease-resistance genes, into the cultivated varieties. The objectives of this work relate to the improvement of *C. arabica* for its resistance to rust and consist of both the characterization of new sources of genetic diversity, and the development of molecular knowledge of a rust resistance locus.

The discovery in New Caledonia of individuals with intermediate phenotypes between *C. arabica* and *C. canephora* motivated the search for new genetic diversity. The study of genesis and gene flows in these interspecific hybrids combined with the evaluation of their resistance to several rust races highlighted new genetic resources for *C. arabica* improvement. Parallel to this study, the genetic mapping of *C. liberica*  $S_{H3}$  rust-resistance locus introgressed in *C. arabica* was carried out, thus opening the way for a marker-assisted selection. With an aim of cloning the rust-resistance genes, the physical map of *C. arabica* homoeologous chromosomal segments to that of *C. liberica* bearing the  $S_{H3}$  factor was carried out. The markers related to the  $S_{H3}$  locus and the comparative genomic analysis with *Arabidopsis thaliana* allowed defining an area from approximately two megabases carrying homoeologous  $S_{H3}$  resistance genes. The establishment of the microsytény between *Coffea* and *Arabidopsis* allowed the transfer of markers to build the physical map, and revealed information on the evolution of these two species. Finally, to open the way to a candidate gene approach and the transfer of new markers to the physical map, the genetic mapping of the chromosomal area related to  $S_{H3}$  was carried out in *Solanum lycopersicum*. Several resistance genes are present in *S. lycopersicum* and *Arabidopsis* orthologous segments. Together these results make it possible to consider the cloning of rust resistance gene(s) in *Coffea*.

**Keywords:** *Coffea arabica*, *Coffea liberica*, *Hemileia vastatrix*, Rust resistance, Natural interspecific hybrids, introgression, gene flow, molecular markers, Evolution, *Arabidopsis thaliana*, Comparative genetics, Microsytény.