

UNIVERSITE de la NOUVELLE-CALEDONIE

**Ecole Doctorale du Pacifique
Milieux Insulaires Ultra-Marins
ED-469**

Doctorat

BIOLOGIE DES POPULATIONS

Céline GOMEZ

**Etude de l'hybridation inter-spécifique entre espèces
du genre *Coffea* en Nouvelle-Calédonie :
Distribution des niches favorables et structuration de
la diversité**

Thèse dirigée par Perla HAMON

Thèse co-encadrée par Valérie PONCET et Marc DESPINOY

Soutenue le 10 décembre 2009

Jury :

Sophie Gerber
Agnès Bégué
Hamid Amir
Michel Petit
Perla Hamon
Alexandre de Kochko
Valérie Poncet
Marc Despinoy

Chargée de Recherche INRA Bordeaux
Chercheur CIRAD Montpellier
Professeur Université Nouvelle Calédonie
Directeur de Recherche IRD, Montpellier
Directeur de Recherche IRD, Montpellier
Directeur de Recherche IRD, Montpellier
Chargée de Recherche IRD, Montpellier
Ingénieur de Recherche IRD, Nouméa

Rapporteur
Rapporteur
Examineur
Examineur
Directeur de these
Examineur
Co-encadrante
Co-encadrant

Remerciements

Ce projet de thèse est le fruit d'une collaboration entre deux unités de l'Institut de Recherche pour le Développement, l'unité Diversité et Adaptation des Plantes Cultivées (DIAPC) et l'unité Expertise et SPAtialisation des Connaissances en Environnement (ESPACE) et plus particulièrement entre deux équipes, Génomique Comparative et Fonctionnelle de l'Adaptation (GECOFA) et l'équipe ESPACE du Pacifique. Ce projet a été soutenu également par la Province Sud de la Nouvelle-Calédonie et notamment par la Direction du Développement Rural (DDR). Cette thèse a été réalisée à l'Université de la Nouvelle-Calédonie de l'Ecole Doctorale du Pacifique. Merci à ces organismes, laboratoires et université de m'avoir permis de réaliser ce projet de thèse. Merci plus précisément à Serge Hamon, directeur de l'unité DIAPC, à Frédéric Huynh, directeur de l'unité ESPACE, et François Mademba-Sy directeur de la DDR.

Cette thèse a débuté à Montpellier au Centre IRD puis à la Maison de la télédétection et enfin à Nouméa au Centre IRD. Ce fut avant tout un défi personnel que de mener à bien un tel projet, qui a été relevé grâce à l'aide d'un grand nombre de personnes. Que ce soit au travers de l'accueil, de l'écoute, de l'aide technique, de conseils scientifiques, de l'accompagnement sur le terrain, du prêt de matériel...toutes ces personnes m'ont apportés du soutien, des sourires et de précieux moments de réconfort. Que toutes les personnes que j'ai croisées pendant ma thèse et qui m'ont apportées leur soutien trouvent ici ma sincère reconnaissance.

Je tiens à remercier tout d'abord Serge et Perla Hamon, Alexandre de Kochko et Valérie Poncet qui m'ont permis de me lancer dans cette aventure dès Montpellier puis en Nouvelle-Calédonie. Merci aussi à Frédéric Huynh et Marc Despinoy pour avoir rendu possible cette riche collaboration avec l'US ESPACE qui m'a beaucoup appris et qui a grandement contribué au façonnage de ma carrière future sur des approches transversales. Merci à Valérie Poncet et Marc Despinoy pour m'avoir encadrée. Merci à Valérie pour m'avoir fait partager ton savoir et ton expérience. Merci d'avoir été disponible quasiment au quotidien même à 22000km pour m'aider dans les situations délicates. Merci pour ton aide sur le terrain pour les échantillonnages, et les observations des floraisons. Et surtout merci de m'avoir appris ta rigueur scientifique qui a toujours fait ses preuves durant les étapes de ma thèse. Marc, merci pour ton soutien et pour m'avoir accompagnée dans le monde, alors nouveau pour moi, de la télédétection et de la spatialisation. Merci aussi pour ta grande disponibilité et toutes les discussions que nous avons eues dans les moments de doute, ton calme m'a souvent apaisée. Grâce à vous deux, j'ai beaucoup appris.

Je tiens à remercier particulièrement Daniel Le Pierrès, qui est l'initiateur du projet en Nouvelle-Calédonie et de la collaboration avec la DDR. Merci d'avoir partagé tes grandes connaissances du terrain en général.

Je remercie aussi Agnès Bégué et Sophie Gerber d'avoir accepté de juger mon travail en qualité de rapporteurs. Merci aussi à tous les membres du jury, Hamid Amir, Michel Petit, Perla Hamon, Valérie Poncet et Marc Despinoy.

Merci à tous les membres de l'équipe GECOFA et ceux du LATICAL (US ESPACE Nouméa). Un grand merci à Joseph (Edouard Oukewen) pour ta gentillesse, ta générosité, ton talent de jardinier et les sorties terrain bien sympathiques. Merci à Monique Martiny pour ton accueil chaleureux au Latical et ton soutien « maternel » toujours avec un grand sourire. Merci à Claudine

Campa pour ta bonne humeur communicative. Merci à Chloé Batisse et Marc Bournier pour leurs travaux qui ont apporté leur pierre à l'édifice. Merci à tous, Pascal Dumas, Jean-Brice Herrenschmidt, Sylvain Broucke, Marie Couderc, Carole Bessière...et tous ceux que j'oublie.

Je remercie aussi tous les membres de la DDR qui ont accompagné le projet, tant au plan administratif, de communication ou technique également. Merci à Sylvie Edighoffer et Danyéla Salmon pour votre gestion du projet Café. Merci à toutes les personnes de la Pépinière de Port-Laguerre pour votre aide quant aux greffes, au terrain parfois. Je remercie particulièrement Corinne Quinty pour son accueil, sa grande bonne humeur, grâce à toi l'ambiance à la pépinière pétille. Merci à Sandra, Anthony et Manuel.

Je tiens aussi à remercier tout particulièrement l'équipe du LSTM à Nouméa qui m'a chaleureusement accueillie pour partager bureaux, labos, blagues et contrepèteries. Un très grand merci à Michel Lebrun, à qui j'accorde une mention spéciale de discussions où tu penses à voix haute, tu as mis en éveil ma curiosité scientifique en me faisant partager tes points de vue sur des sujets divers et variés. Merci pour ton écoute, ton grand soutien et pour tous ces moments chocs et vive les sessions gazette de bab el oued ! Un grand merci aussi à tous les membres de l'équipe qui ont supporté mes moments de « mou » et qui m'ont fait une place à part entière dans l'équipe : Sylvain, Clarisse, Laure, Marc, Philippe et Laurent. Merci à Clarisse pour ces bons moments, notamment à Tahiti et ton aide précieuse pour les manipulations et les commandes. Merci aussi à Bernard Dreyfus, dont les racines sont au LSTM, pour vous être intéressé à mon projet et m'avoir encouragée.

Je tiens également à remercier tous ceux qui m'ont accompagnée scientifiquement, pour la rédaction des articles, pour la mise au point des manipulations et pour tous les aspects techniques et pratiques.

Merci à Ahmed Batti pour m'avoir initiée aux joies du SIG, je ne regrette pas d'avoir fait ce choix là et je te remercie pour les moments passés à m'expliquer la base de la base avec une grande patience.

Merci à Marc Despinoy pour m'avoir initiée à la télédétection et aux traitements d'images avec beaucoup de pédagogie. Merci d'avoir cru en mes compétences et de m'avoir toujours soutenue. Merci à Morgan Mangeas pour m'avoir appris à réfléchir autrement face à un problème statistique et à envisager la modélisation d'une façon agréable. Grâce à toi, j'ai pris un grand plaisir à mettre au point les modèles et ça m'a permis de faire valoir cela lors de ma recherche de post-doc aussi. Un grand merci à vous deux pour votre bonne humeur et votre aide précieuse dans la dernière ligne droite.

Merci à Stéphane Dussert pour tes conseils pour mettre en forme le tout premier article.

Merci à Michel Petit et Christina Corbane pour avoir porté attention à mon travail et m'avoir conseillée sur des aspects techniques en télédétection et sur la façon de mieux manier cette approche transversale.

Merci à Sylvain Santoni, pour ton passage à Nouméa durant lequel tu nous as donné de précieux conseils pour le géotypage dans la bonne humeur.

Merci à Francis Gallois et Jean-Yves Panché pour votre aide technique précieuse concernant mes petits capteurs météo qui ont connu maintes péripéties et la mise au point de la station météo. Merci pour votre accueil chaleureux. Merci à Jean-Louis Laurent (Loulou) et Fabrice Roubaud pour votre aide précieuse lors de l'installation de la station météo, que ce soit pour le calibrage ou l'installation sur le terrain. Cela m'a vraiment été d'une grande aide !

Merci à David Varillon pour m'avoir prêté des capteurs météo quand j'en avais besoin.

Merci à Sylvain Petek pour m'avoir fait de la place dans le congélateur à -80°C.

Merci à l'équipe de Botanique de l'IRD à Nouméa et à Jérôme Muzinger pour ton expertise sur le terrain. Merci de m'avoir régulièrement prêté l'échenilloir pour les récoltes de matériel, merci à Fred.

Merci à l'équipe informatique de l'IRD à Nouméa aussi pour votre aide régulière quand les soucis informatiques ont pointé leur nez.

Merci au personnel du garage de l'IRD à Nouméa.

Merci aussi à tous les membres de l'administration qui me sont venus en aide à plusieurs reprises pour des commandes, des missions. Merci à Marie-Noëlle Segura, Elisabeth Misimoa, Véronique Perrin, Catherine Hartman pour votre aide et votre gentillesse.

Enfin merci à tous ceux qui ont égayé ses longues journées de travail. Merci Hervé pour les séances café ou commérage dans les couloirs ! Merci à mes super « collocatrices » de bureau, Lucie puis Vanesse et enfin Marion, Amélie et Laurent ! Merci à Alex et Nico aussi. Merci à mes collègues thésards avec qui on a souvent partagé nos doutes et nos expériences, merci à Mariko, Anne-So, Bastien, Haizea, Alex, Nico, Laurette, Yohan et Charly! Merci pour les pauses café avec la bande de joyeux lurons, Marcel, Jacob, Joseph, William, Denise et Jenny ! Merci à Djoël pour tes conseils sur la vie tout simplement ! Merci aussi à la bande des africains de Montpellier, vous avez égayé mes journées de dures rédaction, merci Gustave et Deless !

Un grand clin d'œil à tous les amis que j'ai rencontré au cours du chemin en Calédonie et qui m'ont apporté toujours un petit quelque chose à un moment particulier...Milie, Etienne, Jenny, Vanesse, Lucie, Marion, Mat, Amélie, Julien, Oror, Christou, Wadrocal, Shirleen ; Sylvain...Et aussi à ceux qui m'ont toujours soutenu même à 22 000 km... Didine, May, Coco, Magali, Sandra, Marion, ma pronto, Thibaut, Camille, Shakia, Gwen, Nico, Jérôme, Chanti, Chandra, Manu, Pauline...Un profond merci et une spéciale dédicace à Magali, Sandra et Didine pour votre soutien vraiment réconfortant, je vous adore. Merci aux trois cachotiers pour nous avoir fait cette surprise, c'est finalement bien tombé, car vous avez pu m'apporter votre soutien au moment le plus dur.

Un petit mot aussi pour la petite famille Kawa de Petit-Couli pour votre accueil toujours aussi chaleureux. Merci Sonia et Cyprien de m'avoir indiqué des « coins à café » que je ne connaissais pas. Merci à Nounou Edmont, Tony, Niva, Tania, Vany, Yohan, et Sonia, de votre accueil encore et toujours. De belles pensées aux p'tits loups de la famille !

Si je suis arrivée jusqu'ici c'est surtout grâce à vous deux, merci papa, merci maman, du fond du cœur pour tout votre amour, votre soutien constant...pour tout, jusqu'au bout! Merci à toute la family, merci mes sœurs, merci mes cousins, les tontons et les tatas ! Merci ma pronto et ma tataaaaaaaa pour toutes ses belles discussions sur la vie ! Merci à tous, Gignac aura été une escale importante sur la fin de la thèse, on s'y ressource toujours, je pense que tout le monde est d'accord là-dessus !

Et pour finir merci à toi mon Pierro, pour m'avoir accompagnée tout au long de ce défi. Merci pour tout...



La Caf  rie

*Quand le brillant soleil tropical resplendit
 Quand le chant cigalier si trident retentit
 Il est bien doux d'entrer, comme en r  verie,
 Dans la myst  rieruse et fra  che caf  rie*

*Le petit cerisier amoureux des ombrages,
 Charmant arbuste de rapport et d'ornement,
 Le timide arbrisseau qui craint tant les orages
 Mais qui portent en ses fruits des grains d'enchantement*

*   riches caf  iers reluisants de verdure,
 Que vous   tes jolis ! Quand vient la floraison
 Vous vous couvrez de neige,    neige sans froidure,
 Odorante et portant l'espoir de ma maison*

*Et que j'aime vous voir ployant sous vos cerises
 Qui passent si gaiement du vert aux tons diapr  s,
 Pour m  rir en jetant leurs   clats empourpr  s,
 Sous les sombres bois noirs aux   l  gantes frises.*

*Antoine SouryLavergne
 T  moignage, Les morceaux cal  doniens*

Tables des Matières

Introduction	8
I- Cadre général de l'étude et objectifs	9
II – Organisation de la thèse.....	10
 Chapitre 1 – Revue Bibliographique	 12
I – Les bases de la compréhension des patrons de biodiversité	13
1.1 Notions de diversité.....	13
1.2 L'adaptation écologique des espèces	18
II- Notions de biogéographie.....	24
2.1 Modèle de prédiction de distribution des habitats.....	24
2.2 La multiplicité des échelles d'observation	26
2.3 L'application à la structuration des espèces	28
III- Cas des espèces introduites et de leur devenir dans les écosystèmes	31
3.1 Impact anthropique généralisé	31
3.2 Les conséquences : de l'extinction à l'invasion	33
3.3 Clés de compréhension de l'adaptation.....	34
IV- Questions fondamentales abordées.....	36
 Chapitre 2 - Présentation du modèle d'étude.....	 37
I – Le café : origine Afrique.....	40
1.1 Connaissances générales	40
1.2 Caractéristiques génétiques des trois espèces cultivées	45
II- Le café en Nouvelle Calédonie	53
2.1 Historique des introductions et de la culture du café	53
2.2 L'abandon des cultures et la mise en place des zones hybrides à Sarraméa.....	55
III- Objectifs de l'étude	59
 Chapitre 3 - La caractérisation de la structure génétique et de la distribution naturelle de l'espèce <i>Coffea canephora</i> Pierre.....	 60
I – Contexte de l'étude et démarche abordée	61
1.1 – Contexte historique de l'étude.....	61
1.2 - Marqueurs moléculaires.....	62

II – Résultats principaux (<i>Gomez et al. 2009 BMC Evolutionary Biology</i>).....	83
III – Conclusion.....	83
Chapitre 4 - La mise en évidence d'un habitat favorable aux hybridations inter-spécifiques naturelles : détection génétique d'hybrides et analyses environnementales.....	84
I – Contexte de l'étude et démarche abordée	85
1.1 – Diversité introduite et caractérisation des hybrides	86
1.2 – Comment la levée de barrières à la reproduction s'opère t- elle ?	88
1.3 – Importance de caractériser les conditions environnementales favorables.....	94
III– Résultats principaux (<i>Gomez et al. Journal of Applied Ecology</i>)	96
III– Conclusion.....	112
Chapitre 5 - Prédiction de distribution des niches favorables aux hybridations inter-spécifiques entre caféiers : caractérisation fine du couvert forestier et modèle environnemental.....	113
I – Contexte de l'étude et démarche abordée	114
1.1 - Cas particulier des plantations sous couvert	115
1.2 - La caractérisation fine de la canopée	116
1.3 - Mise en place d'un modèle de prédiction	117
II – Résultats principaux (<i>Gomez et al. Remote Sensing of Environment</i>)	117
III - Conclusion	145
Chapitre 6 – Conclusion générale.....	146
I – Synthèse et discussion.....	147
II - Applications	154
III- Perspectives de recherche	156
Références bibliographiques	158

Introduction



I- Cadre général de l'étude et objectifs

Les caféiers sont tous originaires d'Afrique, et seules trois espèces (*Coffea canephora*, *C. arabica* et *C. liberica*) sont utilisées pour la production de café. Ces trois espèces ont été introduites à travers la majorité des pays situés dans la bande intertropicale en Amérique du Sud, en Asie et en Océanie. Elles ont été notamment introduites en Nouvelle-Calédonie dès la fin du 19^{ème} siècle pour leur culture. Or, ces cultures, principalement celles réalisées sous couvert, ont été progressivement abandonnées pour des raisons majoritairement économiques, contraignant les caféiers à survivre dans les conditions naturelles locales.

Les barrières naturelles à la reproduction entre ces trois espèces sont très fortes dans leurs milieux d'origine naturels (le continent africain) : outre le fait qu'elles présentent des aires de répartitions et des exigences climatiques distinctes, un décalage de phénologie renforce leur isolement génétique. Introduites en Nouvelle-Calédonie, ces trois espèces, génétiquement proches, se retrouvent ainsi contraintes à cohabiter dans un milieu dont les conditions environnementales semblent jouer en faveur d'une levée de barrières à la reproduction. En effet, d'après des observations, réalisées par Daniel Le Pierrès dès 1991 sur l'ensemble de l'île, des hybrides inter-spécifiques sont apparus spontanément dans la région centrale et semblent très adaptés à l'environnement qui les a vus naître.

L'objectif de cette thèse est d'étudier les effets de l'introduction de plantes dans un nouvel environnement et de comprendre l'ampleur des levées de barrières à la reproduction en termes de taux d'hybridation et de distribution spatiale.

Autour de cet objectif, s'articulent les différentes questions fondamentales que soulève l'étude d'une telle dynamique après un déplacement d'espèces. Elles concernent la dynamique des niches écologiques et son impact sur les espèces et la caractérisation des facteurs environnementaux favorables ayant autorisé cette hybridation unique au monde. Ceci se fera non seulement à l'échelle de la population mais aussi à une échelle plus large afin de comprendre si le phénomène observé est très localisé ou, *a contrario*, généralisé.

Pour rendre ceci possible, deux approches combinées et complémentaires nous permettant de caractériser le phénomène dans son ensemble ont été abordées :

- l'analyse génétique des populations, sur la base de marqueurs microsatellites (analyses effectuées au niveau des espèces en Afrique et au niveau d'une population tri-spécifique de référence en Nouvelle-Calédonie)
- l'analyse spatiale intégrée, sur la base de systèmes d'information géographiques combinant toutes les variables environnementales et spatiales nécessaires à la compréhension d'un tel phénomène à une échelle régionale (la région de l'étude s'étend sur presque 200km² et présente une importante concentration d'anciennes plantations de caféiers multi-spécifiques présentes dans des habitats bien définis).

II – Organisation de la thèse

Ce mémoire de thèse est structuré en six chapitres.

L'état de l'art est présenté dans le **chapitre 1**. Le modèle d'étude est présenté dans le **chapitre 2**. Ce dernier englobe la description des espèces étudiées, leurs exigences environnementales, leurs distributions naturelles, la dynamique observée en Nouvelle-Calédonie. Ces deux chapitres apportent un cadre au questionnement scientifique et mettent en avant les avancées thématiques et méthodologiques que ce travail transversal engendre.

Le **chapitre 3** présente la première étape de la démarche scientifique qui consiste à décrire la diversité naturelle de chaque espèce en Afrique afin d'évaluer la diversité génétique introduite pour chacune des espèces.

La deuxième étape est la caractérisation génétique des hybrides inter-spécifiques en Nouvelle-Calédonie dans la population tri-spécifique témoin afin de révéler le taux d'hybridation général détecté. Or, pour que des hybridations inter-spécifiques puissent se produire, une levée de barrières reproductives entre espèces doit avoir lieu. L'hypothèse principale réside dans la dynamique des phénologies florales qui seraient perturbées par rapport aux phénologies en Afrique, sous l'influence d'un régime pluviométrique particulier. L'étape suivante est la caractérisation des zones favorables aux hybridations en termes de combinaisons environnementales particulières, sur la base de la compréhension du lien entre phénologies florales et conditions climatiques. Ces deux étapes sont très complémentaires car la mise en évidence de la nature exceptionnelle des habitats présents dans la région de Sarraméa passe non seulement pas la détection génétique des hybrides inter-spécifiques,

mais aussi par la caractérisation des conditions environnementales permettant la cohabitation des espèces et la levée de barrières à la reproduction. L'ensemble de ce travail transversal qui met en avant la modification des niches écologiques, est traité dans le **chapitre 4**.

Enfin la dernière étape, consiste à comprendre l'ampleur géographique du phénomène en caractérisant précisément ces niches favorables et en estimant mathématiquement leur distribution à une résolution spatiale fine. Il s'agit de révéler l'hétérogénéité des habitats dans une région favorable dans son ensemble, afin de détecter des micro-habitats hautement favorables. Ce travail est abordé dans le **chapitre 5**.

Enfin, dans le dernier chapitre, **chapitre 6**, les résultats sont analysés et discutés afin de fournir des éléments de réponses sur la levée de barrières reproductives rendant possible les hybridations entre caféiers et sur les facteurs environnementaux qui l'ont permis. Une conclusion et des perspectives sont proposées.

Chapitre 1 – Revue Bibliographique



I – Les bases de la compréhension des patrons de biodiversité

Les écosystèmes sont le théâtre de l'évolution de la vie. Ils voient naître les organismes, influencent leur structuration en entités biologiques, permettent leur évolution et causent parfois leur extinction. Sur la planète terre, il existe une incroyable diversité d'écosystèmes. Ils sont caractérisés par leurs sols, leurs climats et par la diversité des interactions abiotiques et biotiques qui les font exister. Les organismes qui peuplent ces écosystèmes sont en perpétuel mouvement et en perpétuelle évolution.

La diversité du monde vivant englobe beaucoup de notions, de la diversité écosystémique à la diversité génétique en passant par la diversité spécifique. Les espèces sont au carrefour de toutes ces notions. Pour comprendre le fonctionnement de la dynamique du vivant sur terre, il faut se pencher dans un premier temps sur ce qu'est une espèce et quels sont les fondements de la dynamique des espèces. Dans un deuxième temps, il faut comprendre comment les espèces s'adaptent à leur environnement et quelles sont les bases de leur distribution sur notre planète. Nous aurons alors planté le décor du grand théâtre de la dynamique de la biodiversité sur la planète Terre.

1.1 Notions de diversité

1.1.1 Le concept d'espèce et la spéciation

Depuis toujours les scientifiques se sont chargés de classer le vivant en règnes, ordres, familles, genres et espèces. L'espèce est la plus petite entité reconnue, le taxon à la base de la systématique. Le concept d'espèce a traversé les âges et deux douzaines de définitions différentes ont été énoncées (Hey, 2001). Pourtant entre Darwin, Mayr ou Simpson, les notions abordées pour définir les espèces sont différentes.

Dobzhansky proposa son propre concept d'inter-croisement (Dobzhansky, 1937) qui fut popularisé ensuite par Mayr comme le « concept biologique de l'espèce » appelé ainsi parce que l'inter-croisement est alors considéré comme leur seule réalité biologique reconnue (Mayr, 1982). Pourtant la mise en place de ce concept a surtout suscité des critiques plutôt que de permettre un réel consensus sur la question.

Tableau 1.1 : *Le concept biologique de l'espèce et les alternatives proposées (issue de Coyne et Orr 2004 – p27)*

Basis of concept	Concept	Definition
1. Interbreeding	Biological Species Concept (BSC)	Species are groups of interbreeding natural populations that are reproductively isolated from others such groups (Mayr, 1942)
2. Genetic or phenotypic cohesion	Genotypic Cluster Species Concept (GCSC)	A species is a morphologically or genetically distinguishable group of individuals that has few or no intermediates when in contact with other such cluster (Mallet, 1995)
	Recognition Species Concept (RSC)	A species is that most inclusive population of individual biparental organism which shares a common fertilization system (Paterson, 1985)
	Cohesion Species Concept (CSC)	A species is the most inclusive population of individuals having potential for phenotypic cohesion through intrinsic cohesion mechanisms (Templeton, 1989)
3. Evolutionary cohesion	Ecological Species Concept (EcSC)	A species is a lineage which occupies an adaptive zone minimally different from that of any other lineage in its range and which evolves separately from all lineage outside the range (van Valen, 1976)
	Evolutionary Species concept (EvSC)	A species is a single lineage of ancestral descendant populations or organisms which maintains its identity from other such lineages and which has its own evolutionary tendencies and historical fate (Simpson, 1961; Wiley, 1978)
4. Evolutionary history	Phylogenetic Species Concept 1 (PsC1)	A phylogenetic species is an irreducible cluster of organisms that is diagnosably distinct from other such clusters, and within which there is a paternal pattern of ancestry and descent (Cracraft, 1989)
	Phylogenetic Species Concept 2 (PsC2)	A species is the smallest exclusive monophyletic group of common ancestry (De Queiroz, Donoghue, 1988)
	Phylogenetic Species Concept 3 (PsC3) or genealogical Species Concept (GSC)	A species is a basal, exclusive group of organisms all of whose genes coalesce more recently with each other than with those of any organisms outside the group, and that contains no exclusive group within it (Shaw, 1998)

La première critique a émergé de Simpson (1961) qui propose alors le concept évolutif de l'espèce, basé sur le rôle évolutif d'une espèce dans son environnement. Plusieurs autres ont suivi et tous les concepts sont basés sur des propriétés biologiques différentes (Tableau 1.1). Par exemple, alors que le concept biologique de l'espèce se base sur les propriétés d'isolement reproducteur (Mayr, 1942), le concept écologique de l'espèce se centre sur la problématique d'occupation de niches écologiques distinctes ou dans une zone d'adaptation (van Valen, 1976).

Finalement, la réconciliation des différents concepts passe par la reconnaissance d'un concept plus général que partagent tous les concepts déjà existants (de Queiroz, 2005). Il s'agit du concept lié aux métapopulations (« The metapopulation Lineage Concept of Species »). Les espèces sont communément assimilées à des groupes de populations interconnectés qui constituent une communauté reproductrice et un pool de gène unitaire. Intégrer la notion de métapopulation permet de prendre en compte la structuration de la diversité au sein d'une espèce. De l'individu à l'espèce, tous les gradients de différenciation génétique et écologique sont de ce fait pris en compte. Dans un élan similaire, Templeton (1989) décrit le concept de cohésion en soulignant qu'une combinaison entre la cohésion écologique et reproductrice est importante pour maintenir l'intégrité évolutive des espèces (ce qui rejoint la définition de Simpson).

Dans la même dynamique, il existe plusieurs théories quant à la spéciation qui rejoignent les différents concepts d'espèce. Les formations de nouvelles espèces sont sous l'influence des mêmes pressions évolutives qui causent des changements génétiques au sein des populations (Coyne, Orr, 2004). La théorie des métapopulations permet de comprendre les impacts de chaque force évolutive et leurs impacts couplés (Hanski, 1998; Hastings, Harrison, 1994; Husband, Barrett, 1996). Les forces évolutives sont la sélection naturelle, la dérive génétique, la sélection sexuelle, la mutation et la migration.

Le processus de spéciation s'insère dans cette dynamique évolutive. Les différents modes de spéciation, sous l'influence des pressions évolutives, sont décrits en fonction du contexte géographique du processus de divergence des populations pour les conduire jusqu'à la spéciation (allopatrie, sympatrie ou parapatricie).

Le mode de spéciation le plus classique est la spéciation allopatrique. Intuitivement, il est facile de conceptualiser des barrières géographiques physiques qui restreignent les flux de gènes entre deux populations et qui guident alors les populations sur deux chemins évolutifs différents par l'influence couplée des mutations, de la dérive génétique et de la sélection sexuelle (Turelli *et al.*, 2001). A l'inverse de la spéciation allopatrique, il existe la spéciation sympatrique qui est entièrement indépendante du contexte géographique. La nouvelle espèce occupe les mêmes zones géographiques

que l'espèce ancestrale et se forme au travers de la sélection naturelle divergente à l'adaptation à un habitat alternatif (théorie de la divergence des niches écologiques) (Kawecki, 1997). A l'intermédiaire des deux modes de spéciation précédents, le troisième mode est la spéciation parapatrique. Les espèces n'ont qu'une partie de leur habitat en commun. Les flux de gènes entre les deux espèces sont intermédiaires entre 0% (entre deux populations allopatriques) et 100% (entre deux populations sympatriques). La spéciation est alors due à une balance entre flux de gènes et sélection divergente.

C'est dans ce cadre là que les zones hybrides ont été étudiées initialement (Barton, Hewitt, 1985). L'hybridation est une des dynamiques qui suscite encore et toujours beaucoup d'attention dans la mesure où c'est au travers de ce phénomène qu'il est possible d'étudier toutes les situations entre les deux processus extrêmes, l'un menant à la spéciation hybride et l'autre au renforcement des barrières entre deux espèces.

1.1.2 Isolement reproducteur et barrières à la reproduction

Le questionnement sur la spéciation a mené à s'intéresser aux phénomènes d'isolement reproducteur au travers des barrières à la reproduction. L'étude de la dynamique des barrières à la reproduction est une des clés permettant de reconstruire l'histoire de la spéciation (Orr, Turelli, 2001; Rieseberg *et al.*, 1999b). En effet, durant les processus de spéciation, des barrières à la reproduction se mettent en place. Les barrières reproductives sont classées en fonction de l'étape du cycle de reproduction à laquelle elles interviennent : avant ou après la formation du zygote (fécondation).

L'isolement pré-zygotique inclut l'isolement pré-reproducteur et post-reproducteur (toujours avant la fécondation) : les barrières spatio-temporelles et comportementales ainsi que les compétitions gamétiques et les incompatibilités. Les barrières pré-zygotiques sont souvent décrites comme plus fortes que les barrières post-zygotiques (Kay, 2006; Lowry *et al.*, 2009). Elles sont notamment caractérisées par des différences comportementales chez les animaux (Grant, Grant, 1998). Chez les plantes à fleurs, l'étape de la pollinisation et de la floraison sont des étapes clés pour la reproduction et les barrières interviennent souvent à ces moments là. Les isolements dus aux habitats ou aux variations temporelles de la floraison ont été les premiers à être étudiés comme chez *Arabidopsis thaliana* par exemple (Kuittinen *et al.*, 1997). Beaucoup d'études ont été menées sur les iris quant à la compétition pollinique et le rôle des taux de croissance des tubes polliniques sur l'isolement des espèces du genre *Iris* (Carney, Arnold, 1997; Carney *et al.*, 1994; Carney *et al.*, 1996).

L'isolement post-zygotique concerne à la fois les caractères de viabilité et de fertilité des hybrides. Ceci inclut la mortalité des hybrides ou leur stérilité directe mais aussi la baisse des niveaux de viabilité et fertilité des hybrides au cours des générations suivantes. La théorie du renforcement suppose que si les hybrides ont des niveaux faibles de fertilité et de viabilité alors cela pourra renforcer les barrières d'isolement (Servedio, Noor, 2003). Les réarrangements chromosomiques apparaissent comme l'une des principales barrières à l'échange de gènes comme étudiés chez le Tournesol (Rieseberg *et al.*, 1996). Les raisons de la stérilité ou de mortalité des hybrides ont été recherchées dans les interactions négatives entre deux locus divergents. Plusieurs études ont montré l'effet majeur de deux gènes sur la stérilité des hybrides, comme chez le riz (Li *et al.*, 1997) ou le blé (Hermson, 1963). Ceci relève de la théorie de Dobzhansky et Mueller quant à l'épistasie négative entre deux gènes (Dobzhansky, 1940).

Dans les deux cas d'isolement reproducteur, la part environnementale est forte. Les zones hybrides sont souvent décrites comme des laboratoires naturels pour étudier les barrières aux flux de gènes entre espèces (Barton, Hewitt, 1985).

1.1.3 L'hybridation comme moteur de l'adaptation

Les processus de spéciation peuvent se dérouler alors que des flux de gènes entre les populations divergentes existent. Les conséquences de ces flux de gènes dépendent de beaucoup de facteurs, comme la force de la sélection divergente ou l'ampleur des flux de gènes (Rieseberg *et al.*, 2003).

Pendant longtemps, l'hybridation est restée peu étudiée car l'hypothèse principale consistait à dire que les hybrides sont des intermédiaires entre les deux espèces parentales et ne peuvent donc avoir aucun impact évolutif. Seules quelques études parlaient dès lors de la contribution de l'hybridation à l'évolution adaptative (Arnold, 1997; Grant, Grant, 1998; Lewontin, Birch, 1966). Trois types de modèles ont été discutés quant au fonctionnement et à la dynamique des zones hybrides. La plupart des modèles estiment que les hybrides ont un pouvoir adaptatif plus faible que leur parent car ils seraient contre sélectionnés par l'action des barrières post-zygotiques (Barton, Hewitt, 1985). D'autres supposent que les hybrides peuvent surpasser leurs parents dans certaines zones bien particulières de la zone hybride (Moore, 1977). Enfin les derniers donnent une importance évolutive aux hybrides en supposant que l'hybridation est source de nouveauté évolutive (Arnold, 1997).

L'hybridation peut être un moteur à l'évolution au travers de l'introgession. L'introgession illustre le mélange des génomes par l'introduction d'un gène dans le génome d'une autre espèce au grès des retro-croisements. Ceci doit potentiellement contribuer à l'augmentation de la variabilité génétique et à l'introduction de nouvelles combinaisons géniques (Arnold, 2006). Selon les gènes introgressés, le

potentiel adaptatif de l'hybride varie. Rieseberg (1999a) a étudié les cas d'hybridation à fort potentiel évolutif par la formation d'hybrides transgressifs. Il apparaît que l'observation de phénotypes extrêmes dans les populations d'hybrides est fréquente, montrant que l'hybridation peut permettre des adaptations nouvelles rapides. Ceci peut alors être une explication simple de la divergence des niches et de l'émergence de nouveautés phénotypiques souvent associées aux lignées hybrides dans les zones hybrides (Abbott, 1992; Ellstrand, Schierenbeck, 2000; Rieseberg, 1997).

Les impacts évolutifs de l'hybridation sont nombreux et donnent des clés de compréhension de la dynamique d'adaptation des espèces (Barton, 2001; Seehausen, 2004). Par exemple, lors de changements environnementaux soudains, une levée de barrière à la reproduction peut se dérouler et le taux d'hybridation entre les deux espèces peut alors être très élevé, réalisant une « spéciation inverse » (Seehausen *et al.*, 2008). L'hybridation peut alors créer un mélange généralisé d'espèces (Taylor *et al.*, 2006) mais peut aussi être plus limité dans d'autres cas. Il faut bien sûr rester vigilant, car le fait qu'un génotype hybride semble adapté, peut être un phénomène transitoire sous les effets directs de la recombinaison (Barton, 2001).

La spéciation hybride est une notion qui émerge de l'observation des zones hybrides (Grant, 1981). Il existe la spéciation allopolyploïde où le génome de l'hybride formé est composé des génomes des espèces parentales. Ceci est notamment le cas de l'espèce *C. arabica* allotétraploïde. Il existe aussi la spéciation homoploïde sans changements du niveau de ploïdie (Buerkle *et al.*, 2000). Les espèces isolées par des barrières de stérilité chromosomiques peuvent donner naissance à une nouvelle espèce isolée des espèces parentales par la recombinaison favorable des facteurs de stérilité (Avisé, 2004). Une conséquence des hybridations est la spéciation hybride qui implique un isolement reproductif soudain des hybrides soit par des réarrangements chromosomiques rapides, soit par une divergence écologique, soit par une séparation spatiale (Buerkle *et al.*, 2000). Rieseberg *et al.* (2003) ont pu confirmer que l'hybridation avait un rôle évolutif important dans l'adaptation, au travers de la comparaison phénotypique et génotypique d'hybrides anciens et d'hybrides synthétiques.

La divergence écologique peut s'étudier sur la base de la théorie des niches et de leur dynamique.

1.2 L'adaptation écologique des espèces

Les espèces naissent et s'éteignent selon une dynamique qui dépend des pressions évolutives, des modalités biogéographiques des milieux et des caractéristiques écologiques en termes d'interactions biotiques et abiotiques. L'apport de la composante écologique dans le système d'évolution des espèces est primordial dans la mesure où l'écologie fait le lien entre la biologie des espèces et

l'environnement. Ce sont les bases de la compréhension de l'adaptation des espèces. Les mesures de la biodiversité passent inévitablement par une description des milieux et par une caractérisation de leur hétérogénéité spatiale.

1.2.1 Théories des niches écologiques et leur dynamique naturelle

On considère qu'une espèce est adaptée à son milieu si elle a une croissance positive en termes de descendants. En effet, une des définitions de la valeur adaptative (ou fitness) d'un organisme vivant consiste à dire qu'il est adapté si le milieu et les conditions environnementales lui permettent d'avoir un régime de reproduction efficace lui assurant une progéniture viable et fertile. Pour comprendre l'adaptation des espèces à leur milieu et comment celles-ci se maintiennent ou évoluent en cas de changement de milieu, il est important de se pencher sur la théorie des niches écologiques, développée depuis 1957.

Hutchinson (1957) a été le premier à présenter le concept de niche écologique (Encadré 1.1). Il a distingué dès lors les niches fondamentales, génétiquement et physiologiquement déterminées, des niches réalisées qui incluent la notion écologique du concept. Il appelle niche fondamentale l'ensemble des exigences d'une espèce pour maintenir une croissance positive sans prendre en considération les interactions biotiques. Ceci inclut les exigences écologiques, physiologiques et reproductives. La niche réalisée est la partie de la niche fondamentale dans laquelle l'espèce a toujours une croissance positive étant données les contraintes biotiques rencontrées (prédation, compétition, parasitisme). En effet, comme l'ont montré Volterra et Gause (1934), deux espèces utilisant des ressources communes sous des conditions constantes et dans un milieu limité ne peuvent pas coexister. En conséquence de cette compétition, la niche réalisée est présentée comme étant toujours réduite par rapport à la niche fondamentale, dans la mesure où une espèce n'occupe pas la partie de sa niche fondamentale où les pressions de compétition sont trop fortes. La niche réalisée correspond à la niche observée dans la nature (Silvertown, 2004).

Pourtant, par la suite, les bases de la réflexion écologique sur les impacts des changements des habitats ont été exposées par Holt (1985) puis Pulliam (1988). Ils ont mis en place la théorie appelée « la dynamique source-puits » qui illustre comment la différence dans la qualité des habitats peut affecter la croissance ou la survie d'une espèce. Ils montrent que si une espèce colonise un nouveau milieu qui lui est moins favorable (puits) à partir de son milieu d'origine (source), la niche fondamentale est alors réduite mais la niche réalisée peut être plus large si les pressions des interactions biotiques y sont limitées (Pulliam, 1988). Plus tard, il a été montré que bien que la compétition inter-spécifique ait une influence notable sur la distribution des espèces, d'autres

facteurs comme la dispersion notamment peuvent contrecarrer les effets de la compétition (Pulliam, 2000).

Cette théorie des niches permet d'illustrer la combinaison des facteurs environnementaux ou abiotiques ainsi que des interactions biotiques sur la distribution des espèces. Il existe une deuxième théorie des niches plutôt basée sur les réseaux trophiques en considérant l'impact d'une espèce sur le milieu appelée niche trophique (Austin *et al.*, 1990; Silvertown, 2004).

Ce qui transparaît de ces études montre qu'il est important de considérer à la fois les conditions climatiques et abiotiques d'un milieu mais aussi les changements dans les pressions biotiques pour comprendre la dynamique de distribution des espèces.

Encadré 1.1 : Glossaire

Niche fondamentale : exigences d'une espèce pour maintenir un taux de croissance positif au niveau de la population, sans tenir compte des interactions biotiques.

Niche réalisée : part de la niche fondamentale où l'espèce a une croissance positive étant donné des contraintes biotiques rencontrées telles que la compétition.

Niche trophique : ensemble des dimensions de la niche écologique liées à l'alimentation (modèles proies-prédateurs).

Niche environnementale : ensemble des conditions environnementales liées aux exigences physiologiques d'une espèce pour maintenir un niveau croissance positif.

Niche climatique : aspect de la niche environnementale défini par les limites de la variation climatique.

La taille des niches, la dispersion et la disponibilité des habitats contribuent à la relation entre niche et distribution. Pulliam *et al.* (2000) ont mis en place un modèle simulant les dynamiques des niches et des populations d'une ou plusieurs espèces dans un paysage complexe. Ce modèle a clairement montré que beaucoup de facteurs étaient impliqués dans la distribution des espèces comme la compétition, la dispersion, la taille des niches et la distribution des conditions environnementales dans le temps et dans l'espace.

Les niches et la distribution des espèces peuvent être reliées selon quatre théories écologiques (Pulliam, 2000) : (1) la niche « Grinnellian » qui considère qu'une espèce est présente partout où les conditions lui sont favorables, (2) la niche réalisée qui considère qu'une espèce est exclue de sa niche fondamentale par un compétiteur supérieur, (3) la dynamique « source-puits » qui considère

qu'une espèce peut être présente dans une niche où son taux de croissance est négatif (puits) seulement si un processus constant de migration existe de la source vers le puits et (4) la situation où la dispersion est le facteur limitant qui considère qu'une espèce est souvent absente d'habitats favorables car son taux d'extinction est supérieur à son taux de dispersion. Finalement, pour comprendre la distribution des espèces, il faut intégrer à la fois la théorie des niches de Hutchinson, la théorie des métapopulations et la théorie source-puits.

1.2.2 La théorie de conservation des niches

Le cas de conservation des niches a été démontré à plusieurs reprises sur des modèles biologiques distincts (Pearman *et al.*, 2008b) (Tableau 1.2). Par exemple, Huntley et Webb (1989) ont démontré qu'il était possible de prédire la distribution du hêtre (*Fagus L.*) en Europe en se basant sur les conditions climatiques de sa région de distribution en Amérique du Nord et inversement. Plus récemment, les prédictions géographiques croisées d'espèces sœurs (oiseaux, papillons et mammifères) basées sur leurs niches environnementales respectives, ont montré une forte conservation des niches entre espèces (Peterson *et al.*, 1999). Par ailleurs, Peterson *et al.* (1999) ont aussi montré une faible conservation des niches à l'échelle de la famille qui pourrait indiquer que les phénomènes de spéciation se déroulent dans des dimensions géographiques et que la différenciation écologique apparaît plus tard (ce qui fait échos à la spéciation allopatrique). Un faible changement de niche est aussi attendu quand le taux d'adaptation de l'espèce déplacée est plus bas que son taux d'extinction ce qui est initialement dû à la hauteur des flux de gènes entre les populations « source » et « puits » (Pulliam, 2000).

Tableau 1.2 : Synthèse des études menées quant à la conservation ou la dynamique des niches (issue de Pearman *et al.*, 2008a)

Topic	System	Conclusions	References
Niche conservatism	Beech (<i>Fagus</i>)	Niche conservatism suggested by reciprocal modelling of beech in North America and Europe	(Huntley, Webb, 1989)
	Birds, mammals, butterflies	Niche conservatism displayed by sister taxa with adjacent distributions	(Peterson <i>et al.</i> , 1999)
Niche conservatism and shifts	Sundews (<i>Drosera</i>)	Niche shifts shown by clades, described by ancestral climate envelope reconstruction	(Yesson, Culham, 2006)
Niche shift	Plant phenology	Radiation to new habitats is often accompanied by altered phenology	(Levin, 2006)
	Plant invasions	Ecological and evolutionary processes change during successive stages of plant invasions	(Dietz, Edwards, 2006)
	Invasive species	Niche shift shown by lack of reciprocal predictability between native and invaded ranges	(Broennimann <i>et al.</i> , 2007; Fitzpatrick <i>et al.</i> , 2007)
	Introduced plant	Introduced populations rapidly adapt to conditions at local latitude	(Maron <i>et al.</i> , 2004)
Niche shift and speciation	Dendrobatid frogs	Divergence in climatic niche occurred during speciation of Ecuadorian frogs	(Graham <i>et al.</i> , 2004)

1.2.3 La dynamique des niches : « niche shifts »

Des changements de niches peuvent se produire dans un grand nombre de situations différentes, que ce soit lors d'introductions de plantes dans un nouvel environnement, lors de phénomènes d'invasion biologique, ou lors de phénomènes de spéciation et de divergence de populations (Tableau 1.2) (Pearman *et al.*, 2008b). Broennimann *et al.* (2007) ont montré qu'une espèce envahissante peut occuper une niche climatique différente après son introduction dans une nouvelle région. De la même façon, le cas des populations introduites montre que ces populations s'adaptent rapidement aux conditions locales (changement d'écosystème, changement de latitude) (Maron *et al.*, 2004). D'autre part, il a été montré que l'adaptation rapide à de nouveaux habitats peut être associée à une altération de la phénologie, florale notamment (Levin, 2006). A une échelle évolutive plus large, il existe des modèles théoriques de spéciation qui ont suggéré des changements rapides de niches (Orr,

Smith, 1998). Ceci fait écho à des phénomènes de spéciation sympatrique ou parapatrique (Dieckmann, Doebeli, 1999).

Ces changements rapides sont dus à une adaptation progressive, facilitée par des flux de gènes et des possibilités de nouvelles combinaisons génétiques permettant l'expansion de la niche écologique (Holt *et al.*, 2004). La théorie de modification de niche est présentée ci-dessous (Figure 1.1).

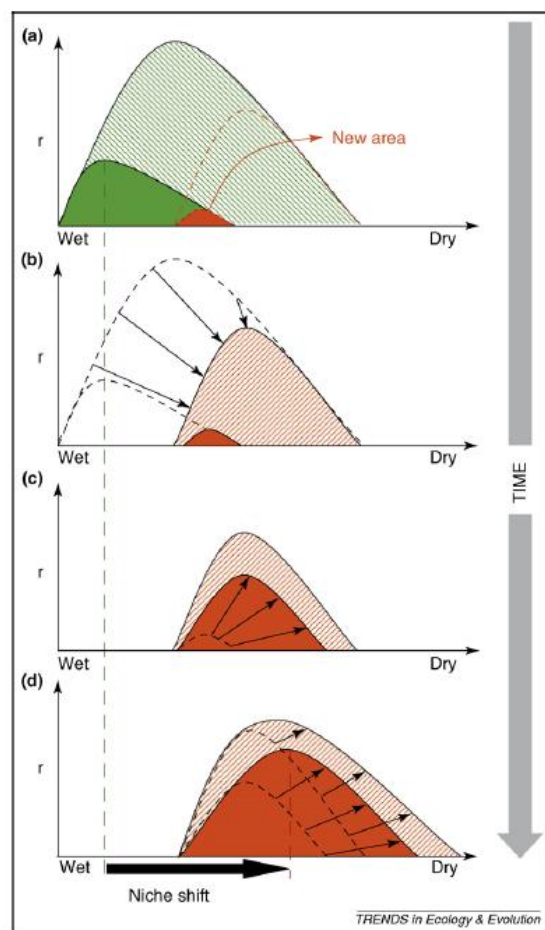


Figure 1.1 :

Relation entre la niche fondamentale et la niche réalisée pendant un événement de fondation et expansion de la niche suivant un gradient environnemental d'humidité (humide-sec). En ordonnée est indiqué le taux d'augmentation des populations. (a) La niche réalisée (vert foncé) de l'espèce en comparaison avec sa niche fondamentale (vert clair); la ligne en pointillé correspond à l'humidité optimale. Un ensemble d'individus défini par une niche réalisée (rouge foncée) et une niche fondamentale (pointillé rouge) va se disperser dans une nouvelle zone aux mêmes conditions; (b) Dans la nouvelle zone, les niches réalisées et les niches fondamentales sont plus petites que celles du milieu d'origine. (c) Ensuite l'espèce étend son utilisation des habitats (niche réalisée) dans des portions de la niche fondamentale qui n'étaient pas occupées dans le milieu d'origine, ce qui peut être dû à une relâche de la pression des compétiteurs dans ce nouveau milieu. (d) Avec le temps, la sélection directionnelle favorise les individus tolérants des milieux plus secs. L'humidité optimale à la croissance de la population est alors déplacée. La modification des niches est ainsi quantifiée.

De plus en plus, des études comparatives qui identifient des groupes d'espèces, soit caractérisés par une conservation de niche, soit par un changement de niche au travers de l'étude d'attributs partagés, sont mises en place par la combinaison des modélisations phylogénétiques et bioclimatiques (Yesson, Culham, 2006).

En effet comme le souligne Pearman *et al.* (2008b), il apparaît que l'application des modèles de prédiction de distribution des habitats est une des clés de compréhension de la dynamique des niches.

II- Notions de biogéographie

Les scientifiques se sont toujours passionnés par la distribution des espèces animales ou végétales à la surface de la terre. Or les études ponctuelles ne sont pas toutes généralisables et des comparaisons multi-sites sont difficiles. C'est pourquoi les outils d'analyses spatiales se sont développés pour permettre des études à des échelles variables selon la biologie de l'espèce étudiée. Ceci prend alors la forme de modèles de distributions spatiales qui combinent tous les facteurs environnementaux et écologiques pour comprendre la dynamique des niches, la distribution des espèces et les processus d'adaptation locale (Ferrier, Guisan, 2006; Guisan, Thuiller, 2005; Guisan, Zimmermann, 2000a; Kearney, Porter, 2009; Pearman *et al.*, 2008c)

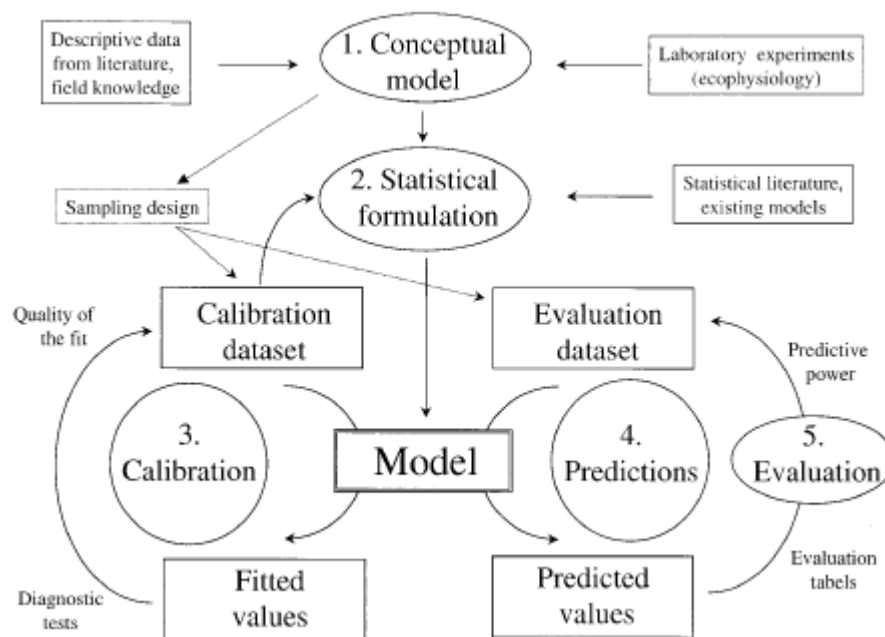
2.1 Modèle de prédiction de distribution des habitats

En combinant des modèles de prédiction de distribution et des études de phylogénie, il serait possible de dépasser la divergence des théories quant à la dynamique des niches en définissant les espèces par une condition *sine qua non* : des niches stables. Ceci constitue l'hypothèse principale de tous les modèles de prédiction de distribution des habitats (Guisan, Thuiller, 2005; Guisan, Zimmermann, 2000). Par définition, les modèles de prédiction de distribution sont basés sur des observations prenant en compte la limitation des ressources et les interactions biotiques. Ils sont donc supposés quantifier la niche réalisée d'une espèce (Guisan, Thuiller, 2005).

Ce sont des modèles statiques et probabilistes car ils établissent une relation statistique entre la distribution géographique des espèces ou des communautés et leur environnement à un instant donné. Il est considéré que les espèces modélisées sont en équilibre avec leur environnement (Guisan, Theurillat, 2000). Quand peu d'études existent sur la réponse dynamique des espèces aux changements environnementaux, les modèles statiques de distribution demeurent la seule approche intéressante pour comprendre les conséquences possibles d'un changement environnemental sur la distribution d'une espèce à large échelle (Woodward, Cramer, 1996). Pour pouvoir projeter un modèle dans le temps ou dans l'espace, il faut qu'il soit construit selon les hypothèses d'un équilibre espèce-environnement. Étonnamment, peu d'études se sont penchées sur les limites de cette hypothèse à la base de tous les modèles de prédiction. Araujo et Pearson (2005) ont examiné la validité de l'hypothèse spécifiant que la distribution des espèces est en équilibre avec le climat actuel. Les réponses aux changements climatiques varient en fonction des organismes étudiés. Les

réponses des plantes et des oiseaux seraient efficacement prédites par des modèles corrélant les distributions actuelles au climat, alors que les reptiles et les amphibiens seraient moins capables de modifier leur distribution et seraient donc plus vulnérables au cours d'un changement environnemental ou climatique rapide. Les distributions des plantes et des oiseaux seraient donc plutôt prévisibles avec des modèles incluant l'aspect dynamique des niches, alors que les distributions d'amphibiens ou de reptiles seraient mieux prédites avec des modèles de niches conservées. Globalement, la conception des modèles de prédiction de distribution comprend toujours cinq étapes incontournables : (1) la formulation conceptuelle du modèle, (2) l'échantillonnage, les échelles spatiales et le contexte géographique, (3) la formulation statistique du modèle, (4) la calibration du modèle et (5) les prédictions du modèle (Guisan, Zimmermann, 2000) (Figure 1.2).

Figure 1.2 : Etapes successives dans la construction d'un modèle (étapes de 1 à 5) quand deux jeux de données sont disponibles : un jeu de données pour calibrer le modèle et un autre pour le valider.



Le choix du modèle statistique en question dépend de la formulation conceptuelle du problème. La revue présentée par Guisan et Zimmermann (2000) réalise un bon tour d'horizon des méthodes statistiques pouvant être appliquées. Bien que ce ne soit pas une étude très récente, elle présente les méthodes statistiques encore utilisées actuellement : régressions généralisées, techniques de classification (CART), enveloppes environnementales, techniques d'ordination (analyses des correspondances) et approches bayésiennes ou réseaux de neurones. Quel que soit le modèle choisi,

il est important de sélectionner les variables les plus influentes comme variables prédictives du modèle (Guisan, Thuiller, 2005). Les réseaux de neurones artificiels apparaissent comme les plus performants pour prédire des distributions d'espèces ayant subi des modifications de niches (Araujo *et al.*, 2005).

Une des principales questions qui se pose quand un modèle de prédiction de distribution est mis en place concerne l'échelle spatiale de l'étude (Guisan, Thuiller, 2005). Les systèmes d'information géographiques permettent de gérer ces changements d'échelles et de combiner un grand nombre de données spatiales, telles que les données issues de télédétection dont le potentiel est très grand quant à la prédiction de variables biophysiques (Foody, 2007).

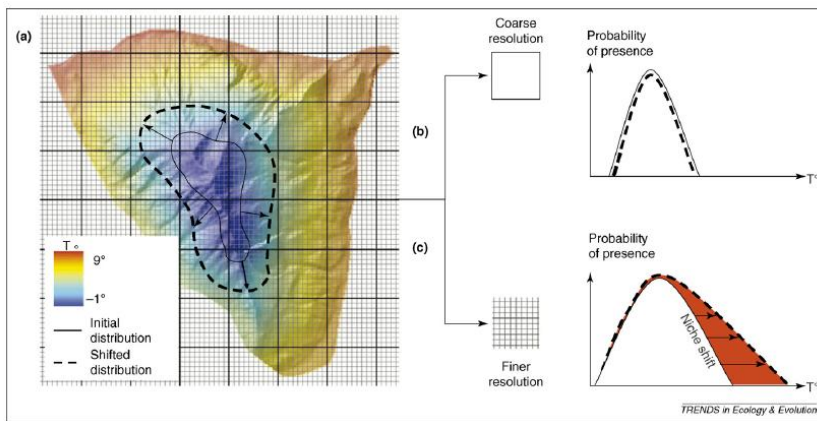


Figure 1.3 : (Pearman *et al.*, 2008c) - Exemple de changements de résolution spatiale. Le gradient de température ne révèle pas le changement de distribution de l'espèce en (a) alors que la modification de la niche apparaît avec une échelle plus fine en (b).

2.2 La multiplicité des échelles d'observation

Le climat est l'une des composantes environnementales principales quand il s'agit de la distribution des espèces. Les premières études sur la compréhension de la distribution des espèces étaient basées sur les gradients climatiques latitudinaux ou altitudinaux (Dobzhansky, 1951). Des programmes de cartographie de distribution d'espèces ont été mis en place considérant les variations climatiques comme l'élément discriminant de l'adaptation des espèces à leur milieu (Booth, Ekeleme, 2002; Booth *et al.*, 2000). Certains modèles de prédiction efficaces se basent sur les enveloppes bioclimatiques des espèces (Hijmans, Graham, 2006; Thuiller *et al.*, 2004). Ces enveloppes bioclimatiques sont définies selon les distributions et les exigences environnementales connues et sont utilisées pour prédire la distribution des espèces dans le futur ou dans le passé. Pourtant, l'utilité des enveloppes bioclimatiques des espèces quant à la prédiction de leur distribution a déjà été contestée (Pearson, Dawson, 2003; Pearson *et al.*, 2004). Il semblerait que le climat ne puisse pas

expliquer toutes les variations dans les distributions et que des données issues d'imageries satellites à plus ou moins haute résolution spatiale (métrique, décamétrique ou kilométrique) puissent apporter une information similaire. En effet, il semblerait que les données de télédétection soient sous-utilisées par rapport à leur potentiel quant à la prédiction des distributions (Zimmermann *et al.*, 2007). Des études très précises de prédictions de distribution à partir de variables prédictives issues de télédétection se sont avérées très efficaces, comme par exemple la modélisation de la distribution des espèces d'arbres en Amazonie (Saatchi *et al.*, 2008).

Les données de télédétection sont donc susceptibles d'apporter non seulement une information relative aux habitats et caractérisant les niches environnementales mais aussi la possibilité de cadrer efficacement les échelles spatiales et temporelles (Foody, 2007; Kerr, Ostrovsky, 2003; Turner *et al.*, 2003). La télédétection a de nombreuses applications écologiques en générant des variables environnementales telles que le détail des habitats (en termes de classification de l'occupation du sol) ou des propriétés biophysiques (comme l'évapotranspiration foliaire par exemple) (Tableau 1.3).

Selon le processus biologique visé et l'espèce étudiée, l'échelle spatiale la plus pertinente ne sera pas la même. La dimension des objets ou des phénomènes observés dirige le choix de l'échelle spatiale. Des études à l'échelle d'un pays sont nombreuses et nécessitent des échelles larges (Elliott *et al.*, 2004a; Kozak *et al.*, 2008a; Mueller *et al.*, 2008). En revanche, des processus biologiques bien particuliers nécessitent des résolutions spatiales bien plus fines et des superficies souvent moins étendues. Par exemple, la compétition inter-spécifique ne peut être détectée qu'à une résolution où les organismes peuvent interagir et entrer en compétition pour les mêmes ressources (Guisan, Thuiller, 2005). Aussi des études ayant pour objectif de prédire la biodiversité d'une région nécessitent la détection potentielle du plus grand nombre d'espèces et donc une très haute résolution spatiale et spectrale (Muller, Brandl, 2009).

De plus en plus, les applications des systèmes d'information géographique et de la télédétection s'élargissent (Turner *et al.*, 2003) (Tableau 1.3). Les applications concernant la dynamique des populations et contribuant à la connaissance en biologie de l'évolution sont celles qui nous intéressent. Celles-ci ont des applications très importantes en termes de gestion et conservation des ressources vivantes.

Tableau 1.3 : Synthèse des études utilisant des outils de télédétection pour caractériser des variables environnementales (issue de Turner et al, 2003). Sensor= Capteur satellitaire; V = Visible ; NIR = Near Infrared ; SWIR = Short Wavelength Infrared; TIR = Thermal Infrared; STRM = Shuttle Radar Topography mission

Ecological variable	Sensor	Spatial resolution	Spectral resolution	Description
Species composition	Ikonos Quickbird Casi Hyperion TM/ETM	< 1-30m	V NIR SWIR	These sensors are being testes for their ability to measure directly canopy community and perhaps species types based upon unique spectral signature
Land Cover	Modis TM/ETM Ikonos Quickbird	<1-1000m	V NIR SWIR TIR	Can discriminate different land surfaces at various resolutions ; land cover classification is considered a first-order analysis for species occurrence
Chlorophylle	Ikonos Quickbird Casi Hyperion TM/ETM Modis Avis	<1-1000m	V NIR SWIR TIR	Measures reflectance to asses presence/absence of vegetation and relative greenness measures enabling detection of ocean and land surfaces chlorophyll useful for calculating productivity and plant health
Phenology	Modis TM/ETM Hyperion Ikonos Quickbird	<1-1000m	V NIR SWIR TIR	Information on leaf turnover and flowering/fruitle cycles can be inferred from comparisons of time series of images. Provides for identification of species tied to certain phonological events
Habitat Structure / Topography	STRM ATM Aster Ikonos Slicer LVIS	90m STRL 30/15m ASTER 1-15m Ikonos, Slicer, LVIS	Microwave for SRTM ; V /NIR / SWIR for others	Digital elevation models derived from radar signals via interferometry (STRM), image stereo pairs (ASTER) or discrete-return LIDAR signals. Many species are constrained by microhabitats resulting from changes in altitude ; elevation also determines waterched flows

2.3 L'application à la structuration des espèces

Pour suivre l'évolution d'une espèce dans un milieu, la notion principale est donc sa capacité d'adaptation. Afin d'étudier la biodiversité, il est nécessaire de prendre en compte toute la variation, c'est-à-dire qu'il faut considérer la distribution spatiale de la diversité génétique. Intuitivement, cette notion prend en considération non seulement les caractéristiques de la plante, mais aussi les caractéristiques des habitats dans lesquels elle évolue. Deux types d'approches semblent s'être distinguées : les approches centrées sur les analyses spatiales (outils d'analyses spatiales pour comprendre la distribution des espèces ainsi que leurs possibles interactions) (Kozak et al., 2008b;

Swenson, 2008) et celles centrées sur les analyses génétiques (se baser sur des analyses génétiques de la structuration et comprendre en quoi le paysage ou l'environnement façonne la structuration) (Antonovics *et al.*, 1994; Manel *et al.*, 2003). Les deux approches sont très différentes, mais peuvent mener à des conclusions convergentes.

2.3.1 L'approche spatiale

Les bases de l'évolution des espèces se situent dans le processus de spéciation qui peut être mieux appréhendé en étudiant notamment des zones hybrides dont l'observation peut commencer spatialement (Kozak *et al.*, 2008b; Swenson, 2008). Pour comprendre les phénomènes de spéciation et plus communément les processus d'hybridation inter-spécifique, il faut comprendre comment se distribuent les espèces concernées et quelles sont les conditions environnementales permettant ces flux de gènes. Au-delà de la distribution des espèces par une modélisation de la distribution des niches écologiques associées, suivre l'évolution d'une ou plusieurs espèces nécessite de prendre en compte la dynamique de celles-ci (Kozak *et al.*, 2008b). L'étude de zones hybrides chez *Ephippiger ephippiger* (un orthoptère) a été menée via des analyses spatiales afin de comprendre si le complexe d'espèces résulte d'une divergence primaire ou d'un contact secondaire entre des populations se dispersant à partir des refuges auxquels elles étaient initialement cantonnées (Kidd, Ritchie, 2000). L'étude des dynamiques de spéciation peut aussi être étudiée via des Systèmes d'Information Géographique (SIG) comme pour les grenouilles et l'étude de la superposition écologique entre des espèces proches (Graham *et al.*, 2004). Ces études peuvent donner des clés quant à l'histoire de la spéciation (Swenson, 2008) (Figure 1.4).

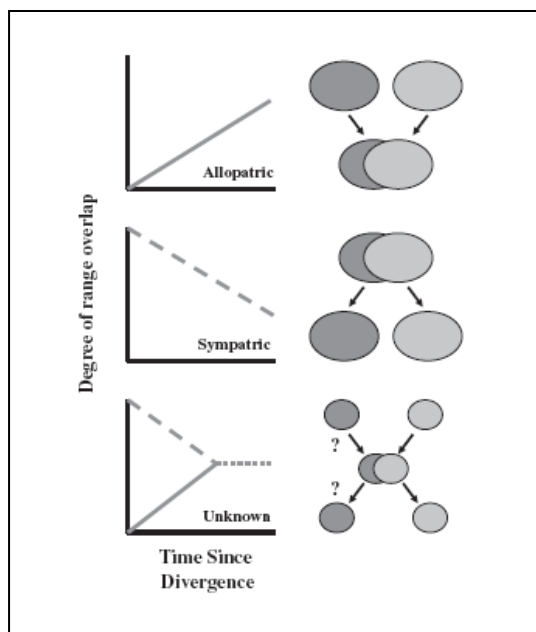


Figure 1.4 : Un modèle graphique pour comprendre en quoi la distribution géographique d'espèces sœurs peut permettre une prédiction du mode géographique de spéciation (Berlocher, 1998) ; Le dernier graphique, montre que dans certains cas les résultats peuvent être ambigus (Losos, Glor, 2003).

L'un des objectifs communément abordé et utilisant de telles approches est la conservation des espèces. La gestion des ressources génétiques est une question phare de notre siècle quand on considère non seulement les destructions d'habitats causées par les activités humaines, mais aussi les changements climatiques qui s'accroissent. Sans information sur la distribution spatiale de la diversité génétique en relation avec l'occupation du sol, il est très délicat de comprendre la variation entre populations et de détecter si la variation génétique est menacée (Ji, Leberg, 2002). Sur la base d'études bien précises de conservation et de gestion des ressources génétiques, l'apport des modèles spatialisés est incontestable, par exemple pour l'exploration de la biodiversité et la conservation du haricot commun *Phaseolus vulgaris* en Amérique du sud (Jones *et al.*, 1997), ou la biogéographie de la cacahuète *Arachis hypogea* dans les Andes (Ferguson *et al.*, 2005), ou encore l'étendue de la distribution du poivre *Capsicum flexiosum* au Paraguay (Jarvis *et al.*, 2005).

2.3.2 L'approche génétique

Parallèlement, de nouvelles méthodes d'analyse de la diversité génétique et de la structure génétique des populations se sont développées et permettent de nos jours d'avoir des visions très fines de la structure des populations. C'est en 1994 qu'Antonovics propose une nouvelle discipline appelée « la génétique communautaire » qui étudie les interactions entre les processus écologiques et évolutifs (Antonovics *et al.*, 1994). Celles-ci façonnent les structures des communautés et leurs dynamiques. Ce sont les bases de la génétique du paysage, discipline à part entière, qui combine deux approches à savoir l'écologie du paysage et la génétique des populations (Manel *et al.*, 2003; Turner, 2001).

Par exemple, une étude combinant la modélisation des niches écologiques et la phylogéographie moléculaire de l'Orge marine (*Hordeum marinum*) a révélé l'histoire évolutive de cette espèce au travers de la différenciation des niches, de la perte de diversité génétique, marques d'un ancien phénomène de spéciation (Jakob *et al.*, 2007). Ces études fines en génétique des populations utilisent de nombreux outils très performants et permettent d'inférer des histoires évolutives qui ont façonné les espèces telles qu'on les observe aujourd'hui.

Les deux types d'approches sont bien sûr complémentaires et des études combinées à différentes échelles (Pavlacky *et al.*, 2009; Pease *et al.*, 2009), avec des analyses génétiques à l'échelle d'une population et des analyses de modélisation des niches environnementales et de prédiction de distribution des espèces à échelle régionale, sont une des clés de compréhension des milieux naturels. En effet, les tenants et les aboutissants de la biogéographie au sens large du terme

permettent d'avoir une véritable vue transversale des problématiques liées à la dynamique de la distribution des espèces et de leurs habitats.

III- Cas des espèces introduites et de leur devenir dans les écosystèmes

Il est souvent considéré que les processus écologiques et évolutifs sont basés sur des échelles de temps très différentes, avec les échelles évolutives bien plus larges que les échelles écologiques. Pourtant, il existe des cas où les échelles se rejoignent et un processus écologique peut avoir des conséquences évolutives directes, entraînant des changements rapides (Thompson, 1998). Naturellement, la structure des métapopulations (comme définies par Levins) peut rapidement se modifier et agir ainsi sur la structure génétique des espèces à différentes échelles géographiques (Hanski, 1998). De nos jours, les impacts anthropiques peuvent modifier les milieux encore plus profondément et impliquer des changements évolutifs très rapides. En effet, en plus de l'hétérogénéité spatiale et temporelle des niches, les pressions anthropiques imposent un stress important notamment aux populations forestières, qui sont alors mises à l'épreuve au niveau de leurs capacités adaptatives (Lexer *et al.*, 2004 ; Mueller-Starck, Schubert, 2001).

3.1 Impact anthropique généralisé

L'une des principales conséquences des activités anthropiques sur la planète est le déplacement d'espèces végétales ou animales lors du transport des hommes ou des marchandises ou intentionnellement pour la culture de certaines espèces dans des régions données du globe.

L'étude des flores régionales montre que beaucoup de plantes sont introduites (Abbott *et al.*, 2003). Par exemple, la flore de grande Bretagne recense approximativement la moitié des espèces comme étant autochtones et l'autre moitié comme étant introduites (Preston *et al.*, 2002). L'introduction d'espèces implique une modification de leurs niches écologiques et ceci peut avoir des conséquences très fortes sur les écosystèmes au travers de dynamiques d'hybridation, d'extinction ou d'invasion (Mack *et al.*, 2000).

Les impacts écologiques des espèces introduites sont variés. En effet, le destin des organismes introduits peut être l'extinction immédiate ou, s'ils s'acclimatent, l'invasion. Les conséquences évolutives de ces introductions peuvent avoir un effet sur le milieu naturel sur des échelles de temps très courtes au travers du processus d'hybridation notamment ou du processus de sélection suite à la

modification de la niche de l'espèce concernée (Abbott *et al.*, 2003b; Allendorf, Lundquist, 2003; Thompson, 1998). Bien sûr, les processus évolutifs naturels sont supposés se dérouler sur des échelles de temps plus grandes, mais les impacts anthropiques sur les écosystèmes accélèrent tous les processus. Ceci est d'ailleurs un point sensible à l'heure où les changements globaux menacent notre planète (Parmesan, Yohe, 2003; Pounds *et al.*, 1999; Ricciardi, 2007; Thomas *et al.*, 2004).

L'élément principal qui détermine le devenir d'une espèce introduite est l'effet de fondation. Ce dernier implique un goulot d'étranglement et donc une diminution de la diversité génétique (Nei *et al.*, 1975). Selon la nature des génotypes introduits, le potentiel de l'espèce dans son nouvel environnement sera variable. La réduction de la diversité allélique et de l'hétérozygotie peut mener à une différenciation rapide des populations introduites et à la fixation d'allèles délétères (Ellstrand, Elam, 1993). Des cas d'introductions multiples ont souvent été recensés et ceci peut limiter la réduction de la diversité si les populations introduites proviennent de différents endroits de la répartition originelle de l'espèce (Lambrinos, 2004). Pour des espèces comme *Bromus tectorum*, introduite en Amérique du Nord, il est clair que ce sont des phénomènes d'introductions multiples qui ont contribué au haut niveau de diversité (Novak, Mack, 1993). Pourtant, il a été démontré que ce n'est pas ce niveau de diversité élevé qui contribue généralement au succès envahissant des espèces introduites. En effet, il apparaît que les effets de fondation avec peu de génotypes très adaptés sont le plus souvent à la base du caractère envahissant de l'espèce introduite (Dlugosch, Parker, 2008).

Un cas particulier qui mérite d'être mentionné ici, concerne les îles. Les conséquences écologiques et évolutives occasionnées par les introductions d'espèces sont par ailleurs très problématiques dans les îles. Les îles ont souvent favorisé l'isolement reproducteur et permis des phénomènes de spéciation (MacArthur, Wilson, 1963; Mayr, 1963), les océans formant des barrières aux flux de gènes. De plus, les conditions abiotiques et biotiques des îles peuvent être très différentes et l'absence de certaines espèces de prédateurs ou d'espèces très compétitives rendent les îles d'autant plus vulnérables aux impacts dévastateurs des introductions (Cox, 1999). L'arrivée des hommes sur certaines îles a contribué à l'extinction de plusieurs espèces d'oiseaux (Blackburn *et al.*, 2004). De nombreuses études relatent les caractéristiques particulières des milieux insulaires en soulignant le fait qu'ils ne sont pas confrontés aux mêmes pressions que les milieux continentaux (Buckley, Jetz, 2007; Kier *et al.*, 2009; Wittaker, Fernandez-palacios, 2007). Les îles connues comme des centres de diversification pour des espèces dont la distribution est limitée, sont souvent associées à un taux d'endémisme élevé (Wittaker, Fernandez-Palacios, 2007).

3.2 Les conséquences : de l'extinction à l'invasion

L'invasion des milieux par des espèces dites exotiques est très certainement le processus qui cause le plus grand nombre de dégâts dans les milieux naturels (Mack 2000, Lambrinos 2004, Suarez, 2007, Ricciardi 2007). Les processus d'invasions sont favorisés par un ensemble de paramètres environnementaux favorables comme des fluctuations climatiques, des changements dans les paysages ou des perturbations physiques (Mack *et al.*, 2000). De plus en plus, il est admis que les populations introduites expérimentent souvent des changements évolutifs rapides à différents niveaux, morphologiques, comportementaux ou à certains traits d'histoire de vie (Lambrinos, 2004). Afin de comprendre et d'appréhender les problèmes de plantes envahissantes, les études se sont beaucoup focalisées sur les effets de fondations.

Toutes les espèces introduites ne deviennent pas envahissantes. Certaines seront éliminées très rapidement par des agents physiques ou biotiques, comme ce fut le cas de la truite *Salmo trutta* introduite dans les îles subantarctiques (Cooper et al 1992). Il est quasiment impossible de quantifier le nombre d'espèces qui ont été déplacées de leur milieu d'origine et qui ont péri. Mais selon les observations d'espèces trouvées en dehors de leur milieu d'origine, il semblerait que le nombre d'espèces introduites qui ont subi un phénomène d'extinction soit assez élevé (Mack, 2000). Les traces des phénomènes d'extinction sont peu nombreuses. Pourtant, des cas où les génotypes introduits ne sont pas adaptés ou rencontrent une compétition biotique trop forte doivent être communs.

Le déplacement d'espèce peut permettre à des espèces génétiquement proches de se rapprocher. Les conditions abiotiques et climatiques du nouvel environnement peuvent permettre une levée de barrières à la reproduction entre ces espèces (Lamont *et al.*, 2003). Le cas des hybridations est central dans la mesure où ces phénomènes mènent à des changements évolutifs rapides (extinction ou même spéciation). En effet, il s'avère que les introductions de plantes peuvent mener à des hybridations naturellement impossibles (Abbott *et al.*, 2003). La question de l'origine des hybridations est souvent posée (Mallet, 2005) car de plus en plus de cas d'hybridation sont relatés. Une des principales hypothèses incrimine les activités anthropiques comme causes de nombreux cas d'hybridations (Allendorf *et al.*, 2001).

Ces hybridations peuvent avoir des conséquences très variées selon le niveau de rapprochement des deux espèces, provoquant parfois l'extinction d'une espèce (Rhymer, Simberloff, 1996) ou menant à un mélange d'hybrides de différentes générations (Ayres *et al.*, 1999; Ayres *et al.*, 2004; Pacifici *et al.*, 2009; Pinto *et al.*, 2005) ou encore augmentant le pouvoir invasif d'une espèce (Ellstrand, Schierenbeck, 2000). Ce phénomène est très fortement dépendant des conditions environnementales du milieu ainsi que des relations phylogénétiques entre les espèces concernées. Dans le cas où les flux de gènes aboutissent à des hybrides viables, le mélange des génomes des espèces peut alors avoir lieu.

3.3 Clés de compréhension de l'adaptation

Les caractéristiques écologiques du milieu, les exigences environnementales des espèces, les interactions biotiques rencontrées et le niveau de diversité génétique des espèces déplacées sont des facteurs à prendre en considération pour étudier l'évolution de l'espèce introduite.

Pour comprendre et analyser les impacts d'une espèce introduite dans son nouvel environnement, l'approche la plus communément utilisée est de comparer ses caractéristiques dans son milieu d'origine et son nouveau milieu (Bossdorf *et al.*, 2005; Darling *et al.*, 2008; Durka *et al.*, 2005a; Genton *et al.*, 2005; Novak, Mack, 1993).

Les espèces introduites et plus particulièrement les espèces devenues envahissantes ne sont pas en équilibre dans le milieu envahi. De ce fait, pour modéliser leur distribution, il est préférable de se baser sur les conditions environnementales de leur milieu d'origine (Peterson, Robins, 2003; Robertson *et al.*, 2004). En comparant des populations indigènes et des populations introduites de la même espèce, il est donc possible de répondre à l'une des principales questions au-delà des questionnements écologiques : « quelle part de la diversité génétique de l'espèce a été introduite ? ». Si l'introduction de l'espèce est accompagnée d'un goulot d'étranglement, les fréquences géniques sont redistribuées et le destin évolutif de l'espèce dans son nouveau milieu en dépend (Zhang *et al.*, 2004). A l'inverse, dans des cas d'introductions multiples, la diversité génétique de l'espèce sera grande et la variation génétique mieux conservée, ce qui peut contribuer au succès de l'espèce dans son nouvel environnement parfois incontrôlable (Darling *et al.*, 2008; Dlugosch, Parker, 2008; Durka *et al.*, 2005a; Genton *et al.*, 2005).

Etudier le devenir d'une population introduite dans son nouvel environnement constitue une possibilité intéressante pour comprendre son potentiel d'adaptation (Hierro *et al.*, 2005). En effet, la plupart des cas d'adaptation contemporains sont associés à des événements de colonisation (Reznick, Ghalambor, 2001). Or, les événements d'introduction provoqués par l'homme sont susceptibles de mimer des processus de colonisation. Lorsqu'une population est déplacée, l'espèce dans son ensemble est alors confrontée à des conditions environnementales hétérogènes ce qui crée des possibilités d'adaptation locale (Reznick, Ghalambor, 2001). Comme le souligne une revue sur les questions conceptuelles que soulèvent les processus d'adaptation locale (Kawecki, Ebert, 2004), le meilleur moyen d'étudier ces processus est une manipulation de transplantations réciproques dans des milieux différents. L'importance d'étudier les performances des plantes d'origines diverses dans des environnements homogènes par le biais de la variation de caractères écologiques est communément admise depuis Wright (1976).

IV- Questions fondamentales abordées

Cette revue bibliographique nous mène à plusieurs questions fondamentales. Elles tournent autour de l'adaptation suite à des déplacements et des introductions d'espèces. Ce sont plus particulièrement les questions relatives à l'hybridation inter-spécifique qui nous intéressent et comment le déplacement d'espèces dans un nouvel environnement peut permettre à certaines d'entre elles naturellement isolées de s'hybrider. L'hybridation apparaît comme l'un des principaux moteurs de l'évolution et cela est d'autant plus vrai à l'heure où les activités humaines modifient grandement les distributions naturelles des espèces et permettent des contacts nouveaux entre espèces génétiquement proches.

D'un point de vue génétique, l'ampleur des flux de gènes et du brassage génétique peut être plus ou moins limitée. Quel est le niveau de diversité introduit par rapport à la diversité connue de l'espèce dans son milieu naturel ? Est-ce que les flux de gènes vont aboutir à des hybrides viables et fertiles ? Quelle est l'ampleur de ces hybridations en termes de mélange des génomes ?

D'un point de vue écologique, lors d'une introduction, une modification des niches écologiques se produit. Quelles sont les conséquences de cette modification de niches sur les traits d'histoire de vie des espèces ou sur leur potentiel d'adaptation au nouvel environnement ? Comment l'environnement agit-il sur la levée de barrières à la reproduction entre espèces originellement isolées ?

D'un point de vue spatial, les hybridations peuvent être limitées à certaines niches particulières, ou plus ou moins généralisées à une région entière. Quelle est la distribution des espèces dans leur nouvel environnement ? Quelle est la distribution des niches favorables à la levée de barrières aux flux de gènes ? Quelle est l'ampleur de ces hybridations en termes de distribution spatiale ?

Enfin globalement, *d'un point de vue appliqué*, ceci peut-il avoir des conséquences sur les ressources génétiques des espèces concernées et sur la mise en place de méthodes de gestion adaptées au contexte socio-économique et environnemental du café en Nouvelle-Calédonie ?

Chapitre 2 - Présentation du modèle d'étude



Le café joue un rôle primordial pour le commerce mondial et pour l'économie de nombreux pays en développement comme le Brésil, le Mexique, la Colombie, ou l'Éthiopie. Le café est au quatrième rang des échanges commerciaux des produits agricoles et au second rang des exportations après le pétrole (Pendergrast, 2009), avec plus de 116 millions de sacs produits pendant l'année caféière 2007/08, en baisse par rapport à la production de 2006/07 qui avait atteint 126,8 millions de sacs.

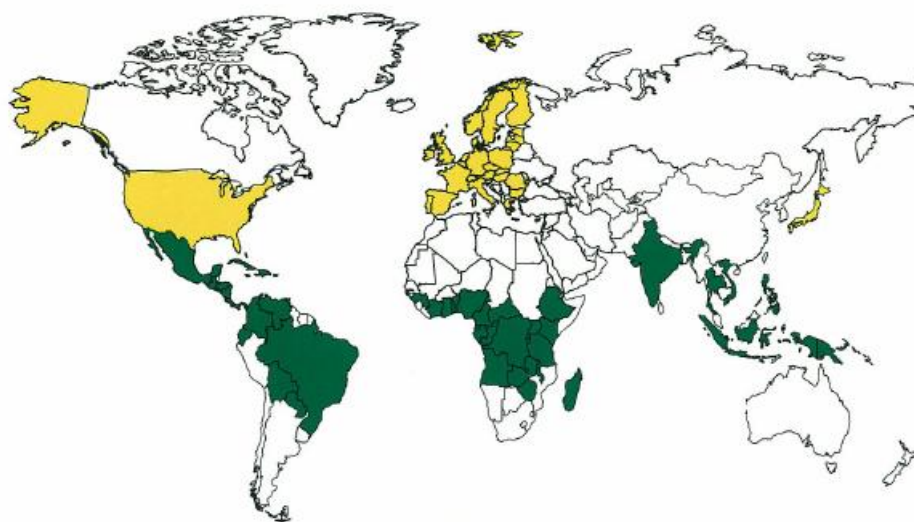


Figure 2.1 : Carte des pays exportateurs et importateurs de café selon le rapport de l'Organisation Internationale du café (OIC) (2007-2008)

Exporting Countries

Angola - Benin - Bolivia - Brazil - Burundi - Cameroon - Central African Republic - Colombia - Congo, Democratic Republic of - Congo, Republic of - Costa Rica - Côte d'Ivoire - Cuba - Dominican Republic - Ecuador - El Salvador - Ethiopia - Gabon - Ghana - Guatemala - Guinea - Haiti - Honduras - India - Indonesia - Jamaica - Kenya - Madagascar - Malawi - Mexico - Nicaragua - Nigeria - Panama - Papua New Guinea - Paraguay - Philippines - Rwanda - Tanzania - Thailand - Togo - Uganda - Venezuela, Bolivarian Republic of - Vietnam - Zambia - Zimbabwe

Importing Countries

European Community (Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, United Kingdom) - Japan - Norway - Switzerland - United States of America

En Amérique latine et en Afrique, beaucoup de pays sont tributaires des recettes d'exportation du café. En effet, la caféiculture a longtemps été plus rémunératrice que les autres activités agricoles (sur une surface donnée). C'est dans ce cadre que des projets de développement de la caféiculture se sont mis en place très tôt, comme par exemple au Togo, au Vietnam ou en Nouvelle-Calédonie, afin que ces pays puissent accroître leurs ressources à l'exportation (Deuss, 1987). Ces projets ont pour

objectif d'atteindre des niveaux de production suffisamment élevés tout en améliorant la qualité du café de façon à obtenir un meilleur prix de vente et améliorer le niveau de vie des paysans. Actuellement, l'Accord International sur le café de 2007 (établi par les 77 membres de l'Organisation Internationale du Café, OIC) œuvre à la mise en place d'une économie caféière durable pour le bénéfice de toutes les parties prenantes et particulièrement des petits caféiculteurs des pays producteurs de café (<http://www.ico.org>). Selon l'OIC, il est difficile d'envisager une activité économique alternative dans des régions productrices de café qui ait un tel impact positif. La culture du caféier, plante pérenne, crée des emplois ruraux stables. Le caféier est un arbuste à feuilles persistantes, un important contributeur à la séquestration du carbone et à la stabilisation des sols.

La culture du café dépend principalement de deux espèces : *Coffea arabica* L. qui représente 61,36% de la production mondiale de café pendant l'année caféière 2007/08 et *Coffea canephora* Pierre (Robusta) qui en représente 38,64%. Une troisième espèce, *Coffea liberica* Hiern, est cultivée si peu qu'elle n'est pas comptabilisée par l'OIC mais est connue pour ses atouts intéressants (tel que sa granulométrie élevée). Ces trois espèces, originaires d'Afrique, ont été introduites dans grand nombre de pays qui produisent ou qui ont produit du café. Dans certains cas, des cultures de café ont été abandonnées mais les espèces cultivées ont la capacité de reprendre des morphologies sauvages et de survivre naturellement. L'étude de l'introduction de ces trois espèces dans certains pays peut permettre d'étudier l'influence du changement de niches environnementales sur leur adaptation au milieu lors d'un arrêt des cultures. Parmi les espèces du genre *Coffea*, nous nous intéressons donc uniquement aux trois espèces cultivées, car ce sont les seules qui ont été déplacées dans le cadre de leur culture et qui peuvent donc être étudiées en dehors de leurs aires naturelles de distribution géographique. La situation au sein du genre est indispensable pour connaître les relations entre les trois espèces étudiées. Nous introduirons ensuite la situation connue en Nouvelle-Calédonie où ces trois espèces ont été introduites et où les cultures ont été ensuite abandonnées.

I – Le café : origine Afrique

1.1 Connaissances générales

1.1.1 Le complexe d'espèces des caféiers

Les caféiers appartiennent à la grande famille des Rubiaceae qui comprend environ 650 genres et 13000 espèces (Rova *et al.*, 2002) et à la sous-tribu des Cofféinées (Chevalier, 1946). Ce sont des arbres ou arbustes à feuilles simples, opposées, à stipules souvent bien développées. Leurs fleurs blanches sont hermaphrodites et groupées en petites inflorescences. Le fruit est une drupe à pulpe sucrée et contient normalement deux graines. La première classification systématique des caféiers fut proposée en 1929 par Chevalier qui la modifiera par la suite en 1946. Cette classification a été améliorée depuis les années 80 couvrant l'Afrique tropicale (Bridson, Verdcourt, 1988 ; Bridson, 1994), et les îles Mascaraïnes (Leroy, 1982). Il existe deux genres proches, le genre *Coffea* L. et le genre *Psilanthus* Hook. F., différenciés sur des critères de morphologie florale en deux sous genres *Coffea* et *Bara Coffea*, *Psilanthus* et *Afri Coffea* respectivement. Au sens strict du terme, seules les espèces du genre *Coffea* sont considérées comme des caféiers.

Le genre *Coffea* recense 103 espèces (Davis *et al.*, 2006). Tous les caféiers sont diploïdes avec $2n = 22$ chromosomes, excepté *C. arabica* qui est allotétraploïde avec $2n = 44$ chromosomes. Cette dernière serait issue de l'hybridation de deux autres espèces du genre ou d'espèces très apparentées à *C. canephora* et *C. eugenioides* (Lashermes *et al.*, 1999).

La reproduction des *Coffea* est soumise à un système d'auto-incompatibilité gamétophytique (SI) monolocus (Coulibaly *et al.*, 2002), ce qui induit un fort taux d'hétérozygotie dans les populations naturelles. *C. arabica* est une exception puisqu'elle est auto-compatible et préférentiellement autogame, tout comme *C. heterocalyx* et *C. anthonyi* (ex *C. sp. moloundou*).

Les travaux de Cros (1995), montrent qu'il existe une corrélation entre taille des génomes et origine géographique des espèces. Il existe en effet un gradient est-ouest : les génomes de l'est étant plus petits que ceux de l'ouest. Cette variation de taille de génomes pourrait être une source de mésappariements des chromosomes en méiose et donc de stérilité. Néanmoins, cette stérilité n'est jamais totale, ce qui sous-entend que la différenciation des génomes est probablement due à des réarrangements chromosomiques. Le gradient est-ouest a été confirmé par la suite (Noirot *et al.*, 2003). L'hypothèse principale est liée aux différences abiotiques des régions d'est en ouest, avec des conditions pluviométriques différentes opposant les espèces xérophytiques de l'est (*C. racemosa*,

C. sessiliflora ou *C. pseudozanguebariae*) aux espèces telles que *C. humilis*, *C. liberica* ou *C. canephora* qui sont adaptées aux forêts humides d'Afrique de l'ouest.

Les premières études sur les relations génétiques des espèces du genre *Coffea* ont porté sur les relations cytogénétiques basées sur le nombre d'hybrides sur 100 fleurs (Charrier, 1978; Louarn, 1992) (Figure 2.2). La tendance la plus marquante est en effet la séparation nette est-ouest, qui rejoint les observations concernant les variations dans les tailles des génomes (Cros *et al.*, 1995; Noirod *et al.*, 2003). Louarn (1992) a montré que sur 100 fleurs, 6 à 18 hybrides se sont formés entre *C. canephora* et *C. liberica*. Plus tard en 1993, Louarn a travaillé sur la viabilité pollinique d'hybrides interspécifiques. Entre 41% et 60% des hybrides F₁ issus d'un croisement entre *C. canephora* et *C. liberica* sont fertiles en termes de viabilité pollinique.

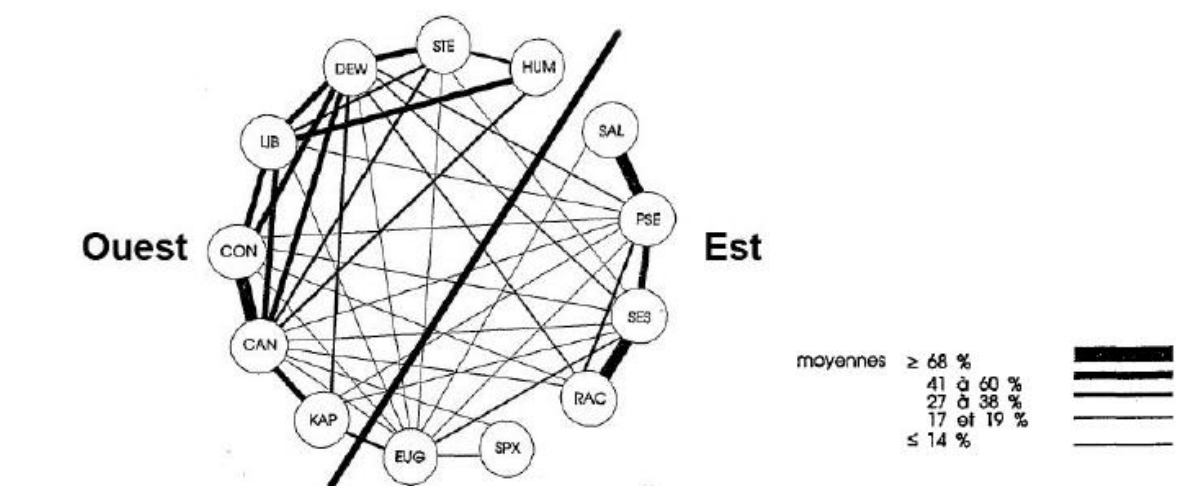


Figure 2.2 : Relations cytogénétiques entre espèces de caféiers selon Louarn (1992). Plus le trait qui relie deux espèces est épais, plus le pourcentage d'hybrides fertiles est élevé, donc plus les espèces sont considérées proches.

La divergence génétique marquée entre les espèces de caféiers d'Afrique de l'ouest et de l'est se traduit aussi par une divergence génomique. En effet, deux paires de chromosomes satellites sont observées chez les espèces d'Afrique de l'est contre une seule paire pour les autres espèces. Cette différence de structure est associée au nombre et à la distribution des locus d'ADN ribosomique 18S et 5S (Hamon *et al.*, 2009).

1.1.2 Relations phylogénétiques

Les relations phylogénétiques entre les espèces de caféiers ont été analysées en travaillant sur les divergences des séquences ITS (Internal Transcribed Spacer) de l'ADN nucléaire ribosomique (Lashermes *et al.*, 1997) ainsi que sur les variations de l'ADN chloroplastique (Cros *et al.*, 1998). La

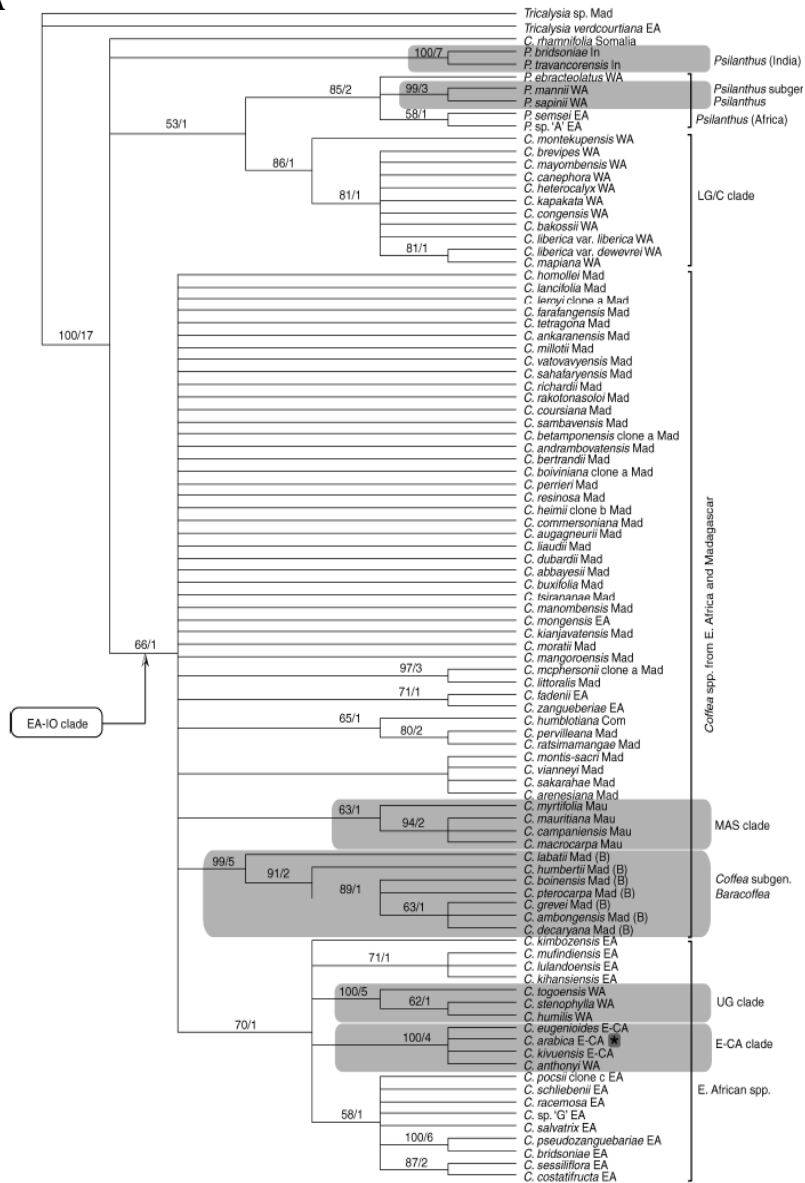
diversité des séquences nucléaires ou chloroplastiques analysées est très faible, ce qui suggère une divergence récente des espèces au sein du genre *Coffea*. Les résultats obtenus par ces deux études indépendantes sont plutôt cohérents quant au regroupement des espèces en quatre groupes majeurs biogéographiques ou clades (Afrique de l'ouest, centre africain, Afrique de l'est et Madagascar). *C. liberica* et *C. canephora* appartiennent au clade ouest et *C. arabica* se retrouve dans le clade ouest selon Cros (1998) alors qu'elle se retrouve dans le clade centre selon Lashermes (1997).

Récemment, une phylogénie plus complète a été construite sur la base des séquences de quatre régions de l'ADN chloroplastique (intron *trnI-F*, espace inter-génique *trnI-F*, intron *rpl16*, espace inter-génique *accD-psa1*) de la région ITS de l'ADN nucléaire (ITS 1/5-8S/ITS 2), et de la totalité des régions séquencées combinées (Maurin *et al.*, 2007). Cette étude confirme la proximité génétique des espèces *C. liberica* et *C. canephora* appartenant au même clade pour les trois analyses menées (clade Lower Guinean/ Congolian). Par ailleurs, l'origine hybride (cas de spéciation hybride allopolyploïde) de *C. arabica* est confirmée par sa position à proximité de *C. canephora* selon l'analyse des ITS et de *C. eugenioïdes* selon l'analyse de l'ADN chloroplastique dans les trois études (Cros *et al.*, 1998; Lashermes *et al.*, 1997; Maurin *et al.*, 2007). Les espèces considérées comme les progéniteurs de l'espèce *C. arabica* sont *C. canephora* et *C. eugenioïdes*. Ceci est cohérent avec la situation récente de sympatrie de ces deux espèces selon la distribution actuelle des espèces (Davis *et al.*, 2006).

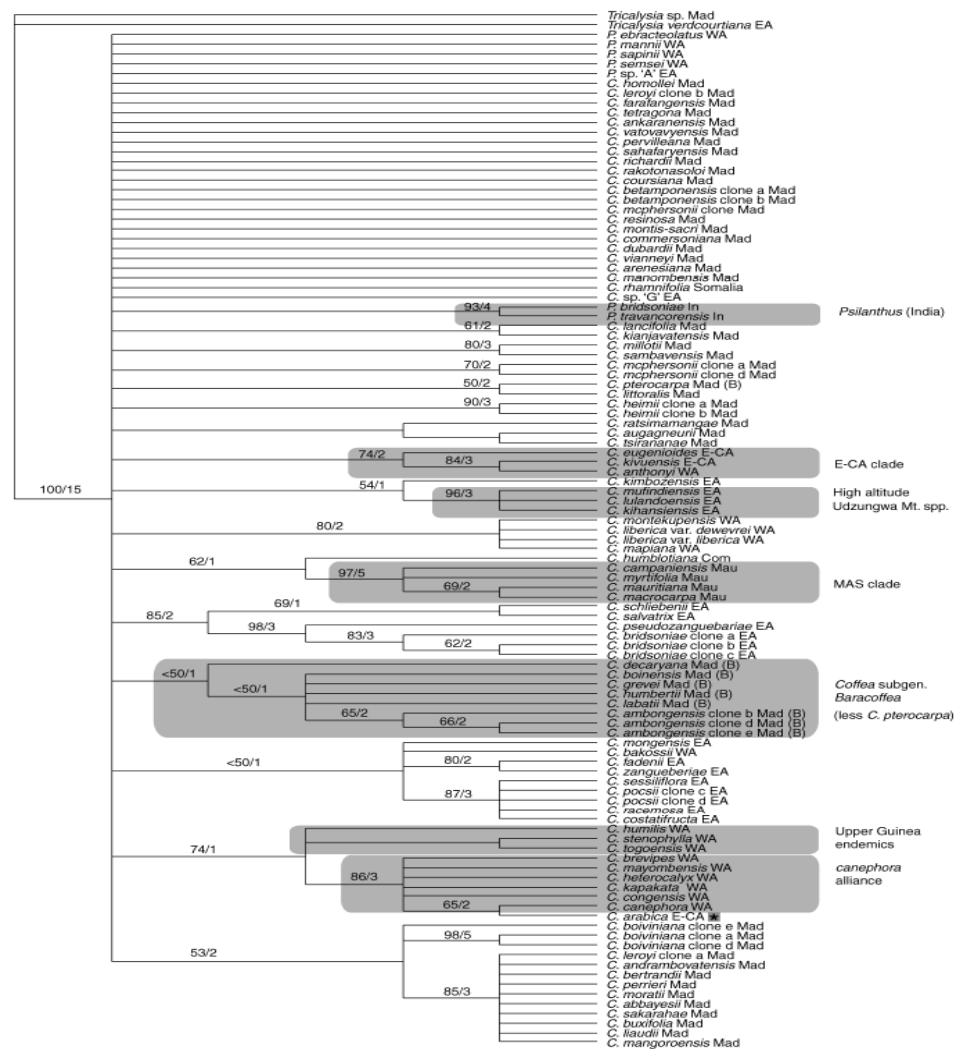
Sont présentés ci-dessous (Figure 2.3) deux arbres phylogénétiques consensus à partir des analyses combinées des plastides (A) et des analyses ITS (B).

*Figure 2.3 : Arbres phylogénétiques consensus établis par Maurin et al. (2007). L'arbre A est basé sur les analyses combinées des plastides et l'arbre B est basé sur l'analyse des séquences ITS. Les valeurs des bootstrap sont annotées au dessus de la branche quand la valeur est supérieure à 50%. E-CA clade = East-Central Africa clade ; LG/C clade = Guinea/Congolian clade ; UG clade = Upper Guinea clade ; MAS clade = Mascarene clade. Régions: EA = East Africa; E-CA= East-Central Africa; Com = Comoros; In = India; Mad = Madagascar; Mau = Mauritius; WA= West Africa. L'analyse des plastides permet des regroupements géographiques forts (bootstrap > 50%) : le clade UG, le clade LG/C et le clade E-CA (qui inclut *C. arabica* et *C. eugenioïdes*). L'analyse des séquences ITS permet le regroupement d'espèces du genre *Coffea* de la région basse Guinée/Congo avec notamment *C. canephora* et *C. arabica* (qui vient d'Ethiopie).*

A



B



1.1.3 Aires de distribution naturelle

Les aires de répartition naturelle des espèces du genre *Coffea* s'étendent d'ouest en est de la Guinée au Soudan et du nord au sud du Cameroun en Angola. Elles se répartissent en trois groupes biogéographiques séparés par le canal du Mozambique et la dorsale de Kivu, à l'est de la République Démocratique du Congo (ex Zaïre) (Berthaud, 1986b; Louarn, 1992; White, 1979a). Les trois espèces cultivées occupent des territoires naturels bien distincts. *C. arabica* est une plante originaire de l'Afrique de l'est, *C. canephora* est retrouvée à l'ouest et au centre. L'espèce *C. liberica* Hiern est scindée en deux sous-espèces : *C. liberica* ssp. *liberica* et *C. liberica* ssp. *dewevrei*. *C. liberica* ssp. *liberica* est retrouvée à l'ouest, et *C. liberica* ssp. *dewevrei* au centre (Figure 2.4). Cette répartition actuelle serait liée à l'histoire évolutive des caféiers, et chaque espèce montre une adaptation particulière à un climat et un environnement spécifique. La dynamique de spéciation des caféiers est fortement liée aux grands changements climatiques (Robbrecht, 1996a) qui ont eu pour conséquence des cycles d'expansions-rétractions des forêts africaines, notamment pendant l'époque glaciaire.

L'espèce *C. canephora* Pierre (ou Robusta) est présente dans des forêts tropicales humides (températures élevées diurnes et nocturnes) à faible voire moyenne altitude (250m à 1500 m). Elle possède une variabilité génétique considérable (Berthaud, Charrier, 1988).

L'espèce *C. arabica* est originaire des forêts d'altitude d'Afrique de l'est (en Ethiopie, au Soudan et au Kenya), et s'est adaptée à plus haute altitude (de 950m à 2000m), où les températures nocturnes sont fraîches. La variabilité des populations naturelles de *C. arabica* a été constatée par les scientifiques ayant visité et exploré la région sud-ouest de l'Ethiopie (Charrier, Berthaud, 1985). Ces forêts sont actuellement menacées de disparition (Davis *et al.*, 2006). Or *C. arabica* est l'espèce la plus cultivée au monde et a été diffusée à partir de quelques génotypes seulement. En effet, des analyses moléculaires récentes ont montré la présence d'un faible polymorphisme au sein des cultivars commerciaux actuels (Anthony *et al.*, 2002). Cet important goulot d'étranglement est encore accentué par la faible variabilité naturelle de l'espèce. Cette espèce produit le café le plus apprécié, en termes de qualité à la tasse, mais présente de grandes sensibilités aux ravageurs de culture.

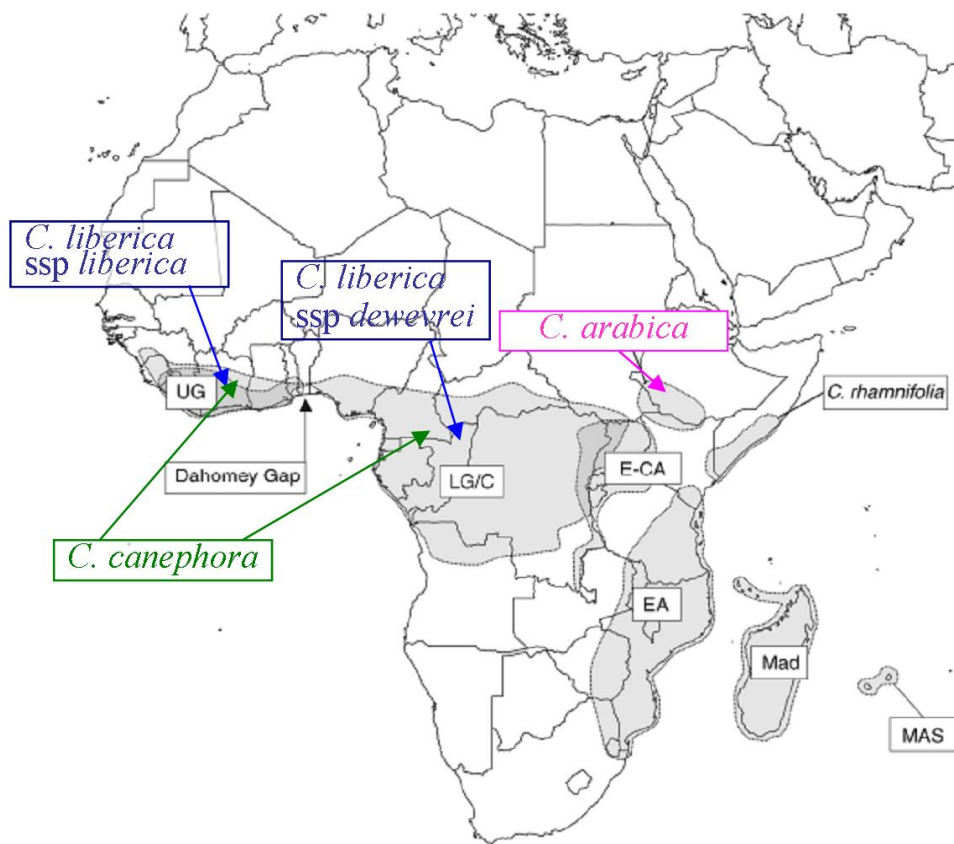


Figure 2.4 : Distributions naturelles des espèces en Afrique selon Maurin et al (2007)

1.2 Caractéristiques génétiques des trois espèces cultivées

1.2.1 Diversité génétique

La richesse du genre *Coffea* a été analysée chez les caféiers endémiques de la région malgache (Charrier, 1978) et d'Afrique occidentale et centrale (Berthaud, 1986) puis en Afrique de l'est (Anthony, 1992). Les premières analyses de diversité étaient basées sur des analyses morphologiques, puis sur des analyses phénologiques (intervalle de temps entre l'induction des bourgeons floraux et la floraison déterminée par le nombre de fleurs ouvertes chaque jour après une pluie inductrice), puis sur des analyses de diversité biochimiques (dosage de la caféine ou d'acides chlorogéniques) ou encore sur la diversité des isozymes (systèmes enzymatiques) (Anthony, 1992).

Concernant l'espèce *C. canephora*, les premières analyses de la diversité génétique ont été faites sur la base de la variabilité isoenzymatiques (Berthaud, 1986) et deux sous-groupes de diversité ont

été révélés : le groupe guinéen et le groupe congolais. Par la suite, en augmentant le nombre de génotypes analysés, le groupe congolais a été subdivisé en deux sous-groupes de diversité SG1 et SG2 (Montagnon *et al.*, 1992a).

La diversité moléculaire chez *C. canephora* a été évaluée ensuite avec des marqueurs RFLP (Restricted Fragment Length Polymorphism). La structuration établie par la diversité des isozymes est conservée mais affinée avec cinq groupes de diversité clairement définis (Dussert *et al.*, 1999) (Tableau 2.1).

Tableau 2.1 : Récapitulatif des groupes de diversité mis en évidence par les différentes études quant à la structuration de l'espèce *C. canephora*, d'après Dussert *et al.* (1999)

(Berthaud, 1986)	(Montagnon <i>et al.</i> , 1992)	(Dussert <i>et al.</i> , 1999)	(Dussert <i>et al.</i> , 1999)	Correspondance Géographique
Isozymes	Isozymes	Isozymes	RFLP	
Guinéens	Guinéens	1	D	Guinée et Côte d'Ivoire
Congolais	SG1	3	A	Façade atlantique de l'Afrique centrale
	SG2	2	E	Bassin du fleuve Congo
			B	Sud Centrafrique, nord RDC
			C	Sud-ouest Centrafrique, nord-ouest RDC, est Cameroun

L'espèce *C. liberica* et ses deux sous espèces ont été étudiées selon le polymorphisme AFLP (Amplified fragment Length Polymorphism) (N' Diaye *et al.*, 2007). Les deux sous-espèces sont autant différenciées génétiquement l'une de l'autre qu'elles ne le sont chacune de l'espèce *C. canephora* (N' Diaye *et al.*, 2005).

L'utilisation des marqueurs moléculaires chez l'espèce *C. arabica* a débuté avec des marqueurs RAPD (Random Amplified Polymorphic DNA) (Lashermes *et al.*, 1996). La diversité génétique sauvage de cette espèce a ensuite été étudiée plus précisément toujours avec des marqueurs RAPD ce qui a montré une nette séparation entre le matériel éthiopien et les variétés Typica et Bourbon (Anthony *et al.*, 2001). Une analyse génétique de l'espèce *C. arabica* a montré que le polymorphisme révélé par les RAPD, les AFLP et les SSR (Simple Sequence Repeat ou microsatellites) est similaire au sein des accessions cultivées de cette espèce (Maluf *et al.*, 2005).

Récemment, une étude basée sur les microsatellites a révélé une différenciation entre les accessions de part et d'autre de la vallée du grand Rift en Ethiopie (Silvestrini *et al.*, 2007).

Actuellement, les marqueurs moléculaires les plus utilisés sont les marqueurs microsatellites. Beaucoup d'études ont déjà été faites sur le genre *Coffea*, quant à la mise au point de marqueurs microsatellites génomiques ou issus de séquences codantes (Expressed Sequence Tag) sur les espèces *C. arabica* ou *C. canephora* et quant au transfert des marqueurs au sein du genre. Une base de données complète recensant tous les microsatellites dessinés à partir de séquences de *C. canephora* ou *C. arabica* a vu le jour récemment : <http://moccadb.mpl.ird.fr/index.php?cat=0> (Plechakova *et al.*, 2009). Le tableau ci-dessous résume l'ensemble des travaux de recherche de microsatellites en vue d'étudier la diversité des espèces du genre *Coffea* (Tableau 2.2).

Alors qu'un grand nombre de marqueurs est maintenant disponible pour toute analyse de diversité, un important travail de sélection pertinent doit être fait avant chaque étude, en fonction des objectifs visés.

Tableau 2.2 : Synthèse des marqueurs microsatellites mis au point sur *C. arabica* ou *C. canephora*

Acronyme utilisé par les auteurs	Type de séquence	Origine de la séquence (<i>Coffea</i>)	Nbre marqueurs	Référence	Taxa testés
CofEST-SSR	EST	<i>C. canephora</i> x <i>C. congensis</i>	9	(Bhat <i>et al.</i> , 2005)	11 <i>Coffea</i> sp.
ES	EST	<i>C. canephora</i>	99	(Poncet <i>et al.</i> , 2006)	4 <i>Psilanthus</i> sp. 7 <i>Coffea</i> sp. 21 Rubiaceae sp.
SSR	EST	<i>C. canephora</i>	10	(Geromel <i>et al.</i> , 2006)	<i>C. arabica</i>
DCM / CofEST-SSR	EST	<i>C. sp.</i>	9	(Aggarwal <i>et al.</i> , 2007)	11 <i>Coffea</i> sp. 4 <i>Psilanthus</i> sp.
Ssr	EST	<i>C. canephora</i>	132	(Lefebvre-Pautigny <i>et al.</i> , 2010)	3 <i>Coffea</i> sp.
M	Genomic	<i>C. arabica</i>	10	(Combes <i>et al.</i> , 2000)	
M	Genomic	<i>C. arabica</i>	17	(Coulibaly <i>et al.</i> , 2003)	2 <i>Coffea</i> sp.
CM	Genomic	<i>C. arabica</i>	9	(Baruah <i>et al.</i> , 2003)	11 <i>Coffea</i> sp. 4 <i>Psilanthus</i> sp.
CFGa	Genomic	<i>C. arabica</i>	34	(Moncada, McCouch, 2004)	<i>C. arabica</i>
M	Genomic	<i>C. arabica</i>	77	(Poncet <i>et al.</i> , 2004)	6 <i>Coffea</i> sp.
DL	Genomic	<i>C. canephora</i>	8	(Leroy <i>et al.</i> , 2005)	<i>C. canephora</i>
M	Genomic	<i>C. canephora</i>	213	(Poncet <i>et al.</i> , 2007b)	3 <i>Coffea</i> sp.
	Genomic	<i>C. arabica</i>	9	(Lashermes <i>et al.</i> , unpublished data)	
Total			638		

1.2.2 Programmes d'amélioration : intra-spécifique et inter-spécifique

Des programmes d'amélioration des deux principales espèces cultivées pour la production de café *C. arabica* et *C. canephora* tentent d'optimiser les caractères agronomiques et la diversité génétique au travers de croisements intra-spécifiques (plutôt pour *C. canephora*) et de croisements inter-spécifiques (plutôt pour *C. arabica*). Les différents groupes de diversité de *C. canephora* ont été initialement caractérisés morphologiquement puis des études de fertilité des descendants issues de croisements inter et intra-groupes ont été menées (Anthony, 1992). Par la suite ces croisements ont été évalués. Les principaux programmes d'amélioration génétique du café arabica se basent sur les croisements interspécifiques contrôlés afin d'introgesser les caractères agronomiques intéressants d'une espèce à l'autre (Leroy *et al.*, 1993; Montagnon *et al.*, 1998; Montagnon *et al.*, 1992).

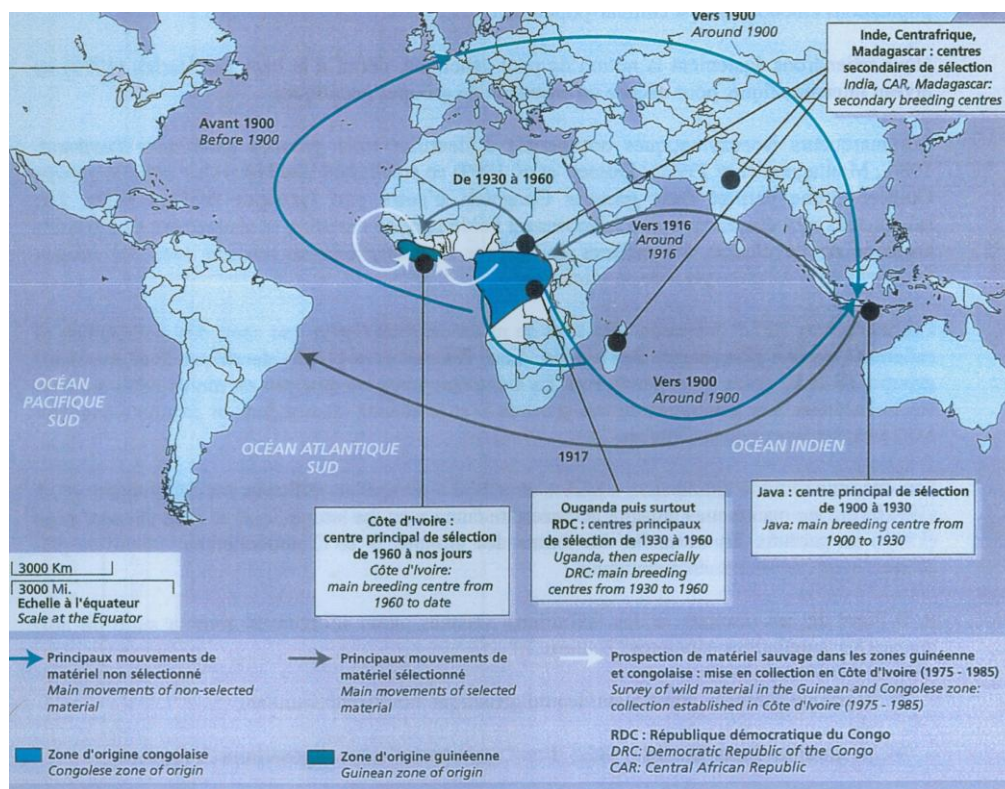


Figure 2.5 : Principaux centres de sélection et mouvements de matériel de *C. canephora* à travers le monde (issu de Montagnon (1998))

Les programmes de sélection intra-spécifique ont été menés sur l'espèce *C. canephora*. Elle a initialement été cultivée surtout au Congo Belge (actuellement République démocratique du Congo) et en Ouganda. Les accessions ont ensuite été transférées à Java, principal centre de sélection de *C. canephora* de 1900 à 1930 (Montagnon *et al.*, 1998) (Figure 2.5). En Afrique, l'utilisation des formes spontanées locales a permis d'élargir l'éventail du matériel mis en culture (Kouilou en Côte d'Ivoire, Niaouli au Togo et Benin, Nana en République Centrafricaine). Le matériel sélectionné à Java a ensuite été réintroduit en Afrique à l'INEAC (Institut National pour l'Etude Agronomique du Congo) qui est alors devenu le principal centre d'amélioration de 1930 à 1960. Bien que sélectionné, le matériel cultivé est resté très proche génétiquement du matériel sauvage mais toute la diversité sauvage n'est pas représentée dans le matériel cultivé (Dussert *et al.*, 1999). Un programme d'amélioration par sélection récurrente réciproque exploitant la complémentarité des groupes guinéens et congolais a été initié en 1984 (Leroy *et al.*, 1993). Les résultats montrent de réelles possibilités d'amélioration des populations de base (Leroy *et al.*, 1997b; Montagnon, 2000).

C. arabica a été améliorée principalement avec des croisements inter-spécifiques. Pour obtenir une résistance durable à la rouille orangée, les chercheurs ont introduit chez cette espèce des gènes de *C. canephora* par hybridation artificielle des deux espèces, suivi de rétro-croisements avec *C. arabica* (Herrera *et al.*, 2002a; Herrera *et al.*, 2002b; Herrera *et al.*, 2004). Une autre stratégie d'introgession a consisté à croiser *C. arabica* avec l'hybride de Timor (hybride naturel entre *C. canephora* et *C. arabica*). Ces deux méthodes ont permis d'obtenir les hybrides « Icatu » au Brésil et « Catimor » (Bertrand, 2002). D'autres croisements inter-spécifiques ont été utilisés pour l'amélioration de *C. canephora* en l'hybridant avec *C. congensis*, pour obtenir l'hybride « Congusta » qui présente un niveau de fertilité élevé.

La mise en culture de *C. arabica* et la sélection ont contribué à une forte perte de diversité génétique initialement pas très élevée car *C. arabica* est une espèce relativement récente (Figure 2.6).

Ces travaux d'amélioration ont apporté des informations intéressantes quant à la différenciation des espèces et à leurs barrières reproductives (niveau de fertilité des hybrides). Qu'en est-il des barrières naturelles aux hybridations inter-spécifiques en termes de floraison essentiellement ?

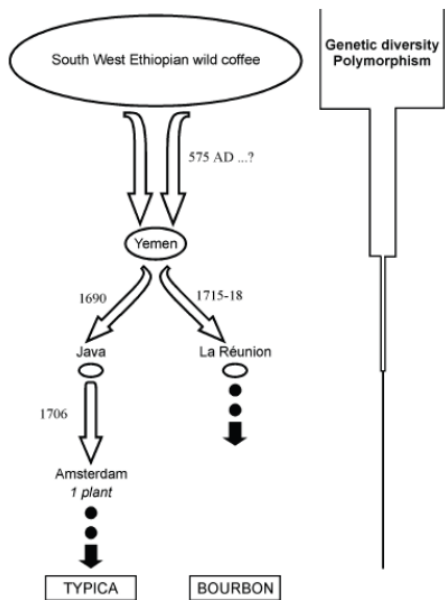


Figure 2.6 : Historique de la culture de *C. arabica* en relation avec la diversité génétique du compartiment cultivé. Tiré de Anthony et al.(2002)

1.2.3 Barrières naturelles à la reproduction

Les principaux freins aux échanges géniques interspécifiques observés interviennent au niveau de la régularité de la méiose ou à cause de gènes de stérilité (Louarn, 1987). Différentes études sur les caféiers ont montré qu'il existe de nombreux freins aux échanges géniques entre les espèces : aires de répartition différentes, asynchronisme des floraisons (Berthaud *et al.*, 1989), etc...

Le premier frein évoqué est la distance géographique entre des populations d'espèces différentes. De par leur adaptation écologique particulière et leurs aires de répartition, certaines espèces n'ont jamais de contacts génétiques avec les autres. L'existence de populations de caféiers multi-spécifiques naturelles a été décrite suite à la série de prospections effectuées avant 1985. Des associations naturelles de *C. liberica* et *C. canephora* ont été observées en Côte d'Ivoire (Ghapleu) ou en association avec d'autres espèces du genre *Coffea* : *C. congensis* en République Centre Africaine (Bangui), ou *C. eugenoides* à Madagascar (Charrier, Berthaud, 1985). Ces observations ont permis de confirmer que les hybridations entre différentes espèces de caféiers dans le milieu naturel étaient très sporadiques avec seulement quelques cas à Madagascar et en Côte d'Ivoire observés dans des lots de graines. Ces hybrides interspécifiques sont supposés ne pas présenter une combinaison viable dans la mesure où aucun hybride interspécifique n'a encore été échantillonné

dans le milieu naturel (Charrier, Berthaud, 1985). Des clés morphologiques permettaient alors de caractériser des hybrides inter-spécifiques.

Le deuxième frein évoqué est l'absence de synchronisme des floraisons. Les floraisons des caféiers sont déclenchées par une pluie en saison sèche. Les floraisons des caféiers sont sous la dépendance de deux processus physiologiques successifs antérieurs à l'ouverture des fleurs : l'induction et le déclenchement. La mise en place de la floraison commence par la mise en place des inflorescences. Trois à cinq bourgeons d'inflorescences se forment à la base des feuilles, avec quatre à cinq fleurs par inflorescence. Les inflorescences se mettent en place puis arrêtent de croître et entrent dans une période d'induction qui peut durer jusqu'à 30 jours et qui correspond à la mise en place des bourgeons floraux (Arcila-Pulgarin *et al.*, 2002). Cette induction nécessite un stress hydrique. L'activation physiologique de la floraison nécessite une pluie d'au moins 5 mm après une période sèche préférentiellement stricte (Le Pierrès, 1995). La durée entre la pluie et l'anthèse est une caractéristique spécifique. Pour ce critère, la très grande variabilité interspécifique rend compte des facultés d'adaptation du genre *Coffea* à des milieux très divers. A Divo, en Côte d'Ivoire, l'ouverture des fleurs (ou floraison), et simultanément l'anthèse, se produisent toujours 5 à 6 jours pour *C. liberica*, 7 jours pour *C. canephora* et 7 à 8 jours pour *C. arabica* après la première pluie suivant une période sèche d'au moins 30 jours (Baranski, 1996) (Figure 2.7). Aussi, *C. liberica* a tendance à fleurir plus tard dans l'année, alors que *C. canephora* fleurit habituellement tôt dans l'année (Charrier, Berthaud, 1985).

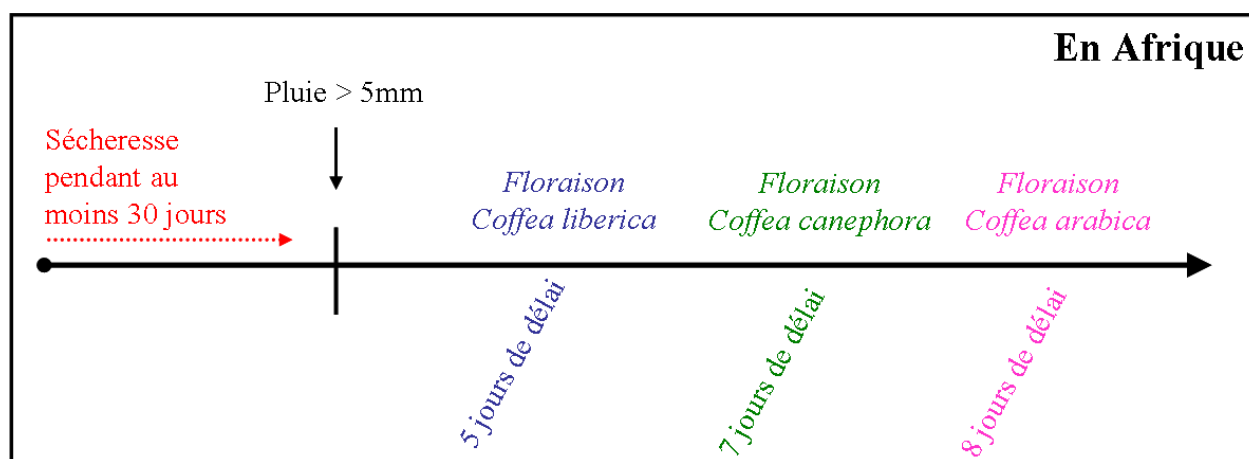


Figure 2.7 : Délais de floraisons entre espèces observés en Afrique après une pluie rompant la saison sèche

Les caractéristiques florales des trois espèces sont différentes et ont été décrites selon plusieurs paramètres morphologiques (Bridson, Vercourt, 1988; Sonke, Stoffelen, 2004; Stoffelen et al., 1997) (Tableau 2.3).

L'anthèse est aussi plus ou moins longue selon les espèces. Il semblerait que l'anthèse soit d'un jour plus longue pour *C. canephora* que pour *C. liberica* (Charrier, Berthaud, 1985). L'anthèse ne dure jamais plus de deux jours. Par contre le stigmate reste réceptif pendant 3 à 5 jours environ pour toutes les espèces.

Des anomalies, comme des déclenchements de floraisons en continu tout au long de l'année, peuvent permettre des floraisons simultanées, notamment quand le régime pluviométrique n'a pas de saison sèche bien marquée, comme il semblerait que ce soit le cas en Nouvelle-Calédonie (Charmetant, Le Pierrès, 1991). C'est d'ailleurs le cas dans beaucoup de régions équatoriales où les saisons sèches et humides ne sont pas bien alternées (Arcila-Pulgarin et al., 2002).

Tableau 2.3 : Caractéristiques florales des trois espèces tirées des références (Bridson, Vercourt, 1988 [1]; Sonke, Stoffelen, 2004 [2]; Stoffelen et al., 1997 [3]) annotées, basé sur la revue bibliographique de Camille Guiziou (communication personnelle).

	<i>C. arabica</i>	<i>C. canephora</i>	<i>C. liberica</i>
Nombre d'inflorescence par axillaire	(1-)2-4 [3]	(1-)2-4(-7) [2]	1-3(-4) [2]
Nombre de fleurs	(1-)4-7 [3]	(3-)4-9 [2]	1-6(-9) [2]
Nombre de calicules par inflorescence	-	1 [2]	2(-3) [2]
Longueur du tube de la corolle (mm)	(5-)9-11 [1]	(5-)6-16 [2]	7-18 [2]
Lobe de la corolle (mm)	9-16 x (2-)4-6 [1]	(5-)8-20 x (1.8)-2.5-4(-5.5) [2]	2-4 [2]
Anthères (mm)	(5-)7-10 [3]	(4-)5-8 [2]	7.5 [2]
Filaments des anthères (mm)	(1-)2-5 [3]	1-5 [2]	4 [2]
Présence d'un calicule autour de l'ovaire	Absent ou juste à la base [3]	Absent ou juste à la base [2]	Absent [2]
Style + stigmate (mm)	(13-)14-32 [3]	(11-)15-30 [2]	18-38

II- Le café en Nouvelle Calédonie

2.1 Historique des introductions et de la culture du café

Les premiers caféiers *arabica*, provenant de l'île de La Réunion, ont été introduits en Nouvelle-Calédonie par des moines Maristes en 1856 (Charles, 1932). De 1860 à 1870, la superficie des caféières s'étend sur 50 hectares. La première vague de colonisation en Nouvelle-Calédonie est essentiellement pénale avec la déportation de prisonniers installés dans des bagnes construits sur l'île. La pénitencière (chargée de gérer les personnes issues de la colonisation pénale) commence à s'intéresser au café et la caféiculture sera vraiment lancée lors de la régression de la canne à sucre. Les prisonniers seront alors chargés des travaux concernant les plantations de café. Le café gagne les villes de Bourail, La Foa, Pouembout et Farino. Cette dernière ville devient le principal centre caféicole de la colonisation pénale. Les exportations commencent en 1890, et la caféiculture s'intègre dans le paysage agricole calédonien. Augustin Bernard commente, en 1965 (Atlas, 1981): « On est revenu récemment au café...Le café est le nickel de l'agriculture ».

Ensuite, la deuxième vague de colonisation fut la colonisation libre, consistant à attirer des familles françaises pour s'installer en Nouvelle-Calédonie. C'est le gouverneur Feillet qui est l'initiateur d'une telle dynamique. En 1884, le décret attribue 110 000 hectares au gouverneur incluant les meilleures vallées de l'île. En 1897, le gouverneur estime que le territoire recèle de 150 000 hectares de bonnes terres et estime que la production de café peut s'élever à 90 000 tonnes. Il substitue donc à la colonisation pénale, la colonisation agricole d'immigrants libres (concessions de 25 hectares). Mais les possibilités agricoles du sol avaient été surestimées. Ainsi, le manque de main d'œuvre se fit sentir et les cours du café chutèrent alors que les prix de la main d'œuvre doublèrent.

En 1910, l'arrivée de la rouille (*Hemilia vastratrix*) sur le territoire ravage ces cultures. Les caféiculteurs introduisent alors de nouveaux génotypes potentiellement plus résistants dès 1915, ainsi que les deux autres espèces cultivées : *C. canephora* depuis le jardin de Buitenzorg à Java et en moindre quantité *C. liberica*. La conjonction de l'introduction de *C. canephora* et *C. liberica* ainsi que la sélection « naturelle » des *C. arabica* sous la pression sélective de la rouille (Pellegrin *et al.*, 1983) permet de surmonter la crise. La maladie s'atténue et la production peut repartir. La production culmine à plus de 2 000 tonnes en 1939. Jusqu'en 1930, le café a été cultivé presque exclusivement par des européens. Dès 1925, la promotion mélanésienne commença, pour intégrer les tribus à l'économie monétaire. Le café va être planté dans les réserves et va servir de

justification économique à leur extension. Le fond caféicole se trouve réactivé avec 2000 hectares de plantations mélanésiennes en 1934. La seconde guerre mondiale a contribué à raréfier la main d'œuvre, par l'embauche des mélanésien dans les camps américains, et par conséquent renchérir son prix par l'inflation. Deux conséquences indirectes des conflits apparaissent, à savoir l'abolition du régime de l'indigénat (les kanaks deviennent des citoyens français à part entière) et l'arrêt des engagements de travailleurs asiatiques. Le métayage se généralise et les planteurs blancs se désintéressent du café et commencent leur exode rural. En revanche, les plantations de café mélanésiennes continuent à être exploitées et ceci servira de justification économique à l'agrandissement des réserves. La part de la production mélanésienne dans la production totale passe de 35% en 1950 à 58% en 1965. Mais cet élan commence à s'essouffler vers 1950 en raison de l'emploi urbain et industriel. L'effondrement est corrélé à l'explosion minière et l'implantation tardive de l'institut Français du Café et du Cacao ne pourra l'enrayer. Même si le déclin du café commença dès 1960, le café resta jusqu'en 1970 la principale production agricole de la Nouvelle-Calédonie. A cette époque il restait 6 000 hectares, avec 70% de *C. canephora* et 30% de *C. arabica*. En 1969, 2000 tonnes de café étaient produites dont 75% provenaient de plantations de colons et 25% de plantations mélanésiennes.

L'opération café a commencé en 1978, sous l'impulsion de Paul Dijoud alors secrétaire d'Etat aux DOM-TOM (Deuss, 1987). Cette opération a consisté à remplacer la caféiculture sous ombrage, majoritairement mélanésienne, aux faibles rendements et confrontée au manque de main d'œuvre, par une caféiculture sans ombrage, utilisant des engrais et susceptible de mécanisation. Ce projet s'insère dans les projets de développement amorcés par l'OIC, offrant des primes sur trois ans et comprenant défrichement, fourniture des plants, des engrais, et entretien. En général, en Nouvelle-Calédonie, la caféiculture s'apparente beaucoup à un programme social, et elle doit être subventionnée par l'Etat pour se maintenir. Mais les mélanésien sont alors confrontés à une culture qui nécessite 80 heures de présence par mois et par hectare, soit beaucoup plus que la caféiculture sous ombrage. En considérant, par ailleurs, la part du temps de travail qui est destinée aux activités coutumières, il n'est pas permis de penser que la caféiculture intensive puisse enrayer le déclin caféicole ou qu'elle puisse offrir au milieu mélanésien une alternative à l'exode rural et à la recherche du salariat. Cette opération fut un échec dont la cause principale est à rechercher dans l'inadéquation de la logique du développement à celle de la société mélanésienne (David, Pillon, 1999; David, Siapo, 1994; Deuss, 1987; Kolher, Pillon, 1983; Pillon, 1985a; Pillon, 1985b).

Actuellement en Nouvelle-Calédonie, deux tendances sont observées. En Province Nord, exclusivement sur la côte est, d'anciennes plantations de *C. canephora* sont toujours cultivées. Dans le même temps, de nouvelles plantations en plein soleil de *C. arabica* sont exploitées. Depuis 1972, un système coopératif, le GAPCE (Groupement Agricole des producteurs de la Côte Est) est mis en place afin que plusieurs petits exploitants mettent leurs efforts en commun pour torréfier et commercialiser leur café. En Province Sud en revanche, la majorité des plantations sont en plein soleil ou avec des ombrages modérés et seule l'espèce de *C. arabica* est cultivée avec plusieurs variétés (Le Roy, bourbon jaune ou bourbon rouge). Alors que certains exploitants optent pour l'intensification et la mécanisation, d'autres continuent de récolter à la main et peuvent enrichir leurs plantations avec des arbres d'ombrage.

En 2008, la production totale de Nouvelle-Calédonie est de 33 tonnes pour *C. arabica* et *C. canephora* confondues. La production en Province Nord de *C. canephora* a augmenté par rapport à 2007 (30 tonnes contre 6 tonnes). Par contre la production de *C. arabica* de la Province Sud a chuté de 15 tonnes à 3 tonnes seulement. Ces chiffres pour *C. arabica* concernent 7 exploitations pour une surface d'environ 23 hectares (rapport de la Direction du Développement Rural ou DDR de la Province Sud de la Nouvelle-Calédonie, daté du 30 mars 2009). Actuellement des plants peuvent être proposés aux agriculteurs ayant déposé un projet de suivi pour commencer leurs exploitations ou tester une nouvelle variété. Tout cela est financé en partie par la DDR et géré par l'UPCC (Unité de Promotion de la Caféiculture Calédonienne) qui rassemble les caféiculteurs indépendants de tout le territoire pour trouver des solutions communes avec des financements communs.

2.2 L'abandon des cultures et la mise en place des zones hybrides à Sarraméa

En Nouvelle-Calédonie, la culture du café s'est étendue sur toute l'île, avec des plantations de *C. canephora* ou *C. arabica*. Finalement, la majorité des vieilles plantations, après le dernier essai de relance en 1980, a été abandonnée. Certaines de ces anciennes plantations sont composées des deux espèces *C. arabica* et *C. canephora* et quelques fois de l'espèce *C. liberica*.

L'environnement au sein duquel évoluent ces anciennes plantations apparaît favorable à deux, voir trois espèces. Selon les études de Pellegrin (1983) et Tercinier (1958), les versants est et ouest ne possèdent pas les mêmes conditions environnementales en termes de climats, de végétation et de sols. *C. canephora*, bien que présente sur toute l'île et n'ayant pas d'exigences environnementales très fortes, semblerait mieux adaptée aux conditions climatiques de la côte est alors que *C. arabica*

s'accommoderait mieux du climat de la côte ouest (Tercinier, 1958). Seul le versant ouest posséderait des caractéristiques environnementales favorables à la fois à *C. arabica* et *C. canephora* (Pellegrin *et al.*, 1983). Par ailleurs, l'absence de pressions de ravageurs de cultures est déterminante pour la survie de *C. arabica*. Il apparaît que la pression de la rouille orangée (*Hemilia vastatrix*) est plus importante sur la côte est de l'île. Les plantations de *C. arabica* y furent complètement décimées lors de la vague de rouille de 1910 alors qu'elles subsistèrent dans les régions de la côte ouest, notamment dans la chaîne montagneuse centrale (Pellegrin *et al.*, 1983). Les dernières anciennes plantations multi-spécifiques se situent dans une zone particulière, sur le versant ouest, dont la dynamique actuelle correspond à celle d'un milieu naturel (Figure 2.8).

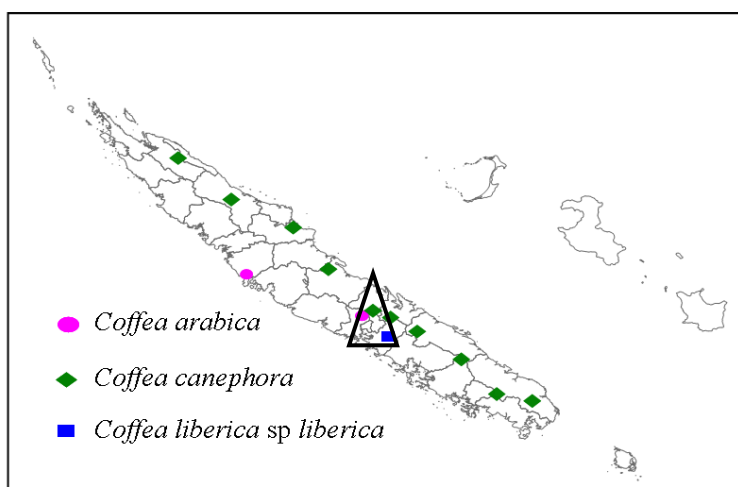


Figure 2.8 : Distribution des trois espèces (au sein de plantations abandonnées) en Nouvelle-Calédonie selon les prospections de Daniel Le Pièrres (1999)

Les prospections en Nouvelle Calédonie ont en effet mené à choisir la région de Sarraméa. Cette région est située sur la côte ouest sur les contreforts de la chaîne montagneuse qui partage la Grande Terre sur toute sa longueur.

Des zones hybrides se sont alors mises en place dans ces anciennes plantations, où les deux espèces *C. arabica* et *C. canephora* (exceptionnellement trois avec *C. liberica*) du genre *Coffea* cohabitent et s'adaptent. Lors des premières prospections dans cette région et ailleurs sur l'île, les premiers hybrides inter-spécifiques furent caractérisés (Jagoret *et al.*, 1999; Le Pièrres, 1999). Des évaluations génétiques des hybrides entre *C. arabica* et *C. canephora* ont été menées afin d'évaluer l'introgression de *C. canephora* sur *C. arabica* (Mahe *et al.*, 2007). Il apparaît que seuls des cas de flux de gènes de *C. canephora* vers *C. arabica* aient été révélés par l'analyse des hybrides à niveaux d'introgression avancés. Les hybrides prospectés au sein de ces anciennes plantations

peuvent être résistants à la rouille et aussi robustes que des *C. canephora*, mais peuvent être improductifs pour des problèmes de stérilité (Le Pierrès, 1999). De nos jours, le seul hybride naturel connu entre *C. arabica* et *C. canephora* est l'hybride de Timor (Bettencourt, 1973).

L'une des caractéristiques importantes de ces anciennes plantations est qu'elles sont situées sous un couvert végétal dense constitué d'arbres indigènes et de légumineuses introduites et ajoutées aux forêts primaires pour la culture du café. Il s'agit d'un système de polyculture dit rustique, comme décrit par Moguel et Toledo (1999), opposé à des systèmes de polyculture plus commerciaux ou des systèmes de monoculture plus intensifs. Le système de plantations rustiques correspond à un ensemble forestier très dense et très hétérogène. La situation à Sarraméa est particulière dans la mesure où cette région bénéficie à la fois des entrées maritimes du fait de sa situation en fond de vallée, mais aussi des influences de la chaîne avec un régime pluviométrique très élevé par sa position en contrebas du plateau de Dogny qui culmine à 1019 mètres d'altitude (Figure 2.9).



Figure 2.9 : Photos de la région d'étude représentant le plateau de Dogny

La zone d'étude est délimitée par l'emprise des images satellites Très Haute Résolution Quickbird. Les analyses spatiales présentées dans cette thèse se basent sur des données spatialisées telles que le Modèle Numérique de Terrain, les réseaux hydrographiques, les interpolations climatiques calculées à partir des stations de Météo France situées à proximité de la zone d'étude (Canala, La Foa, Pocquereux et Col d'Amieu) (Figure 2.10).

Quatre sites d'étude sont choisis comme témoins de la présence d'au moins deux espèces de caféiers et d'hybrides inter-spécifiques pour les modèles environnementaux mis en place. Trois

sites sont bi-spécifiques (Farino, Couli et Dogny) et un site est tri-spécifique (Sarraméa). Ce dernier, par son originalité, est choisi pour effectuer l'analyse génétique de la diversité des espèces introduites ainsi que le calcul du taux d'hybridation.

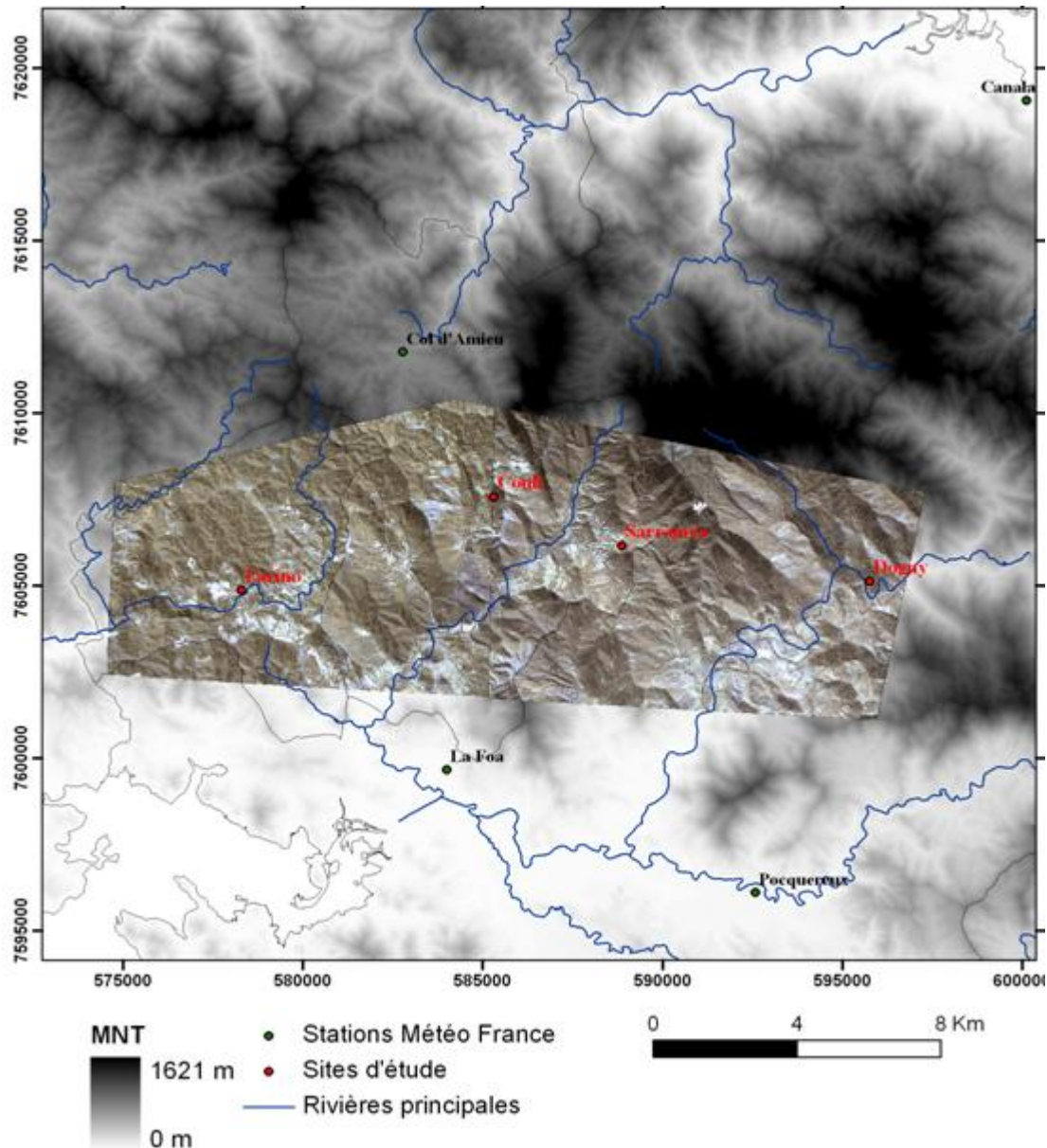


Figure 2.10 : Présentation de la zone d'étude et des données utilisées pour les analyses spatiales

III- Objectifs de l'étude

Le principal objectif de cette étude est de comprendre l'ampleur de la levée de barrières reproductives, en termes de taux d'hybridation et de distribution spatiale, chez des espèces de caféiers introduites en Nouvelle Calédonie. Cette thèse a pour but de contribuer à la compréhension des impacts des introductions de plantes dans de nouveaux écosystèmes. Dans cet objectif, le modèle d'étude « café » s'avère très intéressant.

D'un point de vue génétique, des flux de gènes ayant lieu entre les espèces, la dynamique de l'hybridation mérite d'être étudiée. Pour cela, la caractérisation de la diversité génétique globale des trois espèces étudiées, *Coffea canephora*, *C. arabica* et *C. liberica*, est réalisée afin d'évaluer la part de la diversité génétique introduite en dans la population de Sarraméa en Nouvelle-Calédonie et d'y estimer les flux de gènes.

D'un point de vue écologique, les trois espèces de caféiers ont clairement subi une modification de leurs niches écologiques entre l'Afrique et la Nouvelle-Calédonie. Nous étudions précisément les différences environnementales entre les milieux d'origine des espèces (en Afrique) et la région de Sarraméa en Nouvelle-Calédonie. Nous soulignons l'importance des caractéristiques environnementales de la région d'intérêt en décrivant les niches favorables aux flux de gènes.

D'un point de vue spatial, la distribution des niches favorables à ces flux de gènes entre caféiers peut être très réduite. Afin d'avoir une vision globale, il est important d'analyser cette distribution à l'échelle de la région d'intérêt entière. Cette prédiction de distribution passe par une description fine du couvert forestier sous lequel poussent les caféiers.

Les génotypes hybrides naturels observés seraient adaptés au milieu particulier de la Nouvelle-Calédonie, et pourraient représenter des opportunités de relance de la caféiculture.

D'un point de vue appliqué, ici, l'intérêt des croisements interspécifiques prend tout son sens dans l'ampleur du brassage génétique qu'il représente. Il pourrait alors être possible de combiner amélioration et adaptation, en associant les qualités des différentes espèces renforcées par la diversité et leur adaptation au milieu. C'est alors un modèle intéressant pour une gestion durable des ressources génétiques et la promotion d'une nouvelle caféiculture mélanésienne adaptée au fonctionnement de cette société.

Chapitre 3 - La caractérisation de la structure génétique et de la distribution naturelle de l'espèce Coffea canephora Pierre

Le travail présenté dans ce chapitre a fait l'objet d'une publication :

Gomez, C. et al. (2009) Current genetic differentiation of Coffea canephora Pierre ex A. Froehn in the Guineo-Congolian African zones: cumulative impact of ancient climatic changes and recent human activities. BMC Evolutionary Biology 9, 167



La première étape de notre démarche scientifique consiste à approfondir les connaissances sur la diversité naturelle de chaque espèce étudiée dans le cadre de cette thèse afin d'évaluer la diversité génétique introduite pour chacune dans nos populations de référence. Il s'agit donc, dans ce chapitre, de construire une collection de référence de la diversité de l'espèce *C. canephora*. Les collections de références de *C. arabica* et *C. liberica* ont déjà été mises au point à partir d'études antérieures (N'Diaye *et al.*, 2005 ; Maluf *et al.*, 2005).

I – Contexte de l'étude et démarche abordée

Afin de caractériser la diversité naturelle de l'espèce *C. canephora*, nous disposons d'un échantillon d'accessions sauvages suffisamment représentatif ainsi que d'outils moléculaires adaptés. Nous présentons ci-après ces données complémentaires, la partie centrale de l'étude étant présentée sous forme d'article (cf ci-après) :

- 1) l'historique de l'obtention des accessions et le choix de l'échantillon,
- 2) le choix des marqueurs moléculaires utilisés permettant de révéler le polymorphisme intra-spécifique mais aussi de détecter des hybrides inter-spécifiques pour la suite de nos recherches en Nouvelle-Calédonie et plus particulièrement à Sarraméa.

1.1 – Contexte historique de l'étude

Ce travail se base sur des prospections et des récoltes historiques de matériel. Plus de 700 génotypes sauvages ont été récoltés par l'ORSTOM (l'actuel Institut de Recherche pour le Développement - IRD) en collaboration avec le CIRAD (Centre de coopération Internationale en Recherche Agronomique pour le Développement), la FAO (Food and Agriculture Organization), l'IPGRI (International Plant Genetic Resources Institute) et le MNHN (Muséum National d'Histoire Naturelle) entre 1975 et 1987 dans cinq pays africains : la Côte d'Ivoire, la Guinée, le Cameroun, le Congo et la République Centrafricaine (Berthaud, Guillaumet, 1975 ; Le Pierrès, *et al.*, 1989 ; Berthaud, 1986).

Dans notre étude, nous avons travaillé sur 107 génotypes analysés auparavant par Dussert *et al.* (1999) grâce à l'utilisation de marqueurs moléculaires de type RFLP (Restriction Fragment Length Polymorphism). La diversité naturelle de cette espèce n'avait pas encore été décrite sur la base de marqueurs moléculaires de type microsatellite. Or, les marqueurs microsatellites, dans la mesure

où ils sont basés sur des techniques d'amplification, permettent des analyses avec peu d'ADN. Ils ont aujourd'hui remplacé les RFLP dans les analyses de diversité (Schlotterer, 2004). Ces deux techniques ont leurs avantages et leurs inconvénients, mais toutes deux offrent l'avantage de révéler des marqueurs co-dominants et spécifiques de locus. Les manipulations de type RFLP sont très laborieuses et nécessitent d'avoir extrait une quantité d'ADN importante. L'utilisation de microsatellites nécessite un important travail de séquençage en amont afin de mettre au point les amorces spécifiques, mais une fois mises au point, elles sont très souvent transférables entre les espèces du même genre et révèlent un niveau très élevé de polymorphisme.

La cohérence entre la structuration génétique révélée par les microsatellites et celle déjà révélée par les RFLP a été vérifiée. Cette étude permet notamment une comparaison fine des avantages et des inconvénients des deux types de marqueurs, dont le détail des résultats est présenté dans l'article ci-après. La principale différence entre les deux types de marqueurs concerne leur niveau de polymorphisme. Les estimations de la variabilité génétique sont supérieures pour les microsatellites. Aussi, les microsatellites sont plus performants pour la discrimination des individus à l'intérieur de la population (caractérisation des hybrides entre groupes de diversité). En fonction des résultats obtenus avec les deux types de marqueurs, une collection de référence a été retenue.

1.2 - Marqueurs moléculaires

De nombreux travaux ont permis de mettre au point 638 marqueurs microsatellites sur les espèces du genre *Coffea* (cf. Tableau 2.2, p 47, (Plechakova *et al.*, 2009)), ce sont autant de marqueurs disponibles pour les études de diversité de ces espèces.

Des marqueurs microsatellites développés à partir de séquences génomiques ou de séquences de type EST (Expressed Sequence Tag), ont été sélectionnés sur la base des données acquises par Poncet *et al.* (2004, 2006, 2007) avec le double objectif de :

- révéler le maximum de polymorphisme intra-spécifique (analyse de polymorphisme tel le PIC Polymorphism Information Content ou l'hétérozygotie attendue)
- différencier génétiquement les espèces entre elles (présence d'allèles diagnostiques uniques à une espèce ou gammes de variation de tailles des allèles distinctes entre espèces)

Il est alors possible, avec le même lot de marqueurs microsatellites, de caractériser la diversité naturelle des trois espèces en Afrique, d'évaluer la part de cette diversité introduite et de caractériser les hybrides interspécifiques dans la population de référence à Sarraméa en Nouvelle-Calédonie.

Avec le lot de 26 marqueurs microsatellites sélectionnés, il nous a été possible de différencier précisément les trois espèces *C. canephora*, *C. liberica* et *C. arabica*. La différenciation des deux sous-espèces *C. liberica ssp liberica* et *C. liberica ssp dewevrei* a aussi été permise. La détection et la caractérisation des hybrides entre ces trois espèces est alors possible pour la suite de nos recherches.

II – Résultats principaux

Les objectifs de ce chapitre étaient :

- 1) de caractériser la diversité présente au sein de l'espèce *Coffea canephora* Pierre ex A. Froehn en mettant en évidence les relations entre la variation génétique détectée et les signatures des changements climatiques passés sur l'ensemble de la végétation dans la région Guinéo-Congolaise du continent Africain,
- 2) d'évaluer l'impact des activités anthropiques sur les flux de gènes, la dispersion et la migration du matériel sauvage et cultivé par la détection d'hybrides putatifs entre groupes de diversité,
- 3) de définir une collection de référence comme témoin de la diversité spécifique africaine afin de caractériser la diversité introduite dans la population de référence à Sarraméa.

Les résultats obtenus ont fait l'objet d'une publication :

Gomez, C. *et al.* (2009) Current genetic differentiation of *Coffea canephora* Pierre ex A. Froehn in the Guineo-Congolian African zones: cumulative impact of ancient climatic changes and recent human activities. *BMC Evolutionary Biology* 9, 167

Research article

Open Access

Current genetic differentiation of *Coffea canephora* Pierre ex A. Froehn in the Guineo-Congolian African zone: cumulative impact of ancient climatic changes and recent human activities

Céline Gomez, Stéphane Dussert, Perla Hamon, Serge Hamon, Alexandre de Kochko and Valérie Poncet*

Address: UMR DIAPC, IRD, 911 avenue Agropolis, BP 64501, 34394 Montpellier CEDEX 5, France

Email: Céline Gomez - celine.gomez@ird.fr; Stéphane Dussert - stephane.dussert@ird.fr; Perla Hamon - perla.hamon@ird.fr; Serge Hamon - serge.hamon@ird.fr; Alexandre de Kochko - alexandre.dekochko@ird.fr; Valérie Poncet* - valerie.poncet@ird.fr

* Corresponding author

Published: 16 July 2009

Received: 19 December 2008

BMC Evolutionary Biology 2009, 9:167 doi:10.1186/1471-2148-9-167

Accepted: 16 July 2009

This article is available from: <http://www.biomedcentral.com/1471-2148/9/167>

© 2009 Gomez et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Among *Coffea* species, *C. canephora* has the widest natural distribution area in tropical African forests. It represents a good model for analyzing the geographical distribution of diversity in relation to locations proposed as part of the "refuge theory". In this study, we used both microsatellite (simple sequence repeat, SSR) and restriction fragment length polymorphism (RFLP) markers to investigate the genetic variation pattern of *C. canephora* in the Guineo-Congolese distribution zone.

Results: Both markers were first compared in terms of their informativeness and efficiency in a study of genetic diversity and relationships among wild *C. canephora* genotypes. As expected, SSR markers were found to have a higher genetic distance detection capacity than RFLP. Nevertheless, similarity matrices showed significant correlations when Mantel's test was carried out ($r = 0.66$, $p < 0.0001$). Finally, both markers were equally effective for group discrimination and phylogenetic studies, but SSR markers tended to outperform RFLP markers in discriminating the source of an individual among diversity groups and in putative hybrid detection. Five well defined genetic groups, one in the Upper Guinean forests, the four others in the Lower Guinean forests, were identified, corresponding to geographical patterning in the individuals.

Conclusion: Our data suggested that the Dahomey Gap, a biogeographical barrier, played a role in wild *C. canephora* differentiation. Climatic variations during the Pleistocene and/or Holocene probably caused the subgroup differentiation in the Congolese zone through the presence of a mosaic of putative refugia. Recent hybridization between *C. canephora* diversity groups, both for spontaneous individuals and cultivars, was further characterised according to their geographic dissemination or breeding history as a consequence of human activities.

Background

The species richness in some African zones has attracted attention on the origin of diversification in tropical forests [1-4]. In particular, the Guineo-Congolese regional center

of endemism (Zone I on White's map, 1983) includes about 8,000 spp., about 80% of which are endemic. This zone also corresponds to one of the 34 biodiversity hotspots defined by Mittermeier et al. [5]: the Guinean

forests, divided into the Upper Guinean and Lower Guinean forests by the Dahomey Gap. Analyses of patterns of geographical distribution of species richness and endemism of extant organisms could partly help in understanding these species composition patterns [e.g. [3]]. Species richness in the tropics has been attributed to the gradual accumulation and/or preservation of species over a long geological period in stable equatorial climates (the "museum model") [6-8] and/or to high speciation rates in response to late Tertiary geological events and unstable Pleistocene climates [9,10]. In fact, all major glacial advances in the Arctic resulted in great paleoenvironment and vegetation variations in the African tropics [2]. At least for the most recently evolved taxa, a certain amount of speciation and a great deal of subspeciation is said to have occurred during the late Pleistocene climatic fluctuations, especially since the last severe glaciation which cumulated around 18,000 B.P. [11]. To explain how plants and animal species survived during particularly cold and arid periods, Maley [2] proposed putative refugia, where the respective ancestors of the present species formed isolated populations, leading to allopatric, or geographical speciation [12]. Some of the refugia were located along the West African coast from Sierra Leone to Côte d'Ivoire, from southwestern Cameroon to western Gabon, in the eastern part of the Democratic Republic of the Congo (formerly Zaire) and along the Zaire river [2]. Through a study of the African Rubiaceae genera distribution, Robbrecht [13] gave additional support in favour of the refuge area concept, and demonstrated the importance of fluvial refugia in some taxa.

Molecular phylogenies and species-level diversity could be used to further unravel the forest diversification patterns. A population-genetics rather than a species-level approach has also been proposed to test the role of the relatively recent time frame of Pleistocene events [14]. Climatic changes have affected the genetic structure of many tree species in Europe [e.g. [15-17]]. For Africa, major advances are under way for vertebrates [e.g. [14,18]], but very few studies have focused on tree species of West and Central Africa [19-22], and only one has attempted to establish the relationship between the molecular genetic variation of a tree species, *i.e.* the shea tree (*Vitellaria paradoxa*), and the possible evolution of vegetation after the last glaciations in the Sudano-Sahelian region of Africa [19].

Coffea species (Rubiaceae) are endemic to intertropical forest zones in Africa, Madagascar, Mauritius, Comoros and Réunion [23-27]. There are over 103 accepted species, including the two most commonly cultivated species, *C. arabica* L. and *C. canephora* Pierre ex A. Froehn [28]. Within *C. arabica*, a predominantly self-pollinating (autogamous) species, there is a genetic structure with low dif-

ferentiation between accessions from the east and west side of the Great Rift Valley in Ethiopia as recently revealed with microsatellites [29]. The hybrid origin of *C. arabica* with *C. canephora* and *C. eugenoides* as the likely progenitors [30] was probably recent, and its colonization of Ethiopia likely occurred after the formation of the Great Rift Valley [31]. The systematic position and geographic distribution of *C. arabica* is isolated among diploid *Coffea* species [27,28]. On the contrary, *C. canephora* is one of the two *Coffea* species, along with *C. liberica* Hiern [28,32], with the widest natural distribution area of the genus. Moreover, these species are both diploid, allogamous (self-incompatible), and belong to the same phylogenetic clade [27]. They present overlapping geographical distributions, which extend west to east from Guinea to Uganda, and north to south from Cameroon to Angola [28,33]. This feature represents an exceptional resource for understanding the evolution and adaptation of tropical trees in these regions.

Few studies have been undertaken to assess the genetic diversity of natural *C. canephora* populations. Allozymic surveys revealed marked geographical clustering for two groups: the "Guinean" group, composed of populations from Côte d'Ivoire, and the "Congolese" group, consisting of two subgroups, SG1 and SG2, with populations from the Central African Republic and Cameroon [33,34]. Based on RFLP data, five wild *C. canephora* groups were further distinguished and the diversity structure was found to be conserved even when cultivated material, known as Robusta coffee, was included in the analysis [35]. Due to the increase in the number of microsatellite markers suitable for coffee species analyses [e.g. [36-38]], PCR-based assays are becoming increasingly attractive and compatible with the requirements of evolutionary studies or conservation genetics on large sets of genotypes.

Polymerase chain reaction (PCR)-based marker systems like simple sequence repeats (SSRs) have been widely used in recent years, replacing restriction fragment length polymorphism (RFLP) in DNA fingerprinting [39]. However, few studies have been carried out to compare the efficiency of RFLP and SSR markers for characterizing genetic diversity [40]. Both techniques offer the advantage of implementing nuclear co-dominant, locus-specific markers dispersed throughout the genome. Polymorphism detected by RFLP assays reflects restriction size variations, while microsatellite variants differ in the number of short (1-6 bp) amplified tandem repeats [41]. Their use in the assessment of population genetic structure has both advantages and drawbacks. While RFLPs assay are time consuming and labor intensive, among PCR-based markers, microsatellites require sequence information for primer pair design but, once developed, they are highly transferable across species, especially within the genus

[e.g. [42,43]]. Because of their extremely high level of polymorphism, they are probably the most efficient markers for fingerprinting, assignment tests or paternity analyses [44]. Differences in the resolution power of RFLP and SSR loci are thus expected because of differences in their mutation rates and processes. The simultaneous use of these molecular methods could help in inferring the signature of natural histories of organisms at different time scales, i.e. evolutionary history and historic migration patterns. Moreover, the recent development of Bayesian approaches in assignment tests, as implemented in STRUCTURE[45], has greatly increased the potential for understanding population structure.

In the present study, we investigated *Coffea canephora* genetic diversity across the West and Central African geographic range of the species. Spontaneous individuals are naturally distributed in two postulated refugia areas, i.e. the Upper Guinea and Lower Guinea/Congo regions, separated by the Dahomey Gap. Using both SSRs and RFLP loci, we evaluated the relative efficiency of these DNA-based marker systems, associated with different mutation rates, in resolving genetic diversity, population structure and gene flow among samples of *C. canephora*.

The final aims were:

- (i) to analyse the *C. canephora* genetic variability structure and assess any relationship between the molecular variation of the species and the possible signature of the past evolution of vegetation in the Guineo-Congolian region of Africa
- (ii) to evaluate the impact of human agricultural activities on gene flow, dispersal and migration of both wild and cultivated stocks through the detection of putative inter-diversity group hybrids
- (iii) to define core sets of *C. canephora* accessions that best capture the species diversity in terms of alleles.

Results

Efficiency of RFLP and SSR markers for polymorphism detection

In the 107 sampled *Coffea canephora* accessions (Figure 1, table 1, and table 2), genetic polymorphism indexes (N_a , H_o and H_e) at the 16 SSR loci and 8 RFLP loci were found to be highly variable throughout all geographic samples and types (Table 3). A total of 154 alleles across the SSR loci were detected, giving an average of 9.6 alleles per locus, ranging from 3 at M804 to 20 at M368. In comparison, the RFLP loci analysed gave 59 alleles with an average of 7.4 alleles per locus. The PIC value and gene diversity were quite even for SSR and RFLP loci, i.e. 0.62 vs. 0.59 and 0.65 vs. 0.63, respectively.

Structure of *C. canephora* natural diversity

Dissimilarity matrices were constructed on the whole sample based on shared-allele distances and revealed that the average genetic dissimilarities for RFLP (0.604) and SSR (0.626) markers were very similar. The average genetic dissimilarities obtained for the wild genotypes were also similar for both marker types, RFLPs (0.581) and SSRs (0.595). The correlation coefficient obtained with Mantel's test matrix correspondence test was, indeed, statistically significant ($r = 0.66$, $p < 0.0001$).

Genetic distance estimates obtained from the wild genotypes were used to derive Neighbour-joining trees for both the RFLP and SSR data. The topology of each tree (Figure 2) was unique but both trees discriminated the five *a priori* genetic groups (A, B, C, D, and E) from Dussert et al. [35]. Nevertheless, the SSR tree was not completely congruent with the RFLP tree. One difference was that group C, which occupied an intermediate position on the tree based on SSRs, occurred at a more distant position from groups [B-E-A] within group D based on RFLPs. Another contrast was noted with respect to the A individuals which clustered together near the E group based on SSRs and one of its individuals fell into the group E, based on RFLPs. The internal branches were generally longer for the RFLP tree. Despite these differences, both tree topologies reflected the same distinct clades, corresponding to a geographical patterning in the individuals. Plants from Côte d'Ivoire mostly clustered with plants from Guinea (group D). Individuals from northwest Congo, southeast Cameroon, and southwest Central African Republic fell within the same clade (group C), in line with their geographic proximities in Central Africa. The plants of Central African group C clustered near plants from West Africa (group D). From the eastern part of the geographical distribution, plants from northeast Congo clustered with plants from southeast Cameroon and south Central African Republic (group E), near plants from the southern border of Central African Republic (group B). The genotypes of a population from northwest Congo and a population from southwest Cameroon (group A) clustered near the B and E clades. Several individuals (marked with asterisks) appeared to be classified with groups other than their *a priori* groups. We performed a population structure analysis to further assess the group memberships of the plants and to detect actual migrants, hybrids or misclassified individuals.

The structure analysis using STRUCTURE with admixture showed that the five groups ($K = 5$) were genetically distinct based on SSR data (Figure 3). The results indicated that most of the plants had a high membership in their own cluster, with 97, 97, 69, 90, and 92% mean ancestry for the A, B, C, D, and E groups, respectively. In comparison, with the RFLP data, plants from groups B, C, and D

Table 1: Geographic and genetic origin of the wild *C. canephora* genotypes

Country	Location #	Name	Long.	Latit.	Coll. #	A priori genetic group	Putative hybrids
Côte d'Ivoire ^{1,2}	1	Ira 1	07.40	-07.46	0009 0028	D	
	2	Bafingdala	07.51	-07.41	0056	D	
	3	Ira 2	07.50	-07.42	0105 0159	D D	
	4	Gbapleu	07.34	-08.18	0121 0128	D D	
	5	Gao	06.57	-07.39	0136	D	
	6	Bossematie	06.30	-03.30	0145 0146	D D	
	7	Logbonou	08.04	-05.15	0186 0194	D D	DA DE
	8	Fourougbankoro	08.29	-05.45	0213 0233	D D	
	9	Goazra	07.00	-05.36	0236	D	
	10	Marahoué	06.54	-06.12	0245 0250	D D	
	11	Kouin	07.30	-07.18	0292 0293	D E	
	12	Géoulé	07.36	-07.56	0313 0315	D E	
	13	Gbapleu 1			0319 0321	E E	
	14	Gbapleu 2			0328	D	
	15	Gbapleu 3			0336	D	
	16	Dobia			0345 0350	D D	DEA DE
	17	Sabrégué			0354 0358	D D	DE
	18	Pélézi 2			0362 0395	D D	
	19	Gonaté			0404	D	
Guinea ²	20	Piné	08.57	-08.06	0803	D	
	20	Piné	08.57	-08.06	0808	D	
Cameroon ³	21	Koto	04.22	09.34	0651	E	

Table 1: Geographic and genetic origin of the wild *C. canephora* genotypes (Continued)

	22	Nguila	04.43	11.40	0658	C	DA
	23	Mwamepen	03.29	14.48	0662 0663	E A	AE
	24	Bitonga	03.19	15.16	0664	C	
	25	Boumba	02.03	15.10	0665 0666	E E	
	26	Nguilili	02.04	15.36	0678 0683	C C	
	27	Bombi	04.56	13.10	0685	C	
Central African Republic ⁴	28	Loukoussou			0501	B	
	29	Doungba			0504 0511	E B	
	30	Yombi			0516	E	
	31	Libengué			0518	B	
	32	Carnot			0602 0604	C C	CE
	33	Ndongué			9004	C	
Congo ⁵	34	Ouessou 2	01.35	14.48	0716	C	
	35	Ouessou	01.37	14.53	0721	C	
	36	Sembé-Souanké	01.56	14.11	0723	A	
	37	Souanké-gabon	02.07	14.00	0725 0727	E E	
	38	Boyélé 5			0729 0730	E E	
	39	Impfondo 1			0738	E	
	40	Impfondo 2			0739 0740	E E	

Membership to genetic groups, as previously defined by Dussert et al. [35] and putative hybrids detected in the present study are indicated for each accession.

¹Berthaud [33]; ²Le Pierrès et al. [26]; ³Anthony et al. [24]; ⁴Berthaud and Guillaumet [23]; ⁵de Namur et al. [25]

formed separated clusters with high membership in their own cluster (87, 91 and 88% on average, respectively). However, mixed ancestry between groups A and E at even proportions (A and E contributing to 50 and 44% ancestry on average) was observed, corresponding to a grouping by geographical regions. For this mixed group, we considered that plants possessing <80% ancestry in the [A+E] cluster were putative hybrids. Despite this discrepancy, the level

of admixture per group was very low and similarly estimated by both markers (Table 4). Eight genotypes were identified as admixed by SSRs, with over 20% ancestry from other groups. Five of them were also identified by RFLPs. Some of these putative hybrids, 4/7 and 2/5 detected by SSRs and RFLPS, respectively, were confirmed with the reassignment test implemented in GENECLASS2 (Table 4). The five individuals identified as putative

Table 2: Geographic and genetic origin of the cultivated *C. canephora* genotypes.

Type of introduction	Group name	Donor or collector	Country of origin	Coll #	A priori genetic group	Final group
Donation from field genebanks	"Aboisso"	Aboisso ¹ ,	Gabon	C135	A	A
		Côte d'Ivoire		C307	A	A
				C318	E	DE
				C329	E	AE
	"C10 Man"	Unknown	Rep. of Congo	C429	E	E
				C439	E	E
	"INEAC"	INEAC ⁴ , Rep. of Congo	Rep. of Congo	C002	E	E
				C078	E	E
				C032	E	E
				C054	E	E
				C057	E	AE
				C062	E	E
				C095	E	DE
				C003	A	AE
				C006	E	E
				C015	E	E
Collection in plantations	"Guinea"		Guinea ⁶	0855	D	D-
				0856	D	DE
				0850	E	DE
				0851	E	DE
				0852	E	E
				0909	E	DE
				0915	E	DE
				0848	E	DE
				0853	E	D-
				0881	E	DE
				0882	E	DE
	"Hybrid"		Côte d'Ivoire	C107	D	DE
				C126	E	DE
				C181	D	DE
				C636	D	DE
	"Côte d'Ivoire"		Côte d'Ivoire ^{5,6}	0317	D	DE
				0318	D	DE
				0163	A	AE
				0164	A	A
				C155	A	D
	"Togo"		Togo	0693	A	A
				0695	A	A
	"Tanzania"		Tanzania	0270	Unknown	E-
				0279	E	CE
Unkown	"Robusta AI"		Unknown	C077	A	A

Membership to *a priori* genetic groups, as previously defined by Dussert et al. [35], and final group definition as identified in the present study are indicated.

¹ Introduction at Aboisso (Côte d'Ivoire) by Beynis in 1910, of material cultivated in Gabon [57]

² Introduction at the garden at Bingerville (Côte d'Ivoire) in 1914 of material cultivated in Togo [57]

³ Introduction at CRA Bingerville (Côte d'Ivoire) in 1951, of material selected in Madagascar and originating from Gabon [57]

⁴ Introduction in Côte d'Ivoire in 1935 of material selected at INEAC in the Republic of Congo (Cordier [57])

⁵Le Pierrès et al. [26]; ⁶Berthaud [33]

Table 3: Characteristics of the 16 SSRs and 8 RFLP loci from the total sample (N = 107) and the wild sample (N = 61).

SSR		Total sample N = 107				Wild sample N = 61			
Locus	Map location	N_a	H_O	H_E	PIC	N_a	H_O	H_E	PIC
M257	B	6	0.25	0.55	0.47	6	0.21	0.59	0.52
M809	H	5	0.15	0.25	0.23	4	0.11	0.16	0.15
M804	M	3	0.06	0.07	0.07	1	0.00	0.00	0.00
M821	F	13	0.28	0.83	0.81	13	0.28	0.83	0.81
M314	G	5	0.25	0.49	0.43	5	0.18	0.42	0.36
M368	-	20	0.81	0.93	0.92	15	0.76	0.91	0.91
M764	-	13	0.65	0.83	0.81	11	0.63	0.82	0.80
M394	C	9	0.41	0.80	0.77	8	0.39	0.75	0.71
M779	-	12	0.75	0.85	0.83	12	0.63	0.82	0.80
M782	-	4	0.06	0.25	0.22	4	0.03	0.12	0.11
M495	-	9	0.35	0.65	0.60	9	0.33	0.62	0.57
M259	N	6	0.52	0.59	0.55	6	0.48	0.60	0.56
M367	-	15	0.30	0.88	0.87	12	0.20	0.85	0.83
M755	-	13	0.42	0.78	0.76	11	0.23	0.68	0.65
M387	J	10	0.33	0.75	0.71	8	0.26	0.74	0.70
M856	I	11	0.49	0.88	0.87	11	0.41	0.85	0.84
Total		154				136			
Mean		9.63	0.38	0.65	0.62	8.5	0.32	0.61	0.58

RFLP		N = 107				N = 61			
Locus	LG ²	N_a	H_O	H_E	PIC	N_a	H_O	H_E	PIC
gA13	10	14	0.51	0.73	0.70	11	0.43	0.66	0.61
gA19	9	10	0.43	0.78	0.75	9	0.38	0.72	0.68
gA29	5	5	0.27	0.47	0.44	5	0.26	0.50	0.47
gA10	3	8	0.48	0.78	0.76	7	0.42	0.77	0.74
gA14	3	5	0.30	0.59	0.53	5	0.31	0.56	0.52
gA61	7	8	0.50	0.70	0.66	6	0.48	0.69	0.65
cR167	4	7	0.43	0.60	0.57	7	0.34	0.58	0.55
gA72	7	2	0.16	0.41	0.33	2	0.13	0.34	0.28
Total		59				52			
Mean		7.37	0.38	0.63	0.59	6.50	0.34	0.60	0.56

Linkage group locations for mapped markers are given. The number of alleles (N_a), observed heterozygosity (H_O), expected heterozygosity (H_E , or gene diversity) and PIC (polymorphism information content) values are also given.

¹ Linkage group of (*C. canephora* × *C. heterocalyx*) × *C. canephora* genetic map from Coulibaly et al. [60]

² Linkage group of *C. canephora* genetic map from Paillard et al. [59], with estimated distances: gA10-gA14 = 21 cM, gA61-gA72 = 30 cM

hybrids by structure analysis using either marker type originated from the *a priori* D group of Côte d'Ivoire, and contained >20% group E or group A ancestry (Table 4). On the RFLP tree, these individuals nested within the same clade, with an intermediate position on the between groups [A-B-E] and [C-D]. They occurred at the base of the D clade on the SSR tree. The individual AC663, classified *a priori* in group A, and identified as a putative AE hybrid with RFLPs, fell into the group E cluster in the RFLP tree.

The genetic differentiation of wild *C. canephora* groups, evaluated with F_{ST} values, were found to be higher for

RFLPs compared to SSRs, with 0.50 *vs.* 0.33, respectively (Table 5). Pairwise RFLP-based or SSR-based F_{ST} values ranged from 0.29 to 0.67 and from 0.20 to 0.50, respectively, with substantially lower F_{ST} values within regions than between regions. Group D (West Africa) showed the highest levels of differentiation detected with all other groups with both marker types, whereas the A-B and E-B comparisons (groups from Central Africa), gave the lowest F_{ST} values with RFLP and SSR data, respectively. Mantel's tests revealed that pairwise F_{ST} estimates among samples calculated from SSRs were not significantly differ-

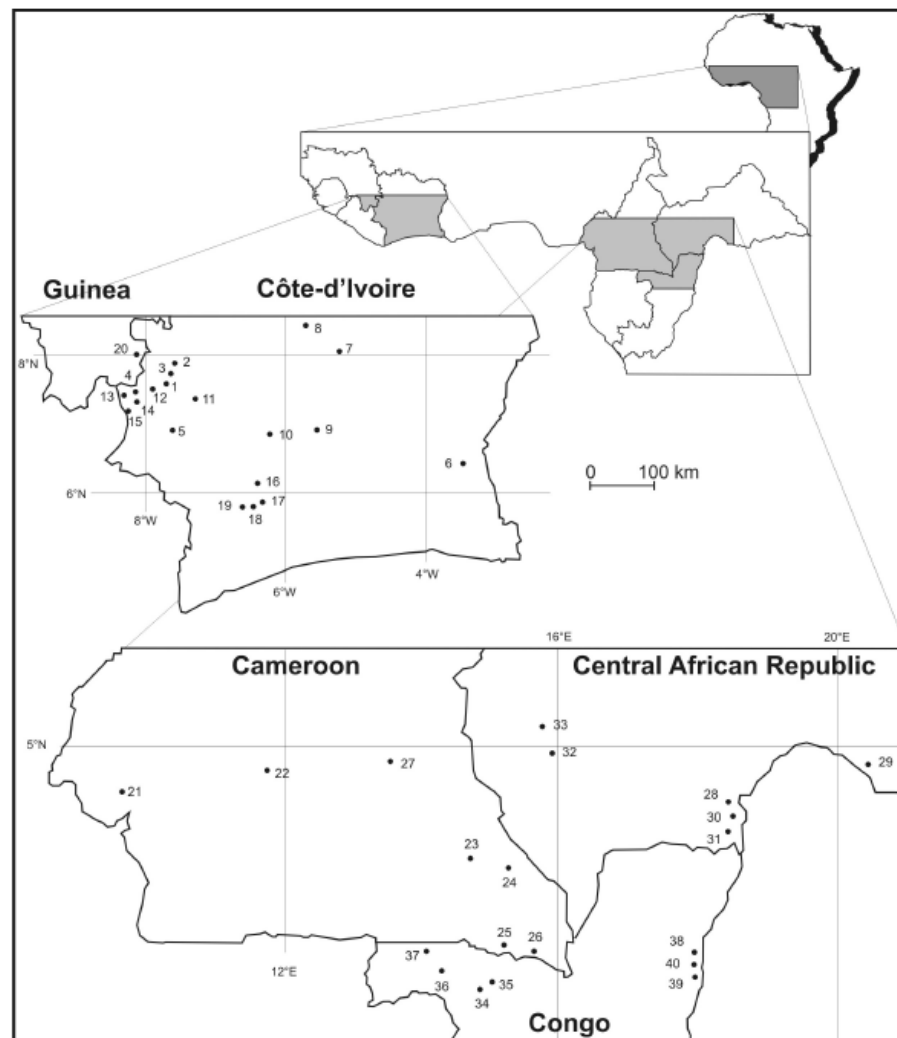


Figure 1
Wild *Coffea canephora* sampling locations in West and Central Africa. Codes are given in Table 1.

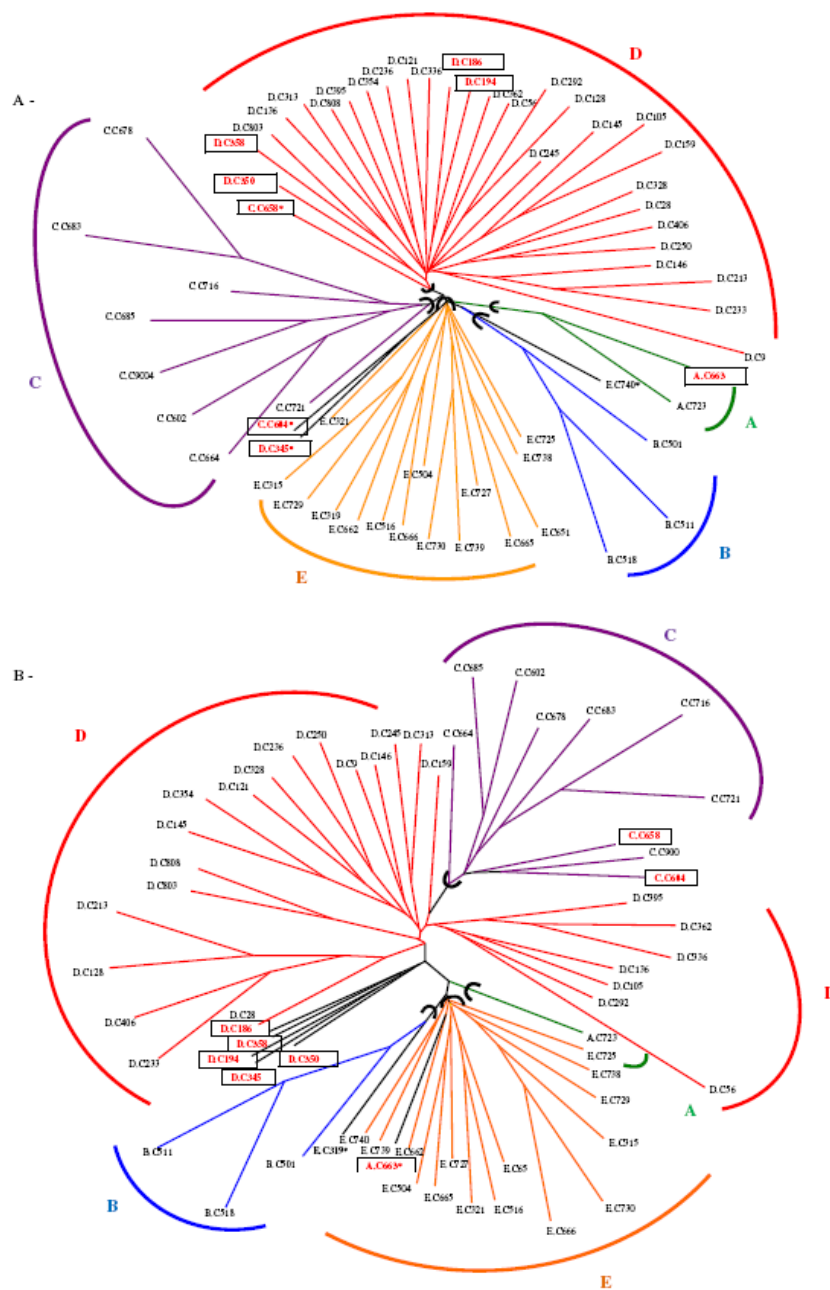
ent from those calculated from RFLPs ($R = 0.72$, $N = 1000$ permutations).

Cultivated sample origin

When the cultivated genotypes were included in the neighbour-joining tree construction, the topology of the trees in five groups could not be identified appropriately with RFLPs and the clusters were even less separated using SSRs (data not shown). Since some of the cultivated plants might be of putative hybrid origin (Table 2, "hybrid" group), we assessed the cultivated genotype origin by population structure analysis.

Assignment tests using GENECLASS were first performed to assess genotypic similarities of the cultivated plants with

respect to the five wild genetic reference groups. Using RFLPs, of the 46 cultivated plants, 10 (22%), 13 (28%) and one (2%) were assigned to groups A, E, and D, respectively, whereas the remaining individuals (48%) were all unassigned. In comparison, SSRs only assigned 8 (17%), 9 (20%) and one (2%) individuals to groups A, E, and D, respectively, whereas 28 (61%) remained unassigned. Considering the overall dataset, assigned cultivated individuals, except for three of them, were correctly classified within their *a priori* group of origin with both RFLP and SSR data. Genetic groups B and C were not represented by cultivated genotypes. None of the *a priori* group D individuals were reassigned to D, suggesting a possible mixed ancestry origin.

**Figure 2**

Unrooted trees of individual wild *Coffea canephora* genotypes using the neighbour-joining method and shared-allele distance among (A) 16 microsatellite loci and (B) 8 RFLP loci. The first letter of the individual labels, A, B, C, D and E, indicates the *a priori* diversity group of Dussert et al. [35]. Individuals who appear to be misclassified in the present tree are marked (*). Circled names represent plants identified as being mis-classified or putative hybrids by the subsequent genetic structure analysis.

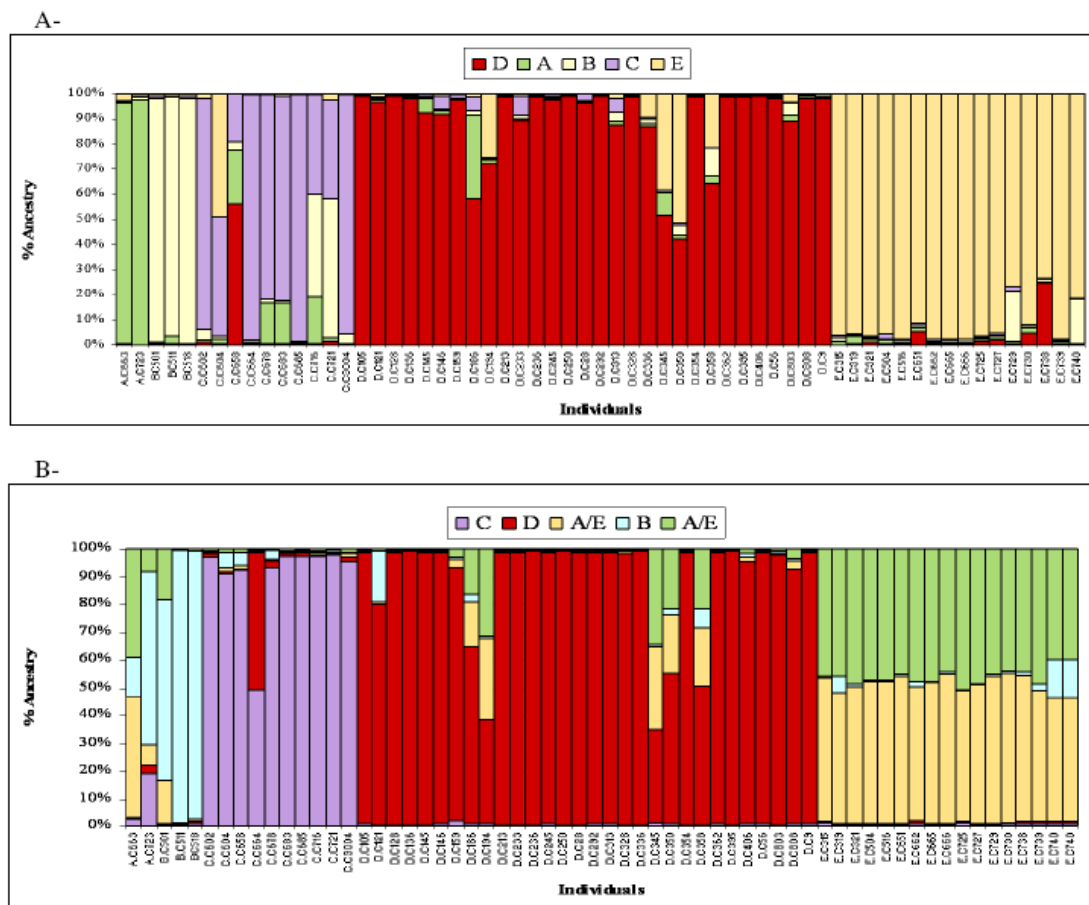


Figure 3
The genetic structure of wild *Coffea canephora* (N = 61) based on (A) 16 polymorphic SSR loci (B) 8 polymorphic RFLP loci. Result of STRUCTURE analysis using K = 5 and assignment to the five *a priori* diversity groups (A to D) from Dussert et al. [35]. The colour bars represent percentages of ancestry from the different groups observed in one individual.

Genetic admixture analysis was conducted with STRUCTURE using the overall dataset with both cultivated and wild plants, while excluding wild individuals from group B. The four groups (K = 4) observed, corresponding to groups A, C, D, and E, were found to be genetically distinct with either RFLPs or SSRs. The mean proportion of population membership of wild genotypes to their own clusters was 72, 84, 90, and 92%, for groups A, C, D and E, respectively, using RFLPs; and 69, 82, 89, and 93% for groups A, C, D and E, respectively, using SSRs. This confirmed that the previously defined genetic groups were still responsible for the observed population structure, even after the addition of cultivated individuals.

Admixture analysis with RFLPs of individual genotypes identified hybrids among individuals from a *a priori* group

A (3/14 = 21%) and a *a priori* group E (8/24 = 33%). All plants from a *a priori* group D were detected as hybrids between groups D and E. In comparison, the SSR analysis detected slightly more hybrids with 4/14 = 29% in a *a priori* group A, 13/24 = 54% in a *a priori* group E, and all plants from a *a priori* group D.

Most of the hybrids 13/17 = 76% and 13/23 = 57%, for RFLPs and SSRs, respectively, were identified as originating from hybridization between a *a priori* groups D (West Africa) and E (Central Africa). These hybrids shared an equivalent average fraction of ancestry from both groups (49%D-45%E with RFLPs and 43%D-48%E with SSRs), suggesting that most of them could be classified as first generation hybrids.

Table 4: Missclassified wild individuals or putative hybrids detected by GeneClass and Structure analyses.

A priori Genetic Group		GeneClass		Structure			
Individuals		RFLP	SSR	RFLP Ancestry (%)	CI	SSR Ancestry (%)	CI
AC663	A	E 1.445		E 43.5 A 39 B 14.2	0–100 0–100 0–61		
CC604	C	-	E 3.224		-	E 49.1 C 47.4	23.3–76.2 18.4–73.5
CC658	C	-	D 5.010		-	D 56.3 A 21.2	24.8–83.6 0–71.9
DC186	D	-	-	D 64 E 16 A 16	36.1–86.5 0–52.8 0–54.2	D 57.9 A 33.5	29.6–84 0–68.3
DC194	D	-	-	D 37.5 E 29.6 A 31.6	0–66.1 0–85.8 0–94.2	D 72.6 E 25.3	48.3–94.4 0–50.3
DC345	D	A 0.579	E 0.526	D 33.7 E 30 A 34.3	0–65.2 0–97.3 0–997	D 51.5 E 38.7	26.7–74.7 12.9–65.2
DC350	D	-	E 1.553	D 54.5 E 20.7 A 21.7	0–87.1 0–75.5 0–91.7	D 42.4 E 51.6	19.8–65.5 14.2–76.9
DC358	D	-	-	D 49.9 E 21 A 21.5	12.1–77.9 0–66.6 0–68.5	D 64.3 E 21.2	42.6–83.7 0–51.2

The alternative group and log10 L_{home}/L_{max} from the GeneClass analysis is given. The Structure results are given for the percentage of admixture (mean ancestry) of each of the five diversity groups and the 90% CI. Cells are kept empty for individual non-detected hybrids, i.e. over 80% membership to their own cluster.

The combined data analysis of cultivated coffee trees using both marker types and both population structure analyses allowed us to identify the putative hybrid origin of many cultivated plants (25/46 = 54%) (Table 2).

Principal coordinate plots representing the genetic similarity between wild and cultivated *C. canephora* for SSRs are presented in Figure 4 on the basis of SSR data. The variance explained by the first two axes was greater for RFLPs than for SSRs (45.2% vs. 16.7%). However, on both plots, wild individuals were genetically distinct, with no overlap of the five genetic groups. The distribution of cultivated individuals overlapped that of wild groups, with putative hybrids located at intermediate positions.

Table 5: Pairwise F_{ST} estimates of genetic divergence in wild *C. canephora* groups (N = 53) obtained with SSRs (below the diagonal) and RFLPs (above the diagonal)

SSR\RFLP	A	B	C	D	E
A	0	0.29	0.50	0.67	0.30
B	0.33	0	0.51	0.67	0.30
C	0.30	0.24	0	0.54	0.41
D	0.50	0.45	0.32	0	0.52
E	0.31	0.20	0.25	0.34	0

Core set selection

We selected core sets of both wild and cultivated accessions from *C. canephora* that capture the maximum number of SSR or RFLP alleles for sample sizes 12 and 25. Core sets of 12 captured 48/59 and 103/154 of the RFLP and SSR alleles, respectively. Gene diversity in each core set was equivalent to that found in the entire sample (Table 6). Each core set contained wild and cultivated accessions from the different groups, and a substantial fraction of putative hybrids.

Discussion

Our analysis of *C. canephora* genetic diversity through RFLP and SSR markers revealed a marked geographic structure in subgroups, which could be interpreted with a broad scope of regional and historical influences. The detection of inter-diversity group putative hybrids allowed us to evaluate the influence of humankind on the original natural distribution and the occurrence of gene flow between wild and cultivated stocks.

The choice of the appropriate marker for any specific study depends principally on the purpose of the research and the biology and genetic structure of the species. The recent development of assignment tests based on Bayesian approaches [45–47] has greatly increased the potential for understanding population structure across its diversity

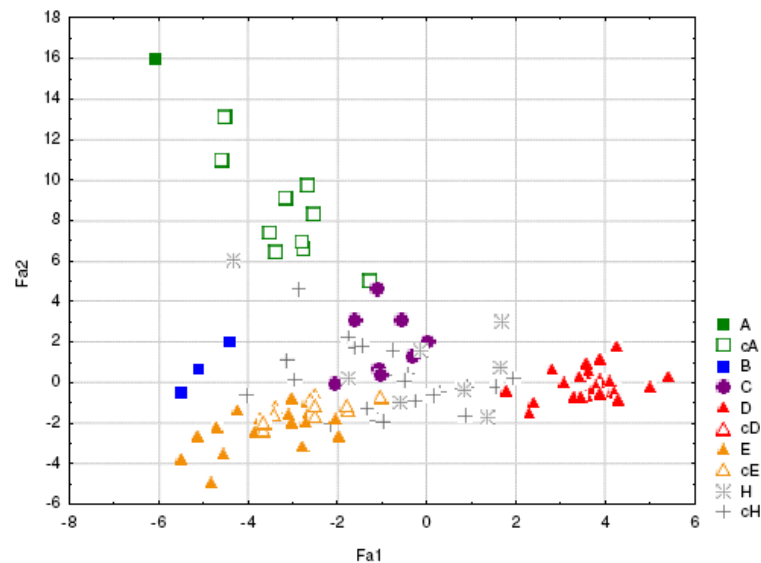


Figure 4

Principal component analysis of wild and cultivated *C. canephora* accessions based on their SSR polymorphism. The first axis represents 10% of the variation and the second one represents 6.5% of the variation. The wild *C. canephora* accessions are represented by colored symbols (A, B, C, D, E). The cultivated accessions assigned to wild genetic groups are represented by empty corresponding symbols (cA, cD, cE). The individuals identified as putative hybrids are represented by a cross mark, H for the wild accession and cH for the cultivated ones.

range. In parallel to our SSR study, we thus re-analysed the RFLP data from Dussert et al. [35] on the same accessions to better compare and assess the associated diversity.

RFLP – SSR comparison

Direct marker system comparison

The genetic data parameters obtained in this study from SSRs were first compared with those obtained with RFLPs.

One of the greatest differences between the two categories of markers concerns their respective polymorphism levels. Estimates of genetic variability H_E and PIC-values, were higher for SSRs than those calculated from RFLPs. This difference in allele variation reflects differences in mutation level: replication slippage is indeed thought to occur more frequently than single nucleotide mutations and insertion/deletion events. RFLP loci are thus characterized by lower mutation rates as compared to SSRs, whose mutation rates can range from 10^{-3} to 10^{-4} depending on their genomic position [48]. Levels of polymorphism detected with RFLP and SSR markers have been compared in soybean [40], also indicating the hypervariability of SSR loci and greater expected heterozygosity.

The higher mutation rates of SSRs also had an effect on genetic distance values. The dissimilarity values taken over

all pairwise comparisons were on average higher for SSRs (0.63) than for RFLPs (0.60). However, the comparison of both dissimilarity matrices revealed that estimates based on RFLPs and SSRs were highly correlated ($r = 0.66$, $p < 0.0001$), indicating congruence between the assays. As a consequence of the mutation rates, the PCO gave a higher percentage of variation explained by the two first axes with RFLPs than for SSRs (45.2% vs. 16.7%), indicating a better separation of groups with fewer multilocus genotypes. However, both plots gave similar resolution in the distinction between individuals of the different groups. Our two NJ trees mainly gave congruent results.

Discriminatory power and assignment capacity

The overall information generated by both markers to facilitate the determination of phylogenetic relationships and classifications, cluster analysis and population structure analysis in the *C. canephora* gene pool was further assessed. Indeed, the differences in mutation rates would also likely affect the power of the different markers to detect population differentiation.

In our data, overall and pairwise F_{ST} estimates obtained with RFLPs were higher than those obtained with SSRs. However, in spite of the differences, SSR and RFLP multi-locus F_{ST} estimates were not significantly different ($R =$

Table 6: Core set of 12 or 25 *C. canephora* genotypes maximising RFLP and SSR diversity.

RFLP	Total:	$N_a = 59$ GD107 = 0.63	SSR	Total:	$N_a = 154$ GD107 = 0.65
Core set of 12	Statut	Core set of 25	Core set of 12	Statut	Core set of 25
_C270	(c) H	_C270	BC511	(+) B	BC511
DCC181	(c) H	DCC181	ECC002	(c) E	ECC002
ECC057	(c) H	ECC057	ECC429	(c) E	ECC429
ECC439	(c) E	ECC439	EC279	(c) H	EC279
CC9004	(+) C	CC9004	ACC135	(c) A	ACC135
EC853	(c) H	EC853	BC501	(+) B	BC501
CC721	(+) C	CC721	_C270	(c) H	_C270
AC164	(c) A	AC164	CC664	(+) C	CC664
EC315	(+) E	EC315	DC9	(+) D	DC9
BC501	(+) B	BC501	ECC095	(c) H	ECC095
EC319	(+) E	EC319	EC851	(c) H	EC851
BC518	(+) B	BC518	DC146	(+) D	DC146
	(c) H	EC848		(+) C	CC716
	(+) E	EC666		(c) A	ACC008
	(c) H	EC915		(+) H	CC658
	(+) H	CC658		(+) H	DC186
	(c) H	EC881		(+) E	EC739
	(+) E	EC651		(c) H	ECC057
	(+) D	DC803		(c) H	EC882
	(+) C	CC716		(+) D	DC136
	(c) H	EC850		(c) A	ACK07
	(c) H	EC882		(c) H	ACK29
	(c) E	ECC429		(+) H	CC604
	(+) E	EC740		(c) A	AC164
	(c) H	EC279		(+) H	AC663
$N_a = 48$ 36.9% GD = 0.68		$N_a = 58$ 59.6% GD = 0.72	$N_a = 103$ 19.6% GD = 0.66		$N_a = 137$ 38.3% GD = 0.69

Corresponding gene diversity (GD) values, cultivated status: wild (+) or cultivated (c), genetic group (A to E) or putative hybrid (H), No. of alleles (N_a) and percentage of total variability revealed in the whole sample (107 individuals) are given.

0.72) when computed over the entire set of samples. In fact, the high level of polymorphism, typical of SSRs, may induce downward bias in the population differentiation estimates. The degree of differentiation assessed through SSRs, and thus the F_{ST} values, are expected to be lower than those calculated using RFLPs [49]. A marked population subdivision similar to that noted in our study was also found at RFLP loci in brown trout [50] and soybean [40]. Moreover, the consistency of our results obtained from both markers is congruent with the findings of the other study comparing SSR and RFLP variation [40].

Finally, the RFLP and SSR multilocus diversity structure analysis divided wild coffee trees into largely concurrent five groups, with main branches on the individual neighbour-joining tree. Comparable diversity structuring was obtained, e.g. in terms of distinguishing geographical origins from West Africa and Central Africa.

At a smaller geographical scale, RFLPs and SSRs showed a different resolution power in detecting the genetic struc-

ture in the wild samples. The Bayesian analysis using STRUCTURE revealed five clearly distinct groups with SSR data, while samples from the southwest Cameroon/northwest Congo region—*a priori* groups A and E—were clustered with the RFLP data. A higher membership in their own cluster was obtained with SSRs (89% vs. 72% mean ancestry over all wild groups with SSRs and RFLPs, respectively). Moreover, the SSR data allowed the identification of slightly more admixed plants, putatively hybrids, both within the wild and cultivated pools. The GENECLASS assignment test, which has been shown to be effective even if clusters are not in HWE, gave congruent results.

In conclusion, SSR markers outperformed RFLP markers in terms of discriminatory power in cluster analyses and assignment tests. Moreover, SSRs were also more efficient in discriminating the source of an individual genotype among putative diversity groups on a local scale. This advantage of SSR markers when analysing genetic affinities at individual levels was also demonstrated in a study

comparing SSR and allozyme markers in brown trout [50].

C. canephora genetic diversity and history

Organisation of C. canephora natural diversity

The genetic structure analysis carried out on the African samples of wild *C. canephora* using RFLPs and SSRs revealed marked separation between the West and Central African samples corresponding, in their composition, to the Guinean and Congolese groups of Berthaud [33]. This marked separation might be related both to the large geographical distances and to historical events. Indeed, the last glaciations and the subsequent migration from various Pleistocene refugia had caused large-scale changes in vegetation patterns, most notably around the Dahomey Gap and Cross River [2]. The western forests from Guinea and Côte d'Ivoire are separated from the Central forests of Cameroon, Central African Republic and Congo by the current 300 km wide Dahomey Gap, which is known to be an important biogeographic barrier. This area is thought to have become an extremely arid and much wider area during the last glacial maxima, around 18,000 BP, separating the forest refuges of southwest Ghana and west Cameroon [2]. Divergence through this geographical isolation might have led to genetic differentiation of the *C. canephora* populations. When analysing *C. liberica*, another *Coffea* species with the same geographical distribution, N'Diaye et al. [32] reported that the two varieties *C. liberica* var. *liberica* Bull. ex Hiern and *C. liberica* var. *dewevrei* (De Wild. & T. Durand) Lebrun, had high genetic differentiation and were characterised by marked reproductive barriers between the two varieties, with a pollen viability of their F_1 hybrids similar to that of inter-specific hybrids. However, although highly differentiated, the West (group D) and Central African (A, B, C, and E) groups of *C. canephora* present fully interfertile individuals (see below). This suggested that the two main *C. canephora* diversity groups do not present a state in the speciation process as advanced as that of the two *C. liberica* varieties. Phylogenetic studies have indeed revealed the role of the Dahomey Gap on lineage origins in the Upper Guinea regions [27,51]. But a similar distinct evolutionary split between populations in Upper Guinea and those of Lower Guinea across the Dahomey Gap has also been observed in shea tree species [19] and in Fire-crested Alethe birds [52].

The refugia scenario concerning the African Guineo-Congolian rainforest is supported by the fact that a similar speciation pattern has been observed in various genera [reviewed in [4]], particularly in African Rubiaceae genera [13]. Isolation of populations in refugia was also suggested to influence divergent adaptations and diversification at the species level [e.g. [14,18]]. Within the Central African zone, the distribution of the four *C. canephora*

groups (A, B, C, and E) showed a complex pattern. This regional pattern of differentiation could be interpreted as evidence of cycles of fragmentation and subsequent expansion of forest habitats. The Congolese region consisted of a mosaic of several Pleistocene refugia [2] that might explain the pattern of genetic diversity in the Central African zone. Although it is difficult to precisely localise the refugia, the reconstructed hypothesized areas [2] were found to be related to species diversification. For example, Anthony et al. [14] suggested a role of these Pleistocene refugia in structuring gorilla genetic diversity. For *C. canephora*, the refugia origin of sub-group E is most likely located in the Congo-Zaire basin (Figure 5). Meanwhile, the *C. canephora* group C might derive from the expansion of the Biafran forest refugia, inside curve of the Gulf of Guinea, from Cross River to Sanaga River (west Cameroon *sensu lato*). The late Holocene phase of dramatic climatic disturbance could also have been favourable for *C. canephora* diversification in Central Africa. Culminating about 2,500 years ago, it led to a catastrophic reduction in central African rainforests, in the region of south Cameroon, south Central African Republic, Gabon and Congo, and still exerts a major influence on the present forest vegetation distribution [53]. This event involved a brutal extension of savannas, favourable for the expansion of pioneering species such as oil palm [54]. The return of wetter conditions favourable for forest re-invasion began around 2,000 years B.P. from residual forests. This climatic disturbance caused fragmentation of the ancient Okoume (*Aucoumea klaineana*) distribution area into two subgroups, as suggested by the two tree varieties observed at the molecular level [21]. This process could also have produced the spatial genetic structuring of *C. canephora* in the Congolese zone. Both the intensity and length of these recent range expansion episodes could explain why coffee trees did not accumulate enough differentiation to lead to reproductive barriers and speciation.

Impact of human cultivation on inter-group gene flow and hybridization

The genetic structure analyses conducted in the present study seem to be efficient for reliable detection of admixtures and individual identification. Indeed, deviations from *a priori* diversity group classifications, are in line with issues concerning the introduction of non-indigenous wild or cultivated *C. canephora* plants or with the putative hybrid origin of cultivars.

Assignment tests confirmed that the landraces have remained genetically very close to the original wild populations. Indeed, since the 19th century, many local cultivars were cultivated spontaneous forms of *C. canephora* directly collected in adjacent forest populations. This was the case in the Democratic Republic of the Congo, in Côte

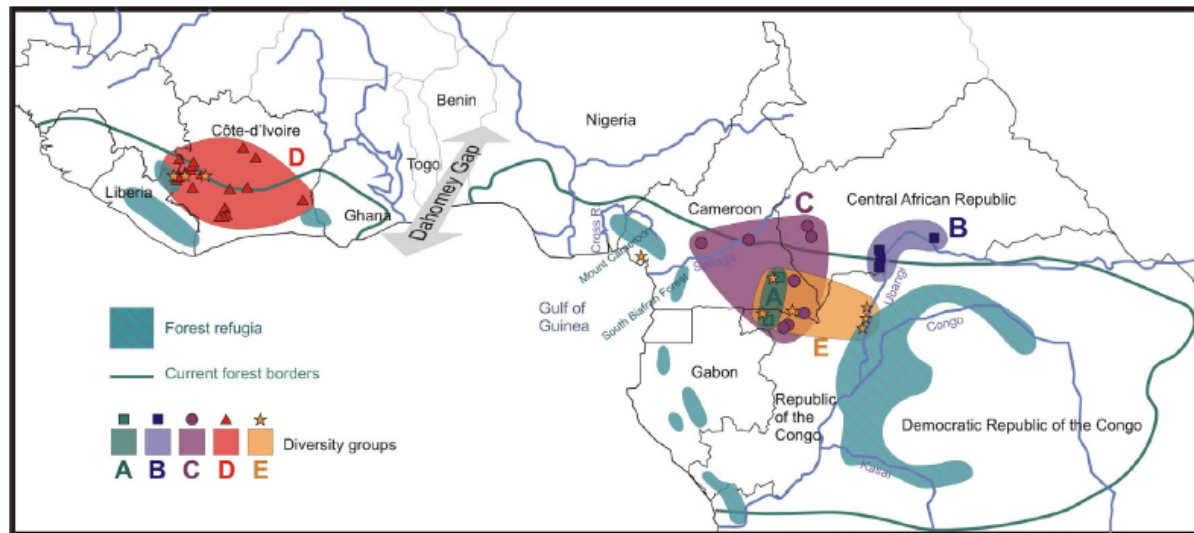


Figure 5
Geographic position of the five major *C. canephora* genetic subgroups. Geographic position of the five major *C. canephora* genetic subgroups collected together with a schematic map of forest refugia over the course of the last major arid phase (c. 18,000 years BP), adapted from Maley [2].

d'Ivoire (for "Kouilou" genotypes), Togo and Benin (for "Niaouli" genotypes), and Central African Republic (for "Nana" genotypes). Even improved material from INEAC (Table 2) has undergone few breeding cycles. Moreover, contact between local spontaneous forms and adjacent plantations could have promoted intermixing of individuals both in the wild and cultivated germplasm. In fact, according to the domestication syndrome definition (Hammer 1984), Robusta coffee trees should be considered as "cultivated" or "semi-domesticated" rather than "domesticated", since no heritable distinctive traits could be discerned when compared to their wild relatives.

The putative hybrid origin of some cultivars is suggested by the present data. In particular, most of the *a priori* group D (from Côte d'Ivoire and Guinea) and some of *a priori* group E (from southeast Cameroon and south Central African Republic) cultivated forms turned out to be putative F_1 hybrids between these two genetic groups. Humans have contributed to this phenomenon since cultivated material originating from Central Africa, in particular from the Congo region, has been introduced in West African countries, such as in Côte d'Ivoire since 1930 (Portères, 1937). The low level of genetic introgression we observed in the putative hybrids suggests a recent contact between the local and exogenous material. Hybrids between the Guinean and Congolese group material had been previously identified by Berthaud [33] in cultivated variants on the basis of their vigour and yield. The use of

intergroup hybrids and the efficiency of reciprocal recurrent selection was subsequently demonstrated [55]. This contrasts markedly with the cultivation history of *C. arabica*, whose varieties originated from a narrow genetic base [56]. Polymorphism was further reduced during selection cycles and by the predominant autogamy of *C. arabica*.

In this study, the putative intergroup hybrids, corresponding to various introgression levels detected and supposedly of wild origin, probably resulted from either culture escapes or from cross-pollinisation with neighbouring plantations.

Consequently, the core sets we propose contain both wild and cultivated accessions from the different groups, and a non-negligible fraction of putative hybrids. They provide good reference sets for further identification of spontaneous or cultivated material.

Conclusion

The high degree of concordance between the SSR and RFLP data for diversity group differentiation indicates that our set of markers provided adequate genome coverage for accurate germplasm characterization. Even though these loci have likely evolved at different rates, they revealed comparable diversity structure with five well-differentiated groups corresponding to geographical patterning in the individuals. The molecular variation was

correlated with the natural distribution in two postulated refugia areas, *i.e.* the Upper Guinea and Lower Guinea/Congo regions, separated by the Dahomey Gap.

A more intense and larger scale sampling would be required for more detailed geographical mapping of the diversity and more genetically precise refugia identification, especially in the Central African mosaic zone.

SSR genotyping provided highly informative data for multi-locus discrimination of individuals and putative hybrid detection. These data could be related to the recent history of coffee agricultural activities. Inter-group hybridizations were detected at a non-negligible frequency, especially between *a priori* group D ("Guinean" in West Africa) and E ("Congolese" in Central Africa). The previous observation of heterosis of Guinean-Congolese cultivars suggests that the high diversity included in these diversity groups represents an excellent genetic reservoir that could be tapped for adapted stock breeding schemes. This study also provides an excellent basis for determining the appropriate scale of wild population conservation and management.

Methods

Study species and sampled genotypes

Coffea canephora is a widespread species throughout west-central Africa [33]. More than 700 wild genotypes were collected by ORSTOM (now IRD, Institut de Recherche pour le Développement, France) in collaboration with CIRAD, FAO (Food and Agriculture Organization), IPGRI (International Plant Genetic Resources Institute), and MNHN (Muséum National d'Histoire Naturelle), between 1975 and 1987, in five African countries: Côte d'Ivoire and Guinea in West Africa; and Cameroon, Congo and Central African Republic in Central Africa [23-26,33]. These genotypes are conserved in the only reference collection for wild forms of *C. canephora*, *i.e.* the Divo collection, Côte d'Ivoire. In parallel, CIRAD assembled a collection of cultivated material, also conserved at the Divo experimental station. This collection contains more than 600 accessions of diverse origins: local varieties and populations, forms taken from village plantations, and selected material [26,33,35,57].

In this study, a total of 107 *Coffea canephora* genotypes (61 wild and 46 cultivated) were selected from the initial set analysed in Dussert et al. [35]. The wild genotypes were sampled in order to have a representation of each of the 40 forest populations studied throughout the range of the species (Figure 1, Table 1). For the cultivated material, a random proportional sampling was done for each of the 10 principal origins identified in collection (Table 2).

Total genomic DNA was extracted according to the method described by Ky et al. [58].

RFLP analysis

Two restriction enzymes were used to digest the genomic DNA: *EcoRI* and *HindIII*. Out of the 26 initially tested probes, eight were retained for their polymorphic and mono-locus characteristics. Selected RFLP probes corresponded to previously mapped loci distributed on six linkage groups of the *C. canephora* linkage map (Table 3) [59].

SSR analysis

Sixteen SSR loci (Table 3) were amplified as previously described [38,42] using a touchdown PCR profile optimised for each set of primers: touchdown 60°C to 55°C or touchdown 55°C to 50°C. PCR products were detected on an IR2 Automated DNA Sequencer (LI-COR, model 4200L-2, Lincoln, NE, USA) using an M13 primer coupled to the infrared tag IRD700 or IRD800 after migration on 25 cm 6.5% KBplus (LI-COR, CAT# 827-05607) polyacrylamide gels. The gel images were processed by SAGA GT™ software (LICOR Biotech) to estimate the size of amplicons according to a 50–350 bp size standard (LI-COR, CAT# 829-05343, 829-05344).

The sixteen SSR loci were evenly distributed throughout the *Coffea* genome and nine of them mapped on nine different linkage groups of the interspecific map [*C. heterocalyx* × *C. canephora*] × *C. canephora* (Table 3) [60].

Data analysis

Genetic diversity

For each SSR and RFLP locus, we assessed genetic polymorphism within total or wild samples by calculating the observed number of alleles (N_a), observed and expected heterozygosity (H_o and H_e), and the polymorphism information content (PIC) using the POWERMARKER v3.25 software [61]. In these analyses, individual wild plants of possible hybrid origin, as determined by genetic structure analysis (see below), were excluded from the wild pool.

Genetic affinities among individual genotypes, Cluster analyses

Neighbour-joining trees were constructed using the shared-allele distances for both the RFLP and SSR data using POWERMARKER v3.25 software [61]. Bootstrapping was conducted with 2000 replicates and the trees were implemented in the PHYLIP package [62] to obtain a consensus tree, visualized in TREEVIEW (taxonomy.zoology.gla.ac.uk/rod/treeview.html). The Mantel matrix correspondence test was used to compare individual genetic distances generated by each marker type using POWERMARKER v3.25.

In order to display patterns in the individual genetic distances, a principal coordinate analysis (PCO) was performed on both the RFLP and SSR dataset. For each individual, we calculated the frequencies of each allele (0, 0.5, and 1) at each locus, and used this data to perform a

principal component analysis (PCA), "French PCA" in STATISTICA V6.1 <http://www.statsoft.com>. This analysis was computed for the whole sample, with cultivated and putative hybrids, as determined by genetic structure analysis (see below), included as additional passive elements.

Genetic structure

For statistical investigation of genetic structure of the *C. canephora* wild sample and detection of intermediate types (hybrids between diversity groups), two different Bayesian analyses were performed.

The genetic structure was first investigated for both RFLP and SSR datasets with the Bayesian approach in STRUCTURE v2.1 [45,47]. Parameters were set at $K = 5$ for the number of groups, 30,000 for the burn-in time and 1,000,000 for the number of runs, with five repetitions. The number of clusters ($K = 5$) was confirmed as the value that maximized the increase in the posterior probability of the $\ln P(D)$ data according to the formula $[\ln P(D)_k - \ln P(D)_{k-1}]$, as suggested by Garnier et al. [63]. We assessed the proportion of q_i membership of each genotype to the five genetic groups, i.e. the proportion of its genome drawn from each group. We assigned each individual genome to one group when the average proportion of membership was $q_i > 0.80$, i.e. over 80% ancestry to their own cluster. In the case of admixed individuals, we jointly assigned them to two or more groups if the proportion of membership to each one was $0.20 < q_i < 0.80$. For each individual, we calculated a 90% CI of the q_i parameter.

We also used the "leave one out" procedure and the method proposed by Cornuet et al. [64]. The Bayesian-based maximum likelihood test implemented in GENECLASS2 V2.0 [46] has been shown to be effective in genotype assignment, even when populations deviate from Hardy-Weinberg equilibrium [65,66]. Each individual to be reassigned was removed from its source group and the frequency estimates of each locus were modified accordingly (Monte Carlo simulations of 1,000 independent individuals for each candidate group). Differences in log-likelihood values were computed to assign individuals to one group with a risk of 0.01. The reference groups used were the *a priori* A, B, C, D, and E groups. This software was also used to assign cultivated plants to the wild genetic groups.

Genetic differentiation

For the analysis of genetic differentiation within the wild sample, individual plants of possible hybrid origin were excluded. The partition of the genetic variation between wild genetic groups for both RFLP and SSR data was estimated with the F_{ST} of Weir and Cockerham implemented in GENETIX [67]. Significance levels of pairwise F_{ST} values were calculated using permutation tests ($N = 1000$). The

correlation between the two F_{ST} matrixes generated by each marker type was investigated by Mantel's test of matrix correspondence in GENETIX.

Core set

To assist in the use or conservation of wild and cultivated *Coffea canephora* germplasm, we defined core sets of accessions that capture the maximum RFLP or SSR diversity using the principal component score strategy (PCSS) [68]. Based on K_{hi-2} distances, a factorial analysis is applied to transform initial data into factor scores. Iterative selection of individuals maximising subset variability is based on their relative contribution to the generalised sum of squares (GSS), expressed in percentage.

Authors' contributions

CG carried out the PCR amplification experiments and the genotyping, participated in the analyses and drafted the manuscript. SD obtained all the specimens, carried out the RFLP experiments and helped in analyzing the diversity data, and also drew up some of the figures. PH, SH and AK helped in analyzing the data. AdK and SH coordinated the project. VP served as the principal investigator of the project, participated in its design and coordination, helped in the data analysis, assisted in the drafting of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

CG received French government funding for her PhD thesis. The authors thank Jérôme Minier for technical support on the SSR experiments and Jean Maley for passionate discussions. We are grateful to Madalena Simas Branco (CIBIO/UP) for her critical review of the manuscript.

References

- White F: The guineo-congolian region and its relationships to the other phytocoria. *Bull Jard Bot Natl Belg* 1979, **49**:11-55.
- Maley J: The African rain forest: main characteristics of changes in vegetation and climate from the upper Cretaceous to the Quaternary. *Proc R Soc Edinburgh* 1996, **104B**:31-73.
- Linder HP: Plant diversity and endemism in sub-Saharan tropical Africa. *Journal of Biogeography* 2001, **28**(2):169-182.
- Plana V: Mechanisms and tempo of evolution in the African Guineo-Congolian rainforest. *Philos Trans R Soc Lond B Biol Sci* 2004, **359**(1450):1585-1594.
- Mittermeier RA, Robles Gil P, Hoffman M, Pil-Grím J, Brooks T, Mittermeier CG, Lamoreux J, Fonseca GAB: Hotspots revisited: Earth's biologically richest and most threatened terrestrial ecoregions. *Mexico City, Mexico* 2004.
- Stebbins GL: *Flowering Plants: Evolution Above the Species Level*. Cambridge, MA 1974.
- Fjeldsa J, Lovett JC: Geographical patterns of old and young species in African forest biota: The significance of specific montane areas as evolutionary centres. *Biodiversity and Conservation* 1997, **6**(3):325-346.
- Evans BJ, Kelley DB, Tinsley RC, Melnick DJ, Cannatella DC: A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. *Mol Phylogenet Evol* 2004, **33**(1):197-213.
- Haffer J: Speciation in Amazonian forest birds. *Science* 1969, **165**(3889):131-137.
- Richardson JE, Pennington RT, Pennington TD, Hollingsworth PM: Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science* 2001, **293**(5538):2242-2245.

11. Mayr E, Ohara RJ: **The Biogeographic Evidence Supporting the Pleistocene Forest Refuge Hypothesis.** *Evolution* 1986, **40**(1):55-67.
12. Salomon M: **Evolutionary biogeography and speciation: essay on a synthesis.** *Journal of Biogeography* 2001, **28**:13-27.
13. Robbrecht E: **Generic Distribution Patterns in Sub-Saharan African Rubiaceae (Angiospermae).** *Journal of Biogeography* 1996, **23**:311-328.
14. Anthony NM, Johnson-Bawe M, Jeffery K, Clifford SL, Abernethy KA, Tutin CE, Lahm SA, White LJ, Udey JF, Wickings EJ, et al.: **The role of Pleistocene refugia and rivers in shaping gorilla genetic diversity in central Africa.** *Proc Natl Acad Sci USA* 2007, **104**(51):20432-20436.
15. Hewitt GM: **Genetic consequences of climatic oscillations in the Quaternary.** *Philos Trans R Soc Lond B Biol Sci* 2004, **359**(1442):183-195.
16. Cottrell JE, Krystufek V, Tabbener HE, Milner AD, Connolly T, Sing L, Fluch S, Burg K, Lefevre F, Achard P, et al.: **Postglacial migration of *Populus nigra* L.: lessons learnt from chloroplast DNA.** *Forest Ecology and Management* 2005, **206**(1-3):71-90.
17. Petit RJ, Aguinalde I, de Beaulieu JL, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, et al.: **Glacial refugia: hotspots but not melting pots of genetic diversity.** *Science* 2003, **300**(5625):1557-1565.
18. Lorenzen ED, De Neergaard R, Arcander P, Siegmund HR: **Phylogeography, hybridization and Pleistocene refugia of the kob antelope (*Kobus kob*).** *Mol Ecol* 2007, **16**(15):3241-3252.
19. Fontaine C, Lovett PN, Sanou H, Maley J, Bouvet JM: **Genetic diversity of the shea tree (*Vitellaria paradoxa* C.F. Gaertn), detected by RAPD and chloroplast microsatellite markers.** *Heredity* 2004, **93**(6):639-648.
20. Lowe AJ, Gillies AC, Wilson J, Dawson IK: **Conservation genetics of bush mango from central/west Africa: implications from random amplified polymorphic DNA analysis.** *Mol Ecol* 2000, **9**(7):831-841.
21. Muloko-Notoutoume N, Petit RJ, White L, Abernethy K: **Chloroplast DNA variation in a rainforest tree (*Aucoumea klaineana*, Burseraceae) in Gabon.** *Mol Ecol* 2000, **9**(3):359-363.
22. Brouat C, McKey D, Douzery EJ: **Differentiation in a geographical mosaic of plants coevolving with ants: phylogeny of the *Leonardoxa africana* complex (Fabaceae: Caesalpinioideae) using amplified fragment length polymorphism markers.** *Mol Ecol* 2004, **13**(5):1157-1171.
23. Berthaud J, Guillaumet JL: **Les caféiers sauvages en Centrafrique. Résultats d'une mission de prospection (Janvier-février 1975).** *Café Cacao Thé* 1978, **22**:171-186.
24. Anthony F, Couturon E, Namur de C: **Les caféiers sauvages du Cameroun. Résultats d'une mission de prospection effectuée par l'ORSTOM en 1983.** *11e Colloque ASIC, Lomé, Togo* 1985:495-501.
25. Namur de C, Couturon E, Sita P, Anthony F: **Résultats d'une mission de prospection des caféiers sauvages du Congo.** *12eme Colloque ASIC, Montreux, France* 1988:397-404.
26. Le Pierrès D, Charmentant P, Yapo A, Leroy T, Couturon E, Bontems S, Tehe H: **Les caféiers sauvages de Côte-d'Ivoire et de Guinée: bilan des missions de prospection effectuées de 1984 à 1987.** *13eme Colloque ASIC, Paipa, Colombie* 1989:420-428.
27. Maurin O, Davis AP, Chester M, Mvungi EF, Jaufeerally-Fakim Y, Fay MF: **Towards a Phylogeny for *Coffea* (Rubiaceae): identifying well-supported lineages based on nuclear and plastid DNA sequences.** *Ann Bot (Lond)* 2007, **100**(7):1565-1583.
28. Davis AP, Govaerts R, Bridson DM, Stoffelen P: **An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae).** *Botanical Journal of the Linnean Society* 2006, **152**(4):465-512.
29. Silvestrini M, Junqueira MG, Favarin AC, Guerreiro O, Maluf MP, Silvarolla MB, Colombo CA: **Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellites markers.** *Genet Resour Crop Evolution* 2007, **54**(6):1367-1379.
30. Lashermes P, Combes MC, Robert J, Trouslot P, D'Hont A, Anthony F, Charrier A: **Molecular characterisation and origin of the *Coffea arabica* L. genome.** *Mol Gen Genet* 1999, **261**(2):259-266.
31. Anthony F, Bertrand B, Quiros O, Wilches A, Lashermes P, Berthaud J, Charrier A: **Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers.** *Euphytica* 2001, **118**(1):53-65.
32. N' Diaye A, Poncet V, Louarn J, Hamon S, Noiro M: **Genetic differentiation between *Coffea liberica* var. *liberica* and *C. liberica* var. *dewevrei* and comparison with *C. canephora*.** *Pl Syst Evol* 2005, **253**(1-4):95-104.
33. Berthaud J: **Les ressources génétiques pour l'amélioration des caféiers africains diploïdes.** Collection Travaux et documents edn. Montpellier, France: Orstom; 1986.
34. Montagnon C, Leroy T, Yapo A: **Genotypic and phenotypic diversity of some coffee groups (*Coffea canephora* Pierre) in the collections – consequences on their use in breeding.** *Café Cacao Thé* 1992, **36**(3):187-198.
35. Dussert S, Lashermes P, Anthony F, Montagnon C, Trouslot P, Combes MC, Berthaud J, Noiro M, Hamon S: ***Coffee, Coffea canephora*.** In *Diversité génétique des plantes tropicales cultivées* Edited by: Hamon P, Seguin M, Perrier X, Glaszmann J-C. Montpellier, France: CIRAD; 1999:175-794.
36. Moncada P, McCouch S: **Simple sequence repeat diversity in diploid and tetraploid *Coffea* species.** *Genome* 2004, **47**(3):501-509.
37. Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, de Kochko A, Hamon P: **SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus.** *Mol Genet Genomics* 2006, **276**(5):436-449.
38. Poncet V, Dufour M, Hamon P, Hamon S, de Kochko A, Leroy T: **Development of genomic microsatellite markers in *Coffea canephora* and their transferability to other coffee species.** *Genome* 2007, **50**(12):1156-1161.
39. Schlotterer C: **The evolution of molecular markers – just a matter of fashion?** *Nature Reviews Genetics* 2004, **5**(1):63-69.
40. Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A: **The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis.** *Molecular Breeding* 1996, **2**(3):225-238.
41. Jarne P, Lagoda PJL: **Microsatellites, from molecules to populations and back.** *Trend Ecol Evol* 1996, **11**(10):424-429.
42. Poncet V, Hamon P, Minier J, Carasco-Lacombe C, Hamon S, Noiro M: **SSR cross-amplification and variation within coffee trees (*Coffea* spp.).** *Genome* 2004, **47**(6):1071-1081.
43. Rallo P, Tenzer I, Gessler C, Baldoni L, Dorado G, Martin A: **Transferability of olive microsatellite loci across the genus *Olea*.** *Theor Appl Genet* 2003, **107**(5):940-946.
44. Estoup A, Tailliez C, Cornuet JM, Solignac M: **Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae).** *Mol Biol Evol* 1995, **12**(6):1074-1084.
45. Falush D, Stephens M, Pritchard JK: **Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies.** *Genetics* 2003, **164**(4):1567-1587.
46. Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A: **GENECLASS2: a software for genetic assignment and first-generation migrant detection.** *J Hered* 2004, **95**(6):536-539.
47. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.** *Genetics* 2000, **155**(2):945-959.
48. Estoup A, Angers B: **Microsatellites and minisatellites for molecular ecology: theoretical and experimental considerations.** In *Advances in Molecular Ecology* Edited by: Carvalho G. Amsterdam: NATO Press; 1998.
49. Rousset F: **Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance.** *Genetics* 1997, **145**(4):1219-1228.
50. Estoup A, Rousset F, Michalakis Y, Cornuet JM, Adriamanga M, Guyonard R: **Comparative analysis of microsatellite and allozyme markers: a case study investigating microgeographic differentiation in brown trout (*Salmo trutta*).** *Mol Ecol* 1998, **7**(3):339-353.
51. Nicolas V, Querouil S, Verheyen E, Verheyen W, Mbumba JF, Dillen M, Colyn M: **Mitochondrial phylogeny of African wood mice, genus *Hylomyscus* (Rodentia, Muridae): implications for their taxonomy and biogeography.** *Mol Phylogenet Evol* 2006, **38**(3):779-793.
52. Sehgal RNM, Jones HL, Smith TB: **Molecular evidence for host specificity of parasitic nematode microfilariae in some African rainforest birds.** *Mol Ecol* 2005, **14**(13):3977-3988.
53. Maley J: **La destruction catastrophique des forêts d'Afrique centrale survenue il y a environ 2500 ans exerce encore une**

- influence majeure sur la répartition actuelle des formations végétales. In *Plant systematics and phytogeography for the understanding of african biodiversity Proceedings of the XVth AETFAT Congress, held in 2000 at the National Botanic Garden of Belgium Volume 71*. Edited by: Robbrecht E, Degreef J, Friis I. Syst Geogr Pl; 2001:777-796.
54. Maley J, Chepstow-Lusty A: *Elaeis guineensis* Jacq. (oil palm) fluctuation in central Africa during the late Holocene: climate or human driving forces for this pioneering species? *Vegetation History and Archaeobotany* 2001, **10**:117-120.
 55. Leroy T, Montagnon C, Cilas C, Yapo A, Charmentant P, Eskes AB: Reciprocal recurrent selection applied to *Coffea canephora* Pierre 3. Genetic gains and results of first cycle intergroup crosses. *Euphytica* 1997, **95**(3):347-354.
 56. Anthony F, Combes MC, Astorga C, Bertrand B, Graziosi G, Lashermes P: The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. *Theor Appl Genet* 2002, **104**(5):894-900.
 57. Cordier L: Les objectifs de la sélection cafeiere en Cote d'Ivoire. *Café Cacao* Thè 1961, **5**(3):147-159.
 58. Ky CL, Barre P, Lorieux M, Trouslot P, Akaffou S, Louarn J, Charrier A, Hamon S, Noirot M: Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). *Theor Appl Genet* 2000, **101**(4):669-676.
 59. Paillard M, Lashermes P, Petiard V: Construction of a molecular linkage map in coffee. *Theor Appl Genet* 1996, **93**(1-2):41-47.
 60. Coulibaly I, Revol B, Noirot M, Poncet V, Lorieux M, Carasco-Lacombe C, Minier J, Dufour M, Hamon P: AFLP and SSR polymorphism in a *Coffea* interspecific backcross progeny [(*C. canephora* × *C. heterocalyx*) × *C. canephora*]. *Theor Appl Genet* 2003, **107**(6):1148-1155.
 61. Liu K, Muse SV: PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 2005, **21**(9):2128-2129.
 62. Felsenstein J: PHYLIP (Phylogeny Inference Package) version 3.6. In *Distributed by the author* Department of Genome Sciences, University of Washington, Seattle; 2005.
 63. Garnier S, Alibert P, Audiot P, Prieur B, Rasplus JY: Isolation by distance and sharp discontinuities in gene frequencies: implications for the phylogeography of an alpine insect species, *Carabus solieri*. *Mol Ecol* 2004, **13**(7):1883-1897.
 64. Cornuet JM, Piry S, Luikart G, Estoup A, Solignac M: New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* 1999, **153**(4):1989-2000.
 65. Rannala B, Mountain JL: Detecting immigration by using multilocus genotypes. *Proc Natl Acad Sci USA* 1997, **94**(17):9197-9201.
 66. Paetkau D, Calvert W, Stirling I, Strobeck C: Microsatellite analysis of population structure in Canadian polar bears. *Mol Ecol* 1995, **4**(3):347-354.
 67. Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F: GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. 4.05th edition. Laboratoire Génome, Populations, Interactions, CNRS UMR Université de Montpellier II, Montpellier (France); 5171:1996-2004.
 68. Hamon S, Dussert S, Deu M, Hamon P, Seguin M, Glaszmann JC, Grivet L, Chantereau J, Chevallier MH, Flori A, et al.: Effects of quantitative and qualitative Principal Component Score Strategies (PCSS) on the structure of coffee, rice, rubber tree and sorghum core collections. *Genetics Selection Evolution* 1998, **30**:237-258.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:

http://www.biomedcentral.com/info/publishing_adv.asp



BioMed Central

III – Conclusion

L'étude de la structuration de la diversité de *C. canephora* offre la possibilité de comprendre comment les changements climatiques ont pu influencer l'évolution et la structuration des espèces en général dans la région Guinéo-Congolaise. Or, comprendre les impacts des changements climatiques passés sur la diversité donne des clés pour appréhender les impacts potentiels de changements climatiques futurs.

L'échantillon de référence de l'espèce *C. canephora* (n = 25) construit au regard de la structuration de l'espèce établie au cours de ce travail, a été retenu. Il a été renforcé par des accessions cultivées supplémentaires de l'espèce.

Des échantillons de référence similaires qui représentent le maximum de la variabilité génétique connue sont mis en place pour l'espèce *C. liberica* avec huit individus de la sous-espèce *dewevrei* et dix individus de la sous-espèce *liberica*. Concernant l'espèce *C. arabica*, compte tenu de la faible variation observée au niveau des microsatellites (Silvestrini *et al.*, 2007), seuls le bourbon pointu et le bourbon rond ont été utilisés comme référence. Tous ces individus de référence ont été génotypés aux locus microsatellites sélectionnés.

Chapitre 4 - La mise en évidence d'un habitat favorable aux hybridations inter-spécifiques naturelles : détection génétique d'hybrides et analyses environnementales

Le travail présenté dans ce chapitre a fait l'objet d'une publication :

Gomez, C. et al. (Sous presse) Favourable habitats for Coffea inter-specific hybridization in central New-Caledonia: combined genetic and spatial analyses. Journal of Applied Ecology



Après l'abandon des plantations de caféiers traditionnelles sous couvert forestier en Nouvelle-Calédonie, les espèces se sont retrouvées contraintes à survivre dans le milieu naturel. Les prospections menées par Daniel Le Pierrès (1999) ont révélé qu'une seule région avait permis aux trois espèces introduites (majoritairement deux espèces *C. canephora* et *C. arabica*) de survivre et de cohabiter dans le même environnement (région de Sarraméa, Figure 2.8, p56). Il s'agit de comprendre comment les paramètres environnementaux de la région de Sarraméa autorisent la cohabitation des trois espèces et favorisent leur hybridation spontanée.

La région d'étude a été prospectée plus finement et quatre anciennes plantations multi-spécifiques présentant des hybrides inter-spécifiques, identifiés initialement selon des critères morphologiques puis confirmés génétiquement (Mahe *et al.*, 2007), ont été sélectionnées comme témoin pour leurs conditions environnementales. Trois d'entre elles présentent les deux espèces majoritaires en sympatrie (Farino, Couli et Dogny) et une présente les trois espèces en sympatrie (Sarraméa). Cette dernière a été choisie comme population témoin pour les études de diversité génétiques et la caractérisation des hybrides.

I – Contexte de l'étude et démarche abordée

Dans ce préambule, nous détaillerons les méthodes et les données complémentaires qui ont permis la caractérisation des conditions environnementales autorisant ces hybridations inter-spécifiques spontanées et qui n'ont pu être que succinctement abordées dans l'article même. Cette démarche se présente en trois étapes successives :

- 1) Dans un premier temps, il s'agit d'utiliser le même jeu de marqueurs que précédemment pour caractériser la diversité introduite pour chacune des espèces et les hybrides inter-spécifiques selon des méthodes d'assignations génétiques.

- 2) Ensuite, il s'agit de caractériser les zones favorables aux hybridations en termes de combinaisons environnementales particulières. Cette étape consiste principalement à comprendre comment la levée de barrières à la reproduction s'opère.

- 3) Enfin, le véritable travail de caractérisation des conditions environnementales favorables est basé sur une expertise comparant les conditions environnementales dans les milieux d'origine des espèces en Afrique avec celles rencontrées dans les potentielles zones hybrides dans la région d'étude. L'expertise combine à la fois des données bibliographiques (Afrique) ainsi que des données de terrain issues d'un important travail de suivi (plusieurs sites témoins au travers de la région d'étude en Nouvelle-Calédonie).

1.1 – Diversité introduite et caractérisation des hybrides

1.1.1 Le jeu de marqueurs utilisé

Le lot de 26 marqueurs microsatellites, sélectionné précédemment, a été utilisé sur les quatre échantillons de référence des trois espèces étudiées (34 *C. canephora*, 2 *C. arabica*, 8 *C. liberica* ssp *liberica*, 10 *C. liberica* ssp *dewevrei*) (Figure 4.1). Sur la totalité des individus des collections de référence (des trois espèces), et la totalité des marqueurs, 244 allèles ont été révélés. Le polymorphisme sur l'échantillon de *C. liberica* ssp *dewevrei* n'est pas représenté car il est montré ci-dessous qu'aucun individu de cette sous-espèce n'est détecté à Sarraméa.

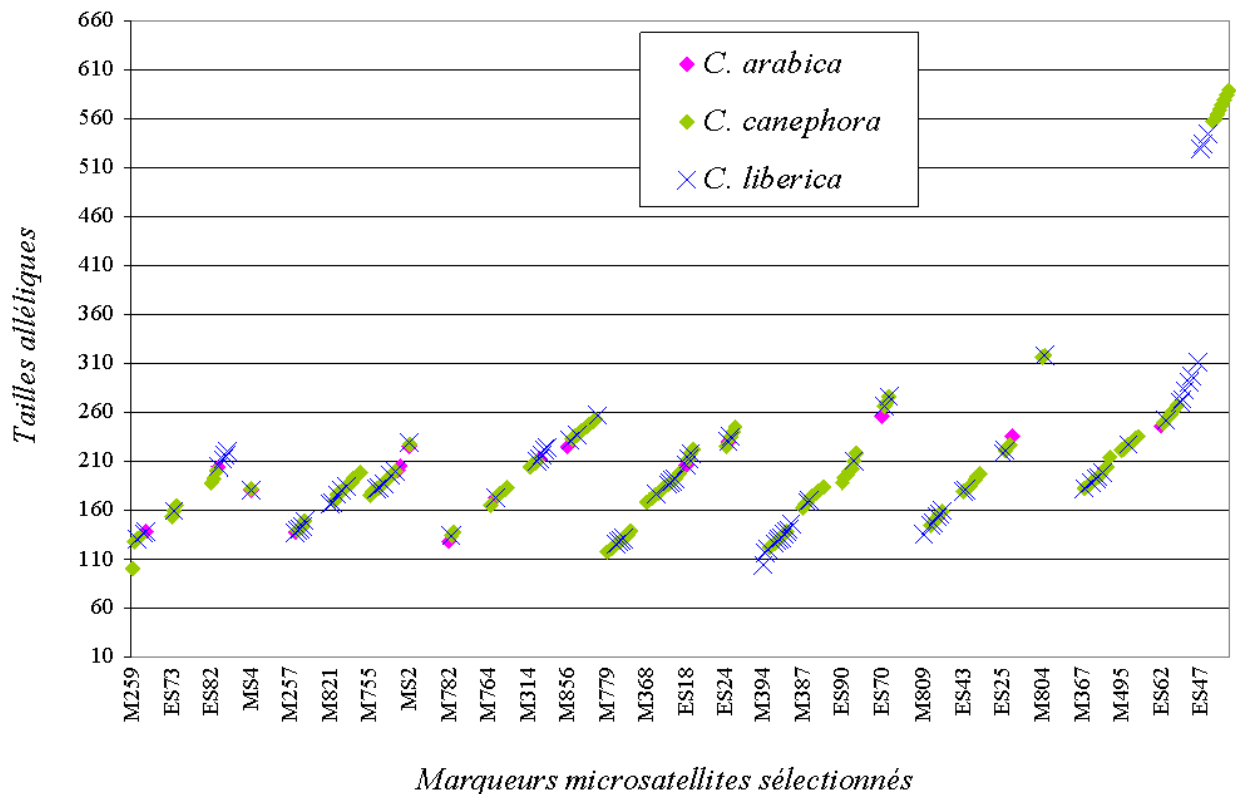


Figure 4.1 Représentation de la diversité allélique révélée par 26 marqueurs microsatellites sur l'ensemble des trois collections de référence. La taille des amplicons est indiquée en paires de bases.

Par ailleurs, plusieurs marqueurs du lot présenté ci-dessus ont des allèles espèce-spécifiques. Globalement, sur l'ensemble des marqueurs sélectionnés et les trois collections de référence, plus de la moitié de la totalité des allèles sont espèce-spécifiques (52%), c'est-à-dire qu'ils n'existent que chez une des trois espèces considérées. Par exemple, le marqueur ES47 distingue 13 allèles

aux tailles d'amplification différentes dont trois sont uniquement représentés par des individus de l'espèce *C. liberica* et six sont uniquement présents chez des individus de l'espèce *C. canephora* (Figure 4.2).

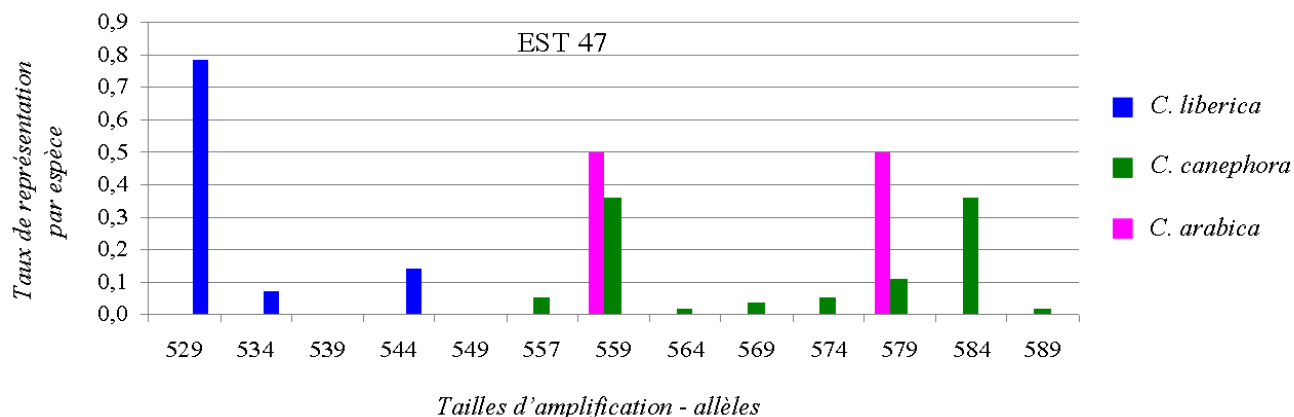


Figure 4.2 Représentation des fréquences des allèles par espèce, révélés par le marqueur ES47 sur les collections de référence des trois espèces

Nous retrouvons bien les caractéristiques pour lesquelles ces marqueurs ont été sélectionnés. Tous les marqueurs présentent un polymorphisme intra-spécifique et certains d'entre eux permettent de différencier les trois espèces grâce à des allèles diagnostiques spécifiques d'une espèce.

1.1.2 Analyses génétiques

Les 95 individus échantillonnés dans la population de Sarraméa, ont été génotypés aux 26 locus microsatellites. Deux types d'analyses sont effectués afin de caractériser la diversité introduite et les hybrides :

- l'identité des individus de Sarraméa est identifiée par des assignations génétiques aux groupes de référence afin d'identifier leur appartenance à une espèce ; les individus non assignés sont supposés hybrides,
- le mélange génétique des hybrides putatifs ainsi que la structuration génétique de la population en clusters sont identifiés par une analyse multilocus.

Dans un premier temps, les assignations génétiques permettent d'identifier l'appartenance d'un individu à des groupes génétiquement distincts. Il existe trois méthodes d'assignation génétique : une méthode basée sur les distances génétiques, une méthode basée sur les fréquences alléliques

des populations de référence et une méthode bayésienne basée sur le calcul de la probabilité qu'un génotype soit résident ou migrant sachant les fréquences alléliques de chaque population. Dans notre étude, nous avons utilisé la méthode bayésienne (Rannala, Mountain, 1997) proposée par le logiciel *GENECLASS2* (Piry *et al.* 2004). Nous avons assigné un échantillon d'individus composé des individus de référence et des individus de la population tri-spécifique aux populations de référence. En fonction du pouvoir d'assignation des individus de référence à leur propre population de référence, un seuil a été fixé (0.2). En deçà de ce seuil, les individus non assignés sont considérés comme des hybrides putatifs. Sur les 95 individus échantillonnés de la population de Sarraméa, sont distingués 67 individus appartenant à l'espèce *C. canephora*, 8 individus de l'espèce *C. arabica* et 16 individus de l'espèce *C. liberica ssp liberica*. Aucun individu de la sous-espèce *C. liberica ssp dewevrei* n'a été détecté.

Dans un deuxième temps, afin de détecter les hybrides inter-spécifiques, le logiciel *STRUCTURE* (Pritchard *et al.* 2000) a été utilisé. Aucun échantillon de référence n'est utilisé, seuls les 95 individus de Sarraméa sont intégrés à l'analyse. Le nombre de groupes génétiques (clusters) présents dans l'échantillon est déterminé et le degré de mélange génétique de chaque individu est estimé. Le nombre de clusters présent dans l'échantillon est testé comme la valeur qui maximise la probabilité *a posteriori*. Dans notre étude, quatre clusters ont été distingués : un correspondant à l'espèce *C. arabica*, un à *C. liberica* et deux à *C. canephora*. Ceci reflète le haut niveau de diversité de l'espèce *C. canephora* introduit en Nouvelle-Calédonie. Trois hybrides inter-spécifiques ont été détectés et présentent des niveaux de mélange génétique supérieur à 20% pour le génome sous-représenté. Un quatrième individu est supposé hybride car il présente 12% de mélange génétique.

1.2 – Comment la levée de barrières à la reproduction s'opère-t-elle ?

Ces trois espèces sont originellement isolées géographiquement en Afrique mais se retrouvent à Sarraméa dans un même environnement. Par ailleurs, leurs phénologies florales sont naturellement décalées ne permettant pas de floraisons synchrones (cf. Figure 2.7, p51). Si on considère les hybrides trouvés au site de Sarraméa, une levée de barrières à la reproduction entre espèces doit s'effectuer notamment au niveau des phénologies florales.

1.2.1 Mesures effectuées et suivi météorologique local

Suivant ce constat, le suivi des floraisons a été réalisé soigneusement sur toute l'année 2008 après avoir précisé le système de classification (stade / densité / localisation des individus selon les espèces). Des passages mensuels, puis hebdomadaires à partir du mois de septembre et journaliers lors des pics de floraisons, ont été effectués.

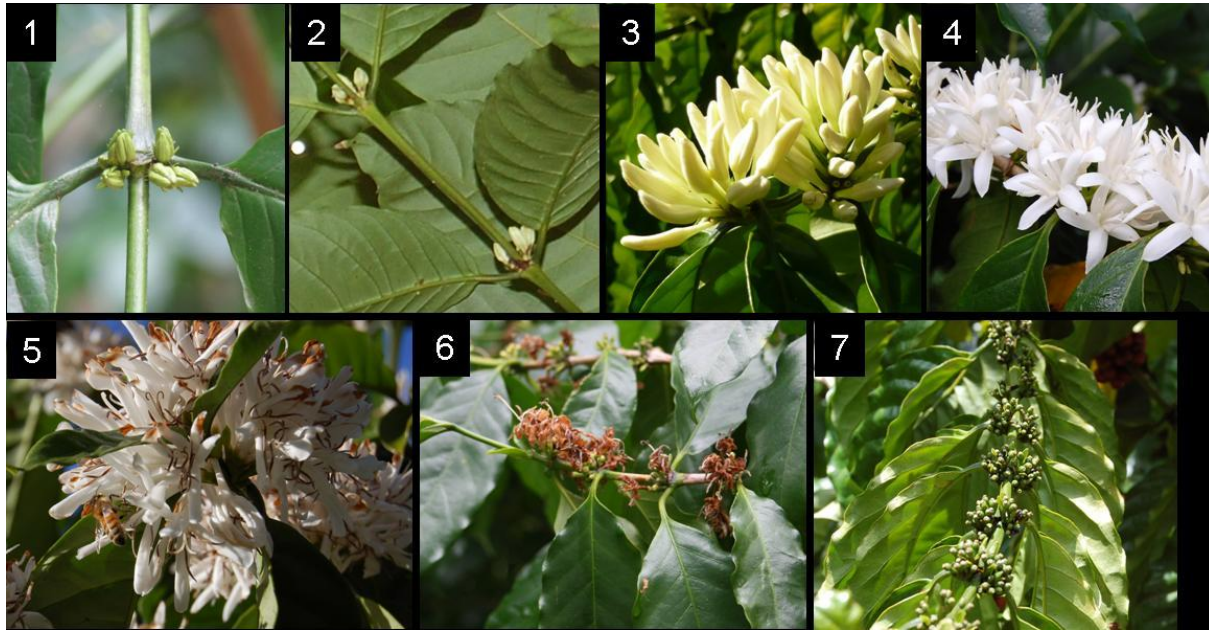


Figure 4.3 : Photographies des stades de floraison annotés

Les stades de floraison suivants ont été relevés (Figure 4.3). Le stade 1 correspond à la mise en place de l'inflorescence. Les bourgeons floraux qui la constituent entrent alors en phase d'induction. L'induction des bourgeons nécessite un stress hydrique d'au moins 30 jours (Arcila-Pulgarin *et al.*, 2002) (*cf.* Figure 2.7). Le stade 2 présente les bourgeons floraux à la fin de la période d'induction. La floraison se déclenche à la suite de la première pluie d'au moins 5mm (Le Pierrès, communication personnelle) qui rompt ainsi le stress hydrique des jours précédents. Le bourgeon devient alors une chandelle (stade 3). Le stade 3 ne dure que quelques jours, parfois même un seul jour. L'anthèse a déjà démarré au stade 3, le pollen est déjà turgescent. L'ouverture des fleurs a alors lieu au stade 4. Les anthères sont bien blanches, les pétales aussi et les fleurs libèrent un doux parfum proche de celui de la fleur d'oranger. Dès le lendemain (stade 5) les anthères semblent grillées, et le pollen n'est alors plus disponible. Les stigmates restent encore réceptifs pendant quelques jours. Le stade 6 correspond au stade où les pétales se fanent. Les

stigmates ne sont alors plus réceptifs. Les fleurs tombent alors et l’ovaire deviendra un fruit s’il a été fécondé (stade 7).

Des classes de densité ont été définies pour l’arbre entier allant de 1 (une fleur à 2-3 sur l’arbre) à 4 (arbre entièrement fleuri). Ce système a été délicat à mettre en place étant donné la différence existant entre espèces en ce qui concerne le nombre d’inflorescences par glomérule ou la taille des fleurs (*cf.* Tableau 2.3). Dix individus de chaque espèce ont été observés, répartis géographiquement dans toute la population tri-spécifique et répartis sur toute la gamme de taille des espèces.

Les données des stations de Météo France, entourant la zone d’étude, ne s’avèrent pas suffisamment proches de notre site d’observation pour obtenir une bonne précision à Sarraméa. Nous avons donc mis en place une station météo performante (Vantage Pro2, Davis) à Sarraméa, nous permettant d’enregistrer des données précises à un pas horaire, à proximité des individus observés. La station météo propose un ensemble de capteurs : température (°C), hygrométrie (%), pression atmosphérique (hPa), direction et force du vent (m/s) et précipitations (mm).

1.2.2 Trois événements de floraisons principaux qui permettent des recouvrements entre espèces

Trois périodes de floraisons majeures ont été observées aux alentours du 22 septembre 2008, du 3 octobre 2008 et du 1^{er} novembre 2008.

Date pluie	Pluie (mm)	Moyenne sur les jours précédents	Lib	Can	Ara
12 sept	11.4 mm	0,47 [0 ; 4,6] mm; 33 j	+10 j	+ 11 j	+ 14 j
26 sept	67 mm	0,0 [0,0 ; 0,2]mm; 11 j	+7 j	+8 j	+10 j
23 oct	58 mm	0,6 [0,0 ; 4,8] mm; 25 j	+9 j	-	+13 j

Tableau 4.1 : Récapitulatif de la saison de floraison 2008 avec la date des pluies déclenchantes, leur intensité (Pluie mm), durée de la saison sèche qui précède (moyenne sur les jours précédents [minimum ; maximum] - durée en jours) et le nombre de jours après la pluie (délais de floraison; Lib = C. liberica ; Can= C. canephora ; Ara = C. arabica)

Les trois floraisons observées suivent de façon évidente une pluie déclenchante (supérieure à 10 mm) précédée d’une période relativement sèche (moins de 5 mm par jour, en moyenne moins de

0.6 mm sur plus de 11 jours). L'espèce *C. liberica* a fleuri deux autres fois à la fin du mois de novembre, puis au milieu du mois de décembre.

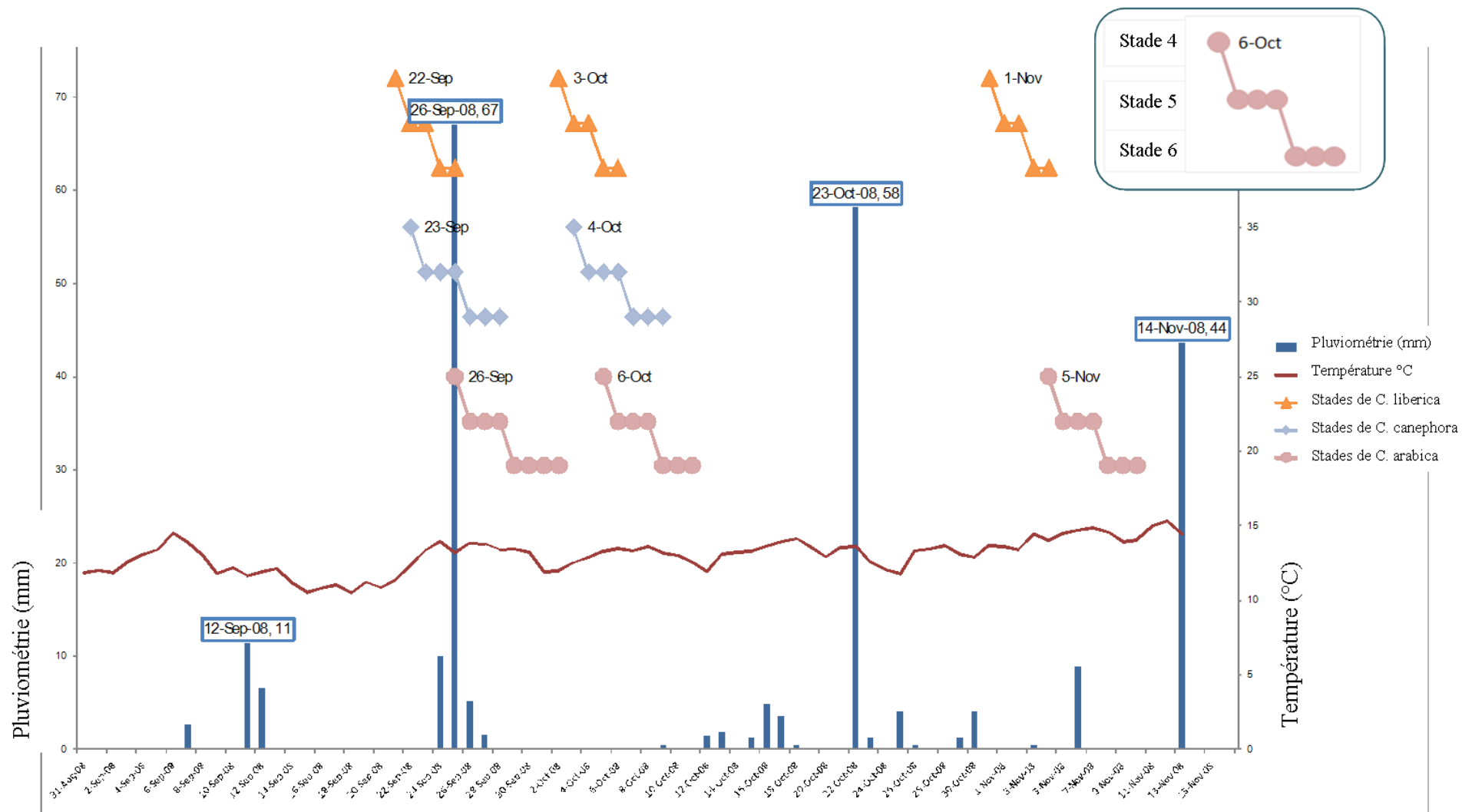
Il est à noter que pour chaque période, l'ordre de floraison habituellement observé pour ces mêmes espèces en Afrique (Berthaud *et al.*, 1989; Charrier, Berthaud, 1985) est conservé. *C. liberica* fleuri la première, suivie de *C. canephora*, et enfin de *C. arabica*. On observe une hétérogénéité au sein même des glomérules, car un même individu peut fleurir plusieurs fois au même nœud (Figure 4.4).



Figure 4.4 : Floraison multiple sur le deuxième nœud d'une branche d'*arabica*. La première floraison seule mènera à des fruits

Le chevauchement des stades d'une espèce à l'autre pour chaque événement de floraison permet les échanges de pollen entre espèce. Si on regarde les séquences chronologiques en Figure 4.5, on observe une concordance des stades 5 de la première espèce florissante avec le stade 4 de la suivante.

Figure 4.5 : Historique des floraisons de la saison 2008 présentant les deux stades de floraisons fertiles (stade 4 et 5) et le stade 6 par espèce en fonction des séquences climatiques (précipitations et température)



Ainsi, des superpositions de floraison entre espèces à Sarraméa sont réelles, comme par exemple *C. liberica* stade 5 avec *C. canephora* stade 4 et *C. canephora* stade 5 avec *C. arabica* stade 4.

Afin de mieux comprendre la distribution des floraisons par espèce et sur l'ensemble de la saison (de septembre à octobre), des moyennes d'intensité de fleurs ouvertes (stade 4) ont été calculées par espèce (= somme des densités du stade 4 pour tous les individus d'une espèce / nombre d'individus observés). Il apparaît nettement que le pic de floraison des individus de l'espèce *C. canephora* se produit au moment de la première floraison majeure, le pic de *C. arabica* se produit au moment de la deuxième floraison majeure et enfin le pic de *C. liberica* se produit au moment de la troisième floraison majeure (Figure 4.4).

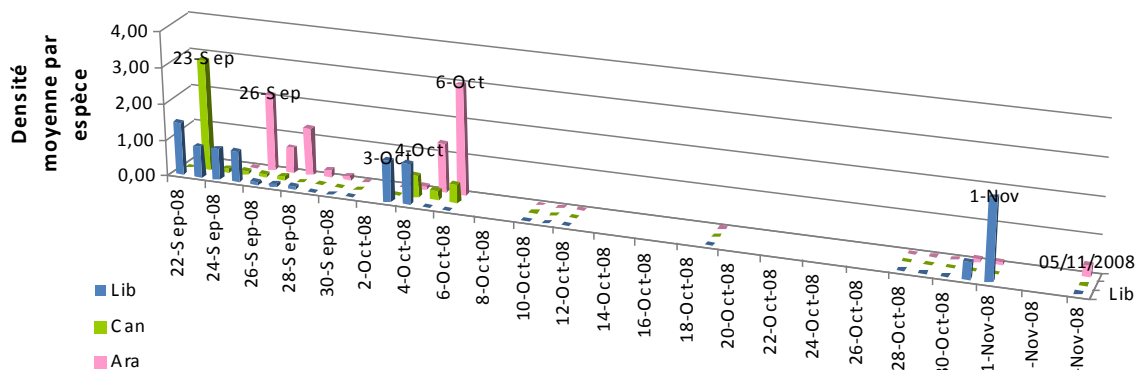


Figure 4.4 Intensités des floraisons observées par espèce

1.2.3 Conclusion

Cette analyse des séquences de floraison a montré que les tendances générales des phénologies florales des trois espèces sont comparables à ce que l'on connaît en Afrique. L'ordre est donc le même qu'en Afrique. Mais les délais ne sont pas figés ce qui permet des recouvrements de floraison entre espèces. L'observation de ces phénologies florales montre qu'elles sont intimement liées aux conditions climatiques et notamment à la pluviométrie.

1.3 – Importance de caractériser les conditions environnementales favorables

La région de Sarraméa est tout à fait originale par rapport au reste de l'île, puisqu'elle est la seule à présenter des espèces de caféiers en sympatrie. Les prospections plus fines au sein même de la région ont permis de révéler une hétérogénéité des conditions environnementales. Il s'agit ici d'explorer le lien existant entre la présence de plusieurs espèces de caféiers et l'organisation spatiale pour mettre en évidence des zones favorables à la présence de caféiers en sympatrie et à leur éventuelle hybridation. Pour cela nous travaillons sur les quatre sites témoins présentés (Farino, Couli, Sarraméa et Dogny) et nous utilisons 20 autres populations multi-spécifiques et présentant une dynamique d'hybridation pour la validation du modèle environnemental mis en place dans ce chapitre.

Cette approche qualitative prend en compte un certain nombre de paramètres environnementaux afin de caractériser le milieu environnant. Les modalités de ces paramètres sont définies en fonction des dires d'experts, de la biologie des caféiers et selon les valeurs témoins de ces paramètres aux 4 sites d'étude: Farino, Couli, Sarraméa et Dogny.

Ce travail croise ainsi des informations biologiques et génétiques (concernant les hybridations interspécifiques) et des données environnementales et géographiques grâce à la spatialisation des données à intégrer au modèle et aux Systèmes d'Information Géographique.

Après l'identification des paramètres influençant la présence des différentes espèces de caféiers en sympatrie, l'objectif est d'intégrer le maximum de données spatialisées obtenues par les institutions locales (Tableau 4.2). En effet, cette démarche est nécessaire pour la construction de l'analyse spatiale par approche multicritère.

Tableau 4.2 : Recensement des données utilisées. DITTT : Direction des Infrastructures de la topographie et des Transports terrestres ; ADRAF : Agence de développement Rural et d'Aménagement Foncier ; DTSI : Direction des Technologies et Services de l'Information ; VALPEDO : Unité de l'IRD spécialisées en pédologie ; MNT : Modèle Numérique de Terrain ; ZE : Zone d'étude restreinte à Sarraméa

Données	Source	Emprise géographique	Précision géographique
BD Carto	DITTT	Grande Terre	1/50000e
BD Topo	DITTT	Grande Terre	1/10000e
Données météo	Météo France	ZE	50m
Cadastres	DITTT/ADRAF	ZE	
MNT	DTSI	Grande Terre	50m
Images Quickbird	DigitalGlobe	ZE	0,66m

Par ailleurs, les caractéristiques climatiques des régions d'origine des espèces, à savoir en Côte d'Ivoire pour *C. canephora* et en Ethiopie pour *C. arabica*, sont importantes pour mettre en évidence les exigences environnementales des caféiers, non seulement en termes de survie mais aussi de reproduction. Il s'agit de croiser les caractéristiques environnementales des sites d'études (4 sites en Nouvelle-Calédonie) avec celles des milieux de référence (Divo et Jimma en Afrique), et de définir des classes allant du favorable au défavorable (Tableau 4.3).

Tableau 4.3 : Valeurs des paramètres aux sites d'étude en Nouvelle-Calédonie et aux sites de référence en Afrique. (P-juin = regain de précipitations au mois de juin par rapport au mois précédent = précipitations de mai – précipitations de juin)

	Farino	Dogny	Couli	Sarraméa	Divo	Jimma
Description	bi-spécifique	bi-spécifique	bi-spécifique	tri-spécifique	Côte d'Ivoire – témoin pour <i>C. canephora</i>	Ethiopie – témoin pour <i>C. arabica</i>
Longitude (IGN72-UTM58S)	578285.49	595763.45	588868.09	585306.10		
Latitude (IGN72-UTM58S)	7604870.8	7605123.3	7606154.9	7607577.3		
Altitude	199	140	248	326	152	1740
Distance à un creek	<100 mètres	<100 mètres	<100 mètres	<100 mètres		
Ecart annuel de température (°C)	8.11	7.59	7.70	7.77	3.2	3.1
P-juin (mm)	-68.97	40.19	-8.4	16	-70	-20
P-juillet (mm)	106.79	-10.74	40.16	17.48	160	65
P-août (mm)	-57.35	-22.63	-162.75	-97.34	55	-80
Maximum précipitations mensuelles (juin-août) (mm)	200	141	284	215	260	240
Cumul annuel pluie (mm)	2368.19	1977.77	3953.84	3088.61	1314	1610

En fonction de cette analyse détaillée, les paramètres caractérisant au mieux les possibilités d'hybridation (notamment les données de pluviométrie qui expliquent les phénologies florales observées) et représentant au mieux les caractéristiques environnementales particulières des quatre sites témoins (l'altitude n'est pas sélectionnée car très hétérogène entre sites) sont intégrés dans l'analyse multicritère. Cette analyse est composée des paramètres environnementaux suivants : l'occupation du sol (issue de l'image à très haute résolution Quickbird), la distance à un cours d'eau temporaire (creek), la pente (issue du MNT), l'écart annuel de température, et les critères concernant la pluviométrie. Le modèle est ensuite construit en utilisant les points « vérités » dont les caractéristiques environnementales sont connues : ce sont les quatre sites témoins (Sarraméa, Couli, Dogny et Szemelweiss) dont nous avons extrait les valeurs de chacun des paramètres du

modèle. Les différentes modalités sont intégrées dans un arbre de décision qui évalue toutes les combinaisons de paramètres possibles en y attribuant une « note arbitraire » selon les dires d'experts. L'ordre des variables est sans importance dans la mesure où l'indice calculé consiste à sommer les notes. Les résultats de l'application de l'arbre de décision sur la zone d'étude sont vérifiés à 20 autres sites géoréférencés qui présentent au moins deux espèces et des hybrides inter-spécifiques.

Ce modèle d'aide à la décision est préliminaire. Les grandes tendances qu'il met en évidence sont toutefois importantes pour la suite de notre recherche.

II– Résultats principaux

Les objectifs de cet chapitre étaient :

- 1) de préciser la nature des hybridations dans une plantation multi-spécifique. Sur les quatre sites témoins, nous avons choisi la population tri-spécifique de Sarraméa comme référence et nous avons analysé la diversité génétique présente et caractérisé les hybrides inter-spécifiques sur la base des marqueurs microsatellites,
- 2) d'évaluer les facteurs environnementaux favorables à l'hybridation par expertise biologique et écologique et selon les valeurs des paramètres aux sites témoins par rapport aux régions d'origine,
- 3) de proposer un modèle de prédiction de distribution des zones favorables aux hybridations.

Les résultats obtenus ont fait l'objet d'une publication, sous presse :

Gomez, C. *et al.* (Sous presse) Favourable habitats for *Coffea* inter-specific hybridization in central New-Caledonia: combined genetic and spatial analyses. ***Journal of Applied Ecology***

Favourable habitats for *Coffea* inter-specific hybridization in central New Caledonia: combined genetic and spatial analyses

Céline Gomez^{1,2}, Ahmed Batti³, Daniel Le Pierrès¹, Claudine Campa¹, Serge Hamon¹, Alexandre de Kochko¹, Perla Hamon¹, Frédéric Huynh³, Marc Despinoy⁴ and Valérie Poncet^{1*}

¹IRD, UMR DIAPC, 911 avenue Agropolis, BP 64501, 34394 Montpellier CEDEX 5, France; ²IRD, UMR DIAPC, 101 Promenade Roger Laroque – Anse Vata, BP A5, 98848 Cedex Nouméa, New Caledonia; ³IRD, Unité ESPACE (S140), Maison de la Télédétection, 500 rue J.F. Breton, Agropolis, 34093 Montpellier Cedex 5, France; and ⁴IRD, Unité ESPACE (S140), 101 Promenade Roger Laroque – Anse Vata, BP A5, 98848 Cedex Nouméa, New Caledonia

Summary

1. Human disturbance of natural habitats has increasingly promoted hybridization between previously isolated species. Coffee trees (*Coffea canephora*, *C. liberica* and *C. arabica*) have been introduced into New Caledonia for cultivation since the 1850s. Many plantations have since been abandoned, leaving the cultivars to evolve in natural conditions. This historical situation and the local environmental conditions have led to gene flow between sympatric species and the creation of hybrid zones which can be considered as a natural laboratory for studying gene flow dynamics. Our objectives were to assess the nature of these hybridizations and to model the distribution of habitats favourable for these hybrid zones at a regional scale.

2. We analysed a tri-specific model population with 26 microsatellite markers using a multilocus analysis approach. We applied Bayesian methods to characterize the species identity of founder individuals and identify inter-specific hybrids. A high level of genetic diversity and inter-specific hybridization was observed in the population. To evaluate the presence of other favourable environments in the local area, a geographic information system was used to spatially differentiate the phenomenon and optimize future hybrid collecting missions. We constructed a predictive distribution model based on observation and environmental expertise (climatic parameters, physical parameters and landcover) at four reference sites of the hybrid zone.

3. At all four sites the *Coffea* species were found to be involved in inter-specific hybridization, with one population, in particular, revealing high genetic diversity. According to the environmental expertise, 40% of the studied region appeared to be favourable for this phenomenon and the model was validated with a prediction accuracy of 79%.

4. *Synthesis and applications.* The use of microsatellite markers was efficient for assessing the nature and extent of coffee tree hybridization events. The environmental expertise and model allowed comparisons between New Caledonia and Africa ecological niches. This methodology may be recommended for better management of *de novo in situ* coffee genetic resources and applied to other plants to provide insight into the impact of adaptation to new environment. Combining environmental and species biological parameters would help in analysing modifications in plant reproductive behaviour induced by global climate change.

Key-words: *Coffea* sp., decision tree, environmental expertise, genetic structure, GIS, inter-specific hybridization, New Caledonia, SSR

Introduction

Human disturbance of natural habitats has often promoted hybridization between previously isolated species. This is

*Correspondence author. E-mail: valerie.poncet@ird.fr

particularly the case for plants which have been introduced in habitats that differ from their original environments. These introductions have disturbed the natural distribution of species and could have led to gene flow between related species (Lamont *et al.* 2003). Hybridization processes and the structure of hybrid zones are valuable models for understanding speciation mechanisms (Lexer *et al.* 2005; Mallet 2005; Gerard *et al.* 2006; Mallet *et al.* 2007). Indeed, they have long been seen as a natural laboratory for studying species in terms of boundaries and dynamics such as in a mixed oak forest (Valbuena-Carabana *et al.* 2005) or among sympatric species of *Rhododendron* (Milne *et al.* 1999). Several crucial questions concern barriers to gene flow between species, and also the fitness of hybrids in later generations, as they are essential for maintenance of the hybrid zone through local adaptation.

When a species is displaced in a new environment, the plant equilibrium could be disturbed and trigger new dynamics (Benkman, Siepielski & Parchman 2008). For example, temporal isolation by divergent floral phenologies can be a strong prezygotic barrier (Rieseberg & Willis 2007). This phenomenon is often linked to rain and/or temperature cycles. Temporal isolation could be modified in cases of habitat disturbance or species introductions, thus promoting hybridization. The adaptation concept is also fundamental as the resulting hybrids could be adapted to the environment in which they have been created – generations are mixed and a hybrid swarm may be observed (Tiranti & Negri 2007).

To gain insight into how environmental factors act on hybridization dynamics, studies should be carried out in regions where hybrid zones are dispersed, within which there are different kinds of interaction between species and variations in environmental factors (temperature, precipitation, soil, associated botanic species, etc.; Guisan & Zimmermann 2000). The environment could be described through analyses of spatial correlations, while taking the environmental factors involved into account, so as to be able to pinpoint favourable areas.

The analysis of species–environment relationships is a central issue in ecology (Anderson, Gomez-Laverde & Peterson 2002; Ferguson *et al.* 2005; Jarvis *et al.* 2005). Climate combined with other environmental factors could be useful for explaining animal and plant distributions (Guisan & Zimmermann 2000), and especially for improving descriptions of species requirements (Booth *et al.* 2000; Booth & Ekelme 2002).

Geographic information systems (GIS) and predictive distribution models are increasingly employed together to assess spatial differentiation in the studied phenomenon (Wilkinson *et al.* 2003; Jarvis *et al.* 2005; Adams *et al.* 2007). GIS tools have two main properties: (i) they integrate large data bases with georeferenced locations and (ii) they rigorously and efficiently quantify spatial patterns (Swenson 2008). Guarino *et al.* (2001) reviewed the uses of GIS in plant genetic resource management. Some studies provide a conceptual framework for using GIS in plant assessments (Guisan & Zimmermann 2000), conservation initiatives, and subsequent use and impact appraisals (Jarvis *et al.* 2005), whereas the combined use of

molecular analysis and GIS modelling for genetic resource management studies is a powerful tool that can be implemented to gain insight into the impact of adaptation and global changes (Elliott *et al.* 2004; Adams *et al.* 2007; Thuiller *et al.* 2008).

Cultivated *Coffea* species, i.e. *C. canephora* (Robusta), *C. arabica* and to a lesser extent *C. liberica*, have been introduced from Africa into other continents where environmental conditions are suitable for their cultivation, i.e. to fulfil their life cycle from germination to fructification. All species of the *Coffea* genus are native to intertropical Africa, Madagascar and the Mascarene Islands, but with different geographical and environmental ranges (Chevalier 1946; Davis *et al.* 2006). *Coffea canephora* and *C. liberica* have a distribution area that ranges from Guinea to Sudan and from Cameroon to Angola. *Coffea canephora* is present in humid tropical forests (with both high nocturnal and diurnal temperatures) at low elevations (250–1500 m). *Coffea liberica* consists of two well-differentiated sub-species, i.e. *liberica* and *deweyi* (N' Diaye *et al.* 2005), and is present in similar ecosystems. Conversely, *C. arabica* naturally grows in East African highland forests (Ethiopia, Sudan and Kenya) from 950 m to over 2000 m, with low nocturnal temperatures but rarely under 0 °C (Davis *et al.* 2006). The elevation effect could be offset by the latitude. These three coffee species are naturally adapted to different environments, which could explain why some traits are differentiated. In particular, high variation in flowering/fructification cycle lengths have been observed. Flowering in *Coffea* is initiated after a dry period of at least 30 days by an adequate rain of at least 5 mm (Le Pierrès 1995). Depending on the species, 4–8 days after this rain, blossoming occurs and styles are receptive for just 1 day. Consequently, as flowering occurs after 5–6, 7 and 8 days for *C. liberica*, *C. canephora* and *C. arabica*, respectively, no natural cross-pollination takes place between these species. However, when the preliminary dry season is interrupted by sporadic rains, flowering could be affected and spread over several months.

In New Caledonia, *Coffea* species have been introduced since 1856 (see Appendix S1 in Supporting Information), and natural hybridizations between the two main species has been noted in abandoned shade plantations (Mahe *et al.* 2007), but only in the central part of the main island. A tri-specific plantation was identified in Sarraamea Valley (21°38'58"S, 165°51'733"E) (Le Pierrès 1999). This former plantation, where three *Coffea* species co-exist with an extraordinary level of regeneration through seedlings, could be likened to a natural laboratory where the diversity has evolved only under the pressure of local environmental conditions.

Now that efficient molecular markers are available, putative hybridization can be assessed in an ecological context. Microsatellite markers are especially relevant for this approach as they provide information on the origin of the introduced genetic variation (Durka *et al.* 2005) and also provide a tool to detect and analyse hybridizations. In coffee, recently developed markers (Poncet *et al.* 2004, 2006, 2007) have been used to characterize natural genetic diversity detected within

C. canephora (Gomez *et al.* 2009), and they have proved to be informative enough to discriminate between *Coffea* species (Cubry *et al.* 2008).

Our specific objectives were to: (i) assess the extent and nature of hybridization events in abandoned plantations where two or three of the cultivated coffee species had been initially introduced. Out of the four reference site populations, the three-species Sarramea population was analysed as a reference population using microsatellite markers; (ii) evaluate, based on both biological and ecological expertise and data obtained at the reference sites, environmental factors favourable for the introduced species' adaptation and hybridization and finally (iii) propose a predictive habitat distribution model based on these variables so as to optimize collecting missions, i.e. to find other sites where similar hybridization phenomena might occur. The final aim of this study was to recommend models for better management of *de novo in situ* coffee genetic resources.

On a larger scale, this approach could be helpful for combining environmental and species biological parameters to analyse modifications in plant reproductive behaviour induced by global climate change.

Materials and methods

LOCATION OF THE STUDY AREA

New Caledonia is located in the southwest Pacific Ocean (21°30'S, 165°30'E) 1200 km east of Australia and 1500 km northwest of New Zealand. Throughout the study area, in the central part of the mountain range, four sites were delineated to gain further insight into coffee inter-specific hybridization ecological requirements (see Fig. S1). These reference sites were Farino (21°39'30"S, 165°45'25"E), Couli (21°38'10"S, 165°49'65"E), Sarramea (21°38'58"S, 165°51'73"E) and Dogny (21°39'15"S, 165°55'78"E). At least two coffee species were present at each site and inter-specific hybrids have been observed (Le Pierrès 1999; Mahe *et al.* 2007). The Sarramea site is the only

one where the three coffee species (*C. arabica*, *C. canephora* and *C. liberica*) coexisted and was chosen for the genetic analysis. Twenty additional sites were used afterwards to test and validate the model. A topographic map, at 1 : 10 000 scale, was used to scan and geo-reference the reference points. Tracking, projected on the Universal Transverse Mercator (UTM) zone 58S grid, was done on the local IGN72 Geographic Coordinate System (Institut Géographique National, France) (see Fig. S1 in Supporting Information).

GENETIC ANALYSIS OF THE SARRAMEA TRI-SPECIFIC POPULATION

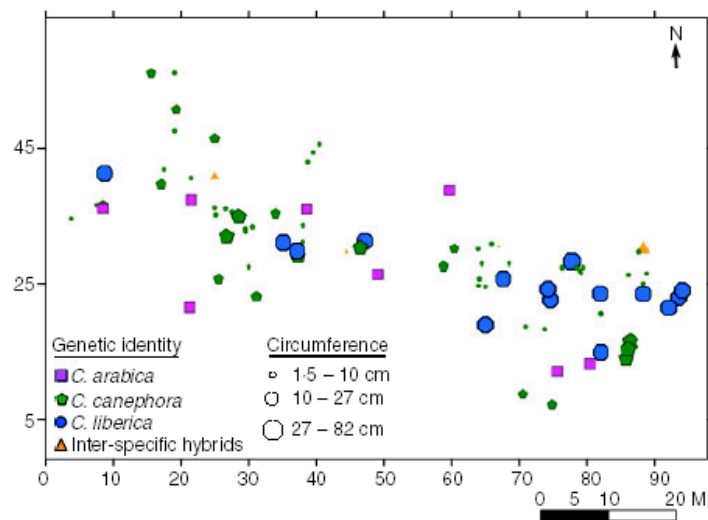
Plant sampling

The Sarramea coffee plantation was probably abandoned in the 1930s. The circumference of individual trees was measured at 1 m height so as to reflect individual age. Hence, 95 individuals were retained, including some of the oldest trees (called founders) and morphologically putative hybrids (Fig. 1). To trace the origin and identity of these introduced individuals, we selected African reference sets of the three species: *C. canephora* ($n = 34$, representative of the five diversity sub-groups) (Gomez *et al.* 2009), *C. arabica* ($n = 2$), *C. liberica* ($n = 18$: ssp. *liberica*, $n = 8$ and ssp. *dewevrei*, $n = 10$) (N'Diaye *et al.* 2007).

Molecular markers

For all plants, total DNA was extracted according to Ky *et al.* (2000). Twenty-six microsatellite loci were selected for being evenly distributed throughout the *Coffea* genome. Nine of them mapped on nine different linkage groups of the (*C. heterocalyx* × *C. canephora*) inter-specific map (Coulibaly *et al.* 2003). The 16 simple sequence repeats (SSR) and 10 expressed sequence tag microsatellite (EST-SSR) loci were amplified according to Poncet *et al.* (2006, 2007). PCR products were detected on an IR2 Automated DNA Sequencer (LI-COR, model 4200L-2, Lincoln, NE, USA) using an M13 primer coupled to the infrared tag IRD700 or IRD800 after migration on 25 cm 6.5% KBplus (LI-COR; CAT#827-05607) polyacrylamide gels. The gel images were processed by SAGA GTTM software (LI-COR Biotech)

Fig. 1. Relative coffee tree position in the former Sarramea plantation. Distances on the x and y axes are given in metres. Species and hybrid genetic identities were defined during the course of this study.



to estimate the size of amplicons according to a 50–350 bp size standard (LI-COR, CAT#829-05343, 829-05344).

Genetic diversity

For each SSR and EST-SSR locus, genetic polymorphism within the samples (the four African reference groups and the New Caledonian sample) was assessed by calculating the observed number of alleles (N_a), observed and expected heterozygosity (H_o and H_e) and polymorphism information content (PIC) using POWERMARKER v3.25 software (Liu & Muse 2005). The number of group-specific (private) alleles was calculated in the African reference groups. The partition of the genetic variation between reference genetic groups was estimated with the F_{ST} of Weir and Cockerham implemented in GENETIX (Belkhir *et al.* 1996–2004). The significance of pairwise F_{ST} values was calculated using permutation tests ($N = 1000$).

Genetic origin and characterization of Saramea trees

To identify the genetic identity (in terms of species) of the 95 individuals studied from the Saramea population, we used the African genotypes ($n = 54$) as a reference set for the assignment analysis.

The first assignment was done with the implemented method described in GENELASS 2.0 (Piry *et al.* 2004). A Bayesian-based approach was used (Rannala & Mountain 1997) with the simulation algorithm based on frequencies with a risk of 0.01. A Saramea individual was considered as correctly assigned to a reference African species when the assignment probability was higher than the minimum self-assignment of African species (0.2).

The genetic structure was also investigated with the Bayesian approach implemented in STRUCTURE v2.1 (Pritchard, Stephens & Donnelly 2000). This approach determines the population structure and estimates the degree of admixture for each individual. This model was implemented on the whole sample (i.e. 95 genotypes from Saramea and 54 of the reference set). This procedure allowed identification

of a given species cluster by comparison with the assignment of reference species genotypes. Selected parameters were 30 000 for the burn-in time and 1 000 000 for the number of runs. The cluster number was tested as the value that maximized the posterior probability of the data $\ln P(D)$ according to the formula $[\ln P(D)_k - \ln P(D)_{k-1}]$, as suggested by Garnier *et al.* (2004). The proportion of membership, q_i , of each genotype to the five genetic groups, i.e. the proportion of its genome drawn from each group was also evaluated. Each individual genotype was assigned to a given group when the average proportion of membership was $q_i > 0.80$, i.e. over 80% ancestry to their own cluster. In case of admixed individuals, genotypes were assigned to two or more groups, if the proportion of membership to each one was $0.20 < q_i < 0.80$. A 90% CI of the q_i parameter was calculated for each individual.

ENVIRONMENTAL CHARACTERIZATION AND PRELIMINARY PREDICTIVE DISTRIBUTION

The aim of the environmental approach was to gain insight into the processes that enhance the coexistence of *Coffea* species and inter-specific gene flows. Based on three levels of data: the history, the ecological demands of coffee species and the environmental conditions for hybridization, a multi-criteria analysis was performed, which included different environmental components: (i) climatic sequences, (ii) soil moisture – through the slope and the temporary hydrographic network and (iii) the shade – through the land cover map (Fig. 2). Decision parameters were determined according to their impact on coffee tree fitness and on genetic barrier effectiveness between species. This expertise was supplemented with environmental data collected from the four reference sites. A predictive distribution model was finally constructed based on the modalities of each decision parameter combined in a decision tree (Fig. 2). The most important part of the spatial analysis was the detection and computation of environmental predictive variables. A preliminary predictive model

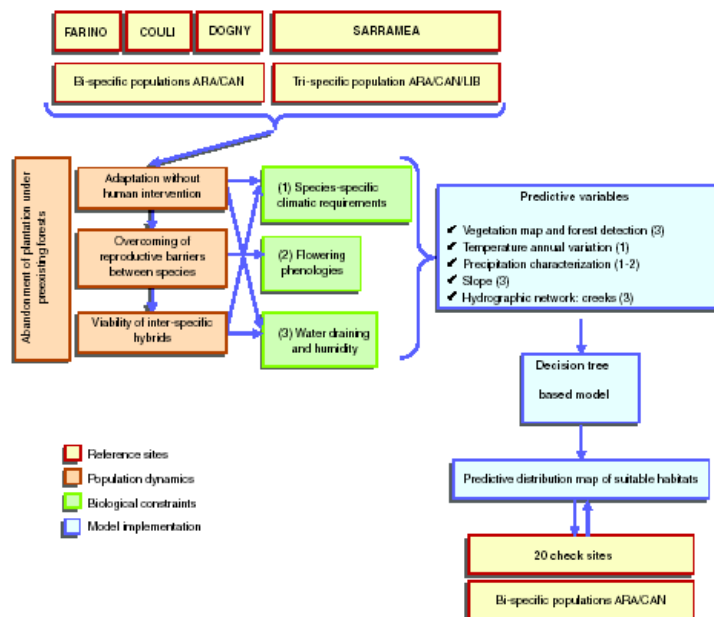


Fig. 2. Flow chart of major steps in generating the model. Among the four reference sites used in the model, only the Saramea site had three coffee species. ARA, *Coffea arabica*; CAN, *C. canephora* and LIB, *C. liberica*. Based on the population dynamics observed and biological constraints, five predictive variables were computed to be implemented in the decision tree.

was implemented on an extraction of the whole study area in order to determine the potential of the method.

Climate characterization

Temperature and precipitation data collected over an 11-year period (January 1996 to March 2007) were obtained from four Meteo France study sites flanking the area of concern: Canala, Col d'Amieu, La Foa and Pocquereux (see Fig. S1 in Supporting Information). Using a regularized spline method (ESRI® ARCMAP™ 9.2), monthly precipitation and temperature data were interpolated between the four sites. This method gives especially good results when only a few points are used for interpolation (Mitas & Mitasova 1988).

The four reference point values were also used to construct climograms based on monthly average rainfalls (x axis) and temperatures (y axis). These graphs were compared to those derived from African regions where the coffee species spontaneously grow: Jimma in Ethiopia for *C. arabica* and Divo in Côte d'Ivoire for *C. canephora* and *C. liberica*.

Slope and the temporary hydrographic network

The optimal coffee hydric equilibrium is acquired with adequate surrounding humidity and good water drainage (Giordanengo 1992). A favourable environment is thus related to the presence of a temporary hydrographic network, i.e. a creek and a land inclination.

The digital elevation model (DEM) of the study area was obtained with a pixel size of 50 m and used to extract the slope and hydrographic network data based on the flow accumulation data. The slope and the flow accumulation were calculated from the DEM with the Hydrology Spatial Analyst Tools in ESRI® ArcGIS™ 9.2. Finally, a map representing the distances to the creeks, so called buffers, was constructed based on the flow accumulation.

Forest distribution, land cover

Most of the abandoned plantations were coffee crops grown under shade trees. Two Quickbird satellite images (acquired on the 18/06/05 for the East zone and on the 16/08/05 for the West zone) were obtained for the study area. The panchromatic and multispectral images were merged with the numerical multiplicative method (Erdas Imagine software), which involves a simple multiplicative algorithm that integrates two raster images. The resulting image was analysed in order to map the forest distribution, thus facilitating implementation of the predictive distribution model for shade grown coffee.

Model generation

An expert model decision tree was constructed using logical combinations of selected parameters as explained above. It is a robust predictive model as it does not have any statistical errors in prediction and represents the best way to explore an ecological situation without admeasured predictions (Zhang, Valentine & Kemp 2005). The hierarchy and weight of the parameters were based on the biological and ecological knowledge of coffee trees and environmental data obtained at the four reference sites. Each predictor variable was coded from very favourable (coded 1) to unfavourable (coded 3). Then all modalities were combined in a decision tree to obtain the susceptibility result map. The final model combined:

- a temperature and precipitation prediction;
- a map of the slope and one of the hydrographic network buffers;
- a forest distribution map.

Finally, each combination code was linked to the modalities, i.e. favourable or unfavourable. All indices were resampled and homogenized to fit with the 50 m spatial resolution of the DEM.

Model validation

The predictive distribution of favourable areas was tested using 20 additional sites distributed throughout the study area. A field survey of these sites was carried out to verify the presence of *C. arabica*, *C. canephora*, and inter-specific hybrids. The model accuracy was assessed by examining the proportion of check points falling into regions predicted as being favourable.

Results

MOLECULAR MARKER EFFICIENCY AND IMPORTANCE OF THE AFRICAN REFERENCE GROUPS

For the four African groups, high genetic polymorphism (N_a , H_o and H_e) was observed at the 16 SSR and 10 EST-SSR loci (see Table S1 in Supporting Information). Of the 54 reference individuals analysed, a total of 244 alleles across all loci were detected, giving an average of 9.38 alleles per locus, ranging from 2 for M804 to 17 for Es62. Furthermore, the microsatellite markers revealed a high intraspecific polymorphism level, with gene diversity (H_e) and PIC values of 0.63 and 0.61, 0.41 and 0.44 and 0.50 and 0.47 for *C. canephora*, *C. liberica* ssp. *liberica* and *C. liberica* ssp. *dewevrei* respectively (Table 1). EST-SSRs showed lower H_e and PIC values of 0.45 and 0.42, 0.33 and 0.3 and 0.30 and 0.26 respectively. When combining both marker types, the reference samples appeared to be highly polymorphic, except for *C. arabica*, which was represented by only two individuals.

The African species and subspecies were well differentiated, as suggested by the high pairwise F_{ST} values, ranging from 0.26 between *C. liberica* ssp. *liberica* and *C. liberica* ssp. *dewevrei* to 0.41 between *C. arabica* and *C. liberica* ssp. *dewevrei* (Table 2). Furthermore, each species/sub-species was distinguished by its own private alleles. More precisely, out of the 244 scored alleles, eight (five SSR alleles and three EST-SSR alleles) were only present in *C. arabica*, 73 (48 SSR and 25 EST-SSR) in *C. canephora*, 21 (13 SSR and eight EST-SSR) in *C. liberica* ssp. *liberica* and 24 (17 SSR and seven EST-SSR) in *C. liberica* ssp. *dewevrei*. Allele range distribution at each locus varied according to the species considered. All EST-SSRs and 75% of the SSRs presented *C. canephora* specific alleles. For the other species, the number of loci with private alleles was lower, with 30%, 30% and 40% EST-SSRs and 25%, 43.8% and 56.3% SSRs revealing *C. arabica*, *C. liberica* ssp. *liberica* and *C. liberica* ssp. *dewevrei* specific alleles respectively.

GENETIC STRUCTURE OF THE SARRAMEA COFFEA POPULATION

Analysis of the 95 Sarramea individuals revealed a total of 186 alleles for both SSR and EST-SSR loci, giving an average of 8.75 and 5.90 alleles per locus for SSRs and EST-SSRs respec-

Table 1. Intraspecific variability in the four African samples

	SSR					EST-SSR				
	N_a	N_{pa}	H_o	H_E	PIC	N_a	N_{pa}	H_o	H_E	PIC
<i>C. canephora</i> ($N = 34$)	126	48	0.42	0.63	0.61	49	25	0.34	0.45	0.42
<i>C. liberica</i> ssp. <i>liberica</i> ($N = 8$)	62	13	0.45	0.41	0.44	27	8	0.27	0.33	0.3
<i>C. liberica</i> ssp. <i>dewevrei</i> ($N = 10$)	70	17	0.39	0.5	0.47	25	7	0.25	0.3	0.26
<i>C. arabica</i> ($N = 2$)	29	5	0.75	N	N	21	3	1	N	N

N_a , number of alleles; N_{pa} , number of private alleles; H_o , observed heterozygosity; H_E , gene diversity or expected heterozygosity; PIC, polymorphism information content; N, not available.

Table 2. Multilocus estimates of F_{ST} calculated for pairs of reference species for the 26 investigated microsatellites

Pair-wise F_{ST}	<i>C. canephora</i>	<i>C. liberica</i> ssp. <i>dewevrei</i>	<i>C. liberica</i> ssp. <i>liberica</i>
<i>C. arabica</i>	Na	Na	Na
<i>C. canephora</i>		0.40197*	0.39615*
<i>C. liberica</i> ssp. <i>dewevrei</i>			0.26644*

* $P < 0.01$.

Na, not available.

tively. Of the 186 alleles detected, 11 were not present in the African reference samples. The PIC values and gene diversity were also high, with an average of 0.66 and 0.69, respectively, for SSR loci and 0.53 and 0.57 for EST-SSR loci. The discrepancy between the level of polymorphism revealed by EST-SSRs and SSRs resulted from the selection of a higher proportion of species-specific markers among the EST-SSRs [see Poncet *et al.* (2006) for further comparison of the two types of markers].

Assignment tests using GENECASS were performed to assess genotypic similarities of the Saramea plants to the African reference samples. Twenty individuals could not be assigned, while 75 individuals were assigned to a species with a probability of > 0.2 . This analysis allowed identification of 58 *C. canephora*, four *C. arabica* and 13 *C. liberica* ssp. *liberica*. No individuals were assigned to *C. liberica* ssp. *dewevrei*.

The multilocus analysis on the whole set of individuals (i.e. Saramea and African individuals) implemented in STRUCTURE was run with $K = 4$. This number of clusters was validated by the formula proposed by Garnier *et al.* (2004): $\ln P(D)_{K-4} - \ln P(D)_{K-3} = 343.6 > (\ln P(D)_{K-5} - \ln P(D)_{K-4} = 161.6$. The clusters were identified according to their African members. One cluster represented *C. liberica* ssp. *liberica*, one *C. arabica*, and finally two clusters represented *C. canephora*, which is known to be highly diversified (Dussert *et al.* 1999). Sixteen additional individuals that had not been previously assigned with GENECASS2 were found to belong to the *C. arabica* (4), *C. liberica* ssp. *liberica* (4) and *C. canephora* (9) clusters. Three inter-specific hybrids were detected with intermediate positions on the trispecific plot obtained by pooling the two *C. canephora* clusters into one (see Fig. S2 in Supporting

Information). These three genotypes were identified as admixed with over 20% ancestry from other groups and derived from backcrosses: 66% *C. arabica*/34% *C. canephora*, 32% *C. liberica* ssp. *liberica*/68% *C. canephora* and 28% *C. liberica* ssp. *liberica*/72% *C. canephora*. A fourth individual had admixture proportions of less than 20%: 88% *C. liberica* ssp. *liberica*/12% *C. canephora*. These results confirmed the assignment analysis implemented in GENECASS2, showing no assignment value for these four putative inter-specific hybrids.

Finally the Saramea sample was composed of 67 *C. canephora*, four *C. arabica*, 17 *C. liberica* ssp. *liberica* individuals and three putative inter-specific hybrids (3% hybridization) (Fig. 1).

In conclusion, the Saramea site appeared to be suitable for species coexistence and hybridization and corroborated the phenomenon also observed at the three other reference sites: Farino, Couli, and Dogny (Le Pierres 1999; Mahe *et al.* 2007).

ENVIRONMENTAL PARAMETER FOR PREDICTIVE VARIABLES

Four climate predictive variables

The study area presents climatic conditions that are favourable for the coexistence of the three species. Indeed, with annual temperature differences between the coolest and the warmest month of over 7 °C at each reference site (8.11 °C at Farino, 7.59 °C at Dogny, 7.70 °C at Couli and 7.77 °C at Saramea), the temperature ranges overlapped those required for *C. liberica* and *C. canephora* (Divo data) and *C. arabica* (Jimma data) (Fig. 3). Similarly, the precipitation variations observed in the New Caledonian region overlapped those of the African sites.

Conversely, dissimilarities in terms of rain distribution throughout the year are likely to be involved in deregulation of each species' phenology, thus weakening the effectiveness of this barrier to hybridization. For example, five to six flowering events were observed in 2008 for the Saramea species with the overlapping of fertile stages for successive species flowerings (data not shown).

Contrary to the African conditions, no complete dry season was observed at the reference sites. Precipitation diminishes markedly between March and June, corresponding to a dry season punctuated with a few minor rainfalls. The subsequent

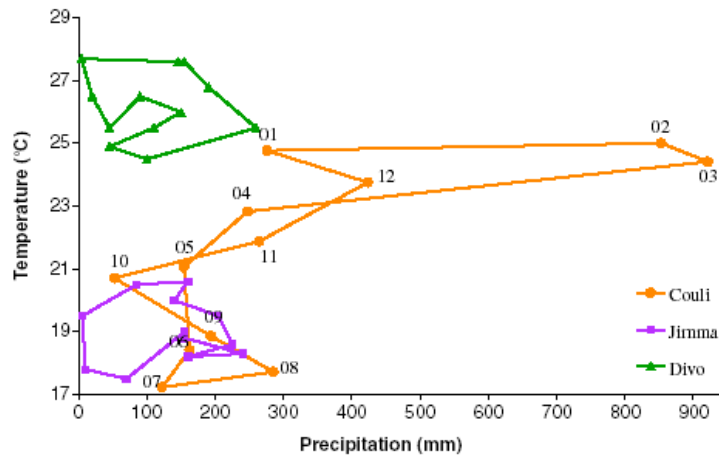


Fig. 3. Climograms based on temperature and precipitation measured in Ethiopia (Jimma) for the *Coffea arabica* reference, in Côte d'Ivoire (Divo) for both *C. liberica* and *C. canephora* references and in New Caledonia (Couli). Month numbers (January = 1 to December = 12) are reported on the Couli climogram.

increase in precipitation observed from June to August is required to activate the flowering process. This phenomenon is revealed by the difference between two successive monthly precipitations. At our reference sites, it was more marked from July to August (85 mm more in August on average) with an increase of 57 mm at Farino, 23 mm at Dogny, 163 mm at Couli and 97 mm at Sarraamea.

Differences in total annual rainfall were also noted with higher values in the Caledonian area sites. Rainfall totals of 2368, 1978, 3954 and 3116 mm were recorded at Farino, Dogny, Couli and Sarraamea, respectively, while 1314 and 1610 mm were noted at Divo and Jimma respectively.

Finally four parameters were selected and coded to be integrated in the model:

- Temp-delta: the temperature difference between the coolest and warmest month: values higher than 7 °C (coded 1) facilitate the species coexistence, while values lower than 7 °C (coded 2) do not.
- Rain-delta: at least one precipitation increase between June and August following a decrease between March and June (coded 1 or 2).
- Rain-delta-quantity: the higher value of the intermonthly variation coded 1 if more than 97 mm and 2 if less.
- Rain-total: the annual quantity of rain coded 1 if more than 2200 mm, 2 if ranging from 1200 to 2200 mm, and 3 if less than 1200 mm.

The unweighted rain parameters were combined to construct a decision rule for precipitation. For each combination, a sensitivity code ranging from 1 (very favourable) to 6 (not favourable), was attributed.

The slope and the distance to the temporary hydrographic network as physical predictive variables

It appeared that the suitable areas, represented by the four reference sites, were located on quite steep slopes and next to a creek. For instance, the Sarraamea coffee population is located on a slope of at least 30% with a maximum distance of 100 m from a creek. Each DEM pixel was classified in a distance cate-

gory: 50 m (coded 1), 100 m (coded 2) and over 200 m (coded 3). The slope parameter was coded 1 if > 30%, 2 if ranging from 10 to 30% and 3 if < 10%.

Distribution of forests as a remotely sensed predictive variable

Based on vegetation indices, the vegetation was easily discriminated. It was coded 1 if there was forest and 2 if not (different vegetation types or no vegetation).

SPATIAL ANALYSIS MODEL

Hierarchy and combination of predictor variables

Building an expert model like a tree decision model involves determining the order of the predictor variables. The most constraining parameter in ecological terms, i.e. forest distribution, was first implemented. Indeed, hybridization between coffee species only happens in coffee plantations grown under shade trees. Terrain parameters, including distance to a creek and slope, were then implemented. Finally, the climatic parameters were included, including the annual temperature gap and the precipitation sensitivity code, according to the parameters presented before, i.e. Temp-delta, Rain-delta, Rain-delta-quantity and Rain-total (Table S2).

Predictive areas

The final model combined five unweighted predictor variables (slope, distance from a creek, annual temperature gap, precipitation and forest presence). The decision tree gave the parameter combinations, the calculated sensitivities ranging from 5 to 15. For example, in a forest with a creek located at less than 100 m, a slope of at least 30%, an annual temperature gap of at least 7 °C and a precipitation sensitivity of 1, a sensitivity to hybridization events of 5 was obtained. The susceptibility map of sympatric sites present is given in Fig. 4. In the study area, 63% was covered with forest, 60% of which

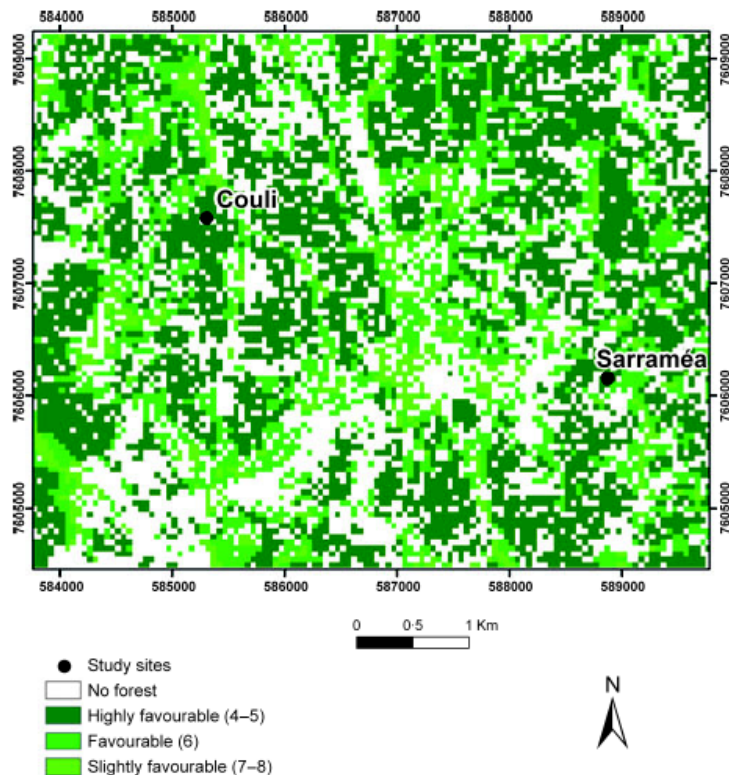


Fig. 4. Suitability map produced for the Sarramea and Couli area according to the predictive model based on coffee environmental expertise.

appeared to be favourable for coffee tree sympatric populations and hybridization events.

Spatial distribution validation

Finally, the overall accuracy was high. Seventy-nine per cent of the check sites were located in either highly favourable areas (74%) or favourable areas (5%). Only 21% of the check points were located in no forest zones. This mislocation was probably an artefact caused by the low spatial resolution of the model imposed by the DEM at 50 m, i.e. there was probably a forest cover underestimation due to the reduction in image resolution when overlapping the Quickbird image (0.70 m resolution).

Discussion

HIGH LEVEL OF INTRODUCED *COFFEA* DIVERSITY

Plant species could become established in a new range of environments according to different scenarios depending both on the introduction history and prevailing evolutionary forces. Several factors could influence the gain or loss of genetic diversity between introduced and source populations (e.g. Durka *et al.* 2005; Benkman *et al.* 2008). For instance, Novak, Mack & Soltis (1993) observed a reduction in genetic variability across *Bromus tectorum* populations in North America but an increase in genetic variation within each population compared to native (Eurasian and African) populations. This result first

suggested a founder effect involving an overall bottleneck, but also many introduction events reinforcing within-population genetic diversity. After multiple introductions, a high amount of diversity is established in the new populations, thus increasing the adaptation potential and saving the species from a bottleneck effect. Consequently, multiple introductions could contribute to the adaptation and also the potential invasion success of a species. Similarly, Durka *et al.* (2005) demonstrated that the invasion success of self-compatible garlic mustard *Alliaria petiolata* in North America was not the consequence of reduced genetic variability but of a substantial molecular genetic variation within and among the introduced populations.

Coffee species have been introduced all over the world, where the environmental conditions are *a priori* suitable for their cultivation. However, each introduction event differs in terms of the genetic variation introduced. For instance, Prakash *et al.* (2005) studied the Indian robusta (*C. canephora*) gene pool to assess its breeding potential. Comparative analysis revealed the higher diversity in representatives from a core collection than in the Indian gene pool. On the contrary, our results from the New Caledonian population revealed a high level of diversity, which could be explained particularly by the introduction of polymorphic founder individuals from three species. Indeed, compared to the genetic diversity naturally existing in Africa and represented by reference samples of the three species, the introduced diversity planted in the studied site was not significantly different. In particular, at the

intraspecific level, the introduced *C. canephora* accessions represented a significant part of the diversity described by Gomez *et al.* (2009) for three of the four African sub-groups (data not shown).

INTER-SPECIFIC HYBRIDIZATION IN A NEW ENVIRONMENT

The introduction of a species in a new environment where formerly interacting organisms and climatic conditions are modified could have considerable impacts on both the species diversity and the community structure (Benkman *et al.* 2008). Moreover, when related species are found in a common environment, new interactions like hybridization could occur, resulting occasionally in gene flow (e.g. Abbott *et al.* 2003). Some studies have focused on hybridization between introduced and native species (e.g. Abbott *et al.* 2003; Riley *et al.* 2003; Fortune *et al.* 2008). However, very few studies have been undertaken, to our knowledge, on actual hybridization between two introduced species previously isolated from each other. In our study, the three coffee species, i.e. *C. canephora*, *C. liberica*, and *C. canephora*, are geographically or ecologically isolated from each other in Africa, and their co-existence in the same population has made their survival and hybridization even more challenging and exceptional.

Inter-specific gene flows are currently well studied and provide a better understanding of the hybridization levels in relation to ecological factors. For instance, Gow, Peichel & Taylor (2006) found different hybridization rates between three-spined sticklebacks *Gasterosteus aculeatus* according to sampling locations and phenotypic adaptation. Furthermore, Milne *et al.* (1999) suggested that inter-specific gene flows might be favoured by human habitat disturbance, while natural habitats provide effective barriers to hybridization. In New Caledonia, evidence was obtained on gene flow between recently introduced coffee species, with over 3% inter-specific hybridization detected. These species were intentionally introduced for cultivation and the chosen habitats had been initially modified by human activities. Since their abandonment, natural conditions have allowed both the preservation of species integrities and mixed reproduction. Indeed, the population dynamics are enhanced through both inter-specific and intra-specific mating. Although initially coffee cropping in the valley was relatively widespread, the environmental heterogeneity generated variations in coffee tree abundance and fitness among the habitats, with optimal adaption only in suitable environmental niches.

FAVOURABLE ECOLOGICAL FACTORS

To understand the climatic requirements in terms of favourable factors for adaptation is an important goal in resource conservation management. As forest trees are generally not domesticated and are perennial, studying their adaptation dynamics through the genetic potentials and the ecological context enables us to gain insight into the long-term implications.

For coffee trees, an environmental expertise was developed to characterize the topography and climate as key factors for the development, growth and breeding of reference populations. In particular, a comparison of climograms in an original environmental range of species with the climograms of the area of interest in New Caledonia confirmed the favourable climatic features in the Sarramea zone. The annual temperature and rainfall cycle fulfils the three coffee species demands and allows their coexistence. The precipitation prediction variables were specifically defined according to the potential effect on the flowering phenologies (Godoy *et al.* 2009).

A PREDICTIVE DISTRIBUTION OF SUITABLE ZONES

Geographic information systems and mathematical models have been jointly used when introducing spatial differentiation in the habitat distribution with a predictive distribution model (Guisan & Zimmermann 2000). Logistic regression or generalized linear models are commonly used as mathematical models (Gavashelishvili & Lukarevskiy 2008; Kaczensky *et al.* 2008; Peltzer, Ferriss & FitzJohn 2008). Despite their performance, these parametric models have two main drawbacks: (i) if some assumptions are violated, it is difficult to extrapolate the significance of the results and (ii) if the sample size is too small, significance tests could be impacted (Vayssières, Plant & Allen-Diaz 2000).

The model used in this study was entirely based on environmental expertise, and no statistical uncertainty on parameter combinations was introduced. The variables were selected according to knowledge of native species–habitat associations and agronomy, as well as on the basis of field experience at the reference sites in the Sarramea region. The habitat distribution was investigated through a classification tree. These methods have applications in various fields such as plant pathology, wildlife management or landscape classification (e.g. Zhang *et al.* 2006; Pesch *et al.* 2008). Free of distributional assumptions, Vayssières *et al.* (2000) found that models based on classification trees were significantly more efficient than mathematical ones (logistic regression models) in most cases, in terms of intuitive, easy to develop and interpret, and accurate tools. This method was well suited to fulfilling our objectives and the overall 79% prediction accuracy confirmed the importance of focusing on this region to study coffee species interactions. This study was the first attempt to model coffee habitats in connection with natural hybridization, giving a first quantitative estimation of habitat suitability for coffee multi-species sympatry. Our model provided a preliminary tool for identification of potential hybrid zones. The interaction factor has been studied with a classification tree and insight gained from this model will help us to further develop an improved parametric model in the future.

Implications

Our study provides a good base for biological and ecological knowledge about coffee trees in New Caledonia, an area known to be an important biodiversity hotspot (Myers *et al.*

2000). Like many other tropical islands, its environmental patterns are very complex and forests of the Sarramea region are characterized by a relatively high diversity index (Jaffré & Veillon 1995). Thanks to the integration of biological and ecological knowledge about coffee trees and environmental conditions in the study area, the potential spatial distribution of favourable niches (i.e. areas influenced by the same environmental factors) was proposed. This zone is of exceptional significance as no similar phenomenon of spontaneous inter-specific hybridization has been noted elsewhere in the world, except for a rare event observed in Timor leading to the so-called 'Timor hybrid' (Cramer 1957). The predictive model that we implemented allowed us to check the application of such a combination of parameters, and was confirmed with the high proportion of check sites falling into regions predicted as being favourable. Furthermore, the prediction ability was stable over time. Indeed, contrary to weedy or fast growing species, perennial species (Elliott *et al.* 2004; Peltzer *et al.* 2008) have few inter-annual variations in abundance or seed production. So the predictive distribution obtained represents long-term information and allows long-term conservation and management over decades.

Similar situations could be encountered elsewhere in the world where the environmental ranges are complex and the biodiversity rich. For example, such an approach could be used to predict the habitat suitability of naturalized plants or spread of invasive species in response to the environment, and might help in understanding niche dynamics in the climate change context (Pearman *et al.* 2008).

Acknowledgements

The authors gratefully acknowledge the financial support of the DDR (Direction du Développement Rural), Province Sud of New Caledonia (Convention 531-PVF/DDR).

References

- Abbott, R.J., James, J.K., Milne, R.I. & Gillies, A.C. (2003) Plant introductions, hybridization and gene flow. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **358**, 1123–1132.
- Adams, J.R., Lucash, C., Schutte, L. & Waits, L.P. (2007) Locating hybrid individuals in the red wolf (*Canis rufus*) experimental population area using a spatially targeted sampling strategy and faecal DNA genotyping. *Molecular Ecology*, **16**, 1823–1834.
- Anderson, R.P., Gomez-Laverde, M. & Peterson, A.T. (2002) Geographical distributions of spiny pocket mice in South America: insights from predictive models. *Global Ecology and Biogeography*, **11**, 131–141.
- Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. & Bonhomme, F. (1996–2004) *GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations*. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier, France.
- Benkman, C.W., Siepielski, A.M. & Parchman, T.L. (2008) The local introduction of strongly interacting species and the loss of geographic variation in species and species interactions. *Molecular Ecology*, **17**, 395–404.
- Booth, T.H. & Ekeleme, F. (2002) Identifying climatically suitable areas for growing particular trees in Africa: an example using *Grevillea robusta*. *Agroforestry Systems*, **54**, 41–49.
- Booth, T.H., Jovanovic, T., Old, K.M. & Dudzinski, M.J. (2000) Climatic mapping to identify high-risk areas for *Cylindrocylindrium quinquevittatum* leaf blight on eucalypts in mainland South East Asia and around the world. *Environmental Pollution*, **108**, 365–372.
- Chevalier, A. (1946) Ecologie et distribution géographique des caféiers sauvages et cultivés. *Revue de Botanique Appliquée d'Agriculture Tropicale*, **26**, 81–94.
- Coulibaly, I., Revol, B., Noirot, M., Poncet, V., Lorieux, M., Carasco-Lacombe, C., Minier, J., Dufour, M. & Hamon, P. (2003) AFLP and SSR polymorphism in a *Coffea* interspecific backcross progeny [(*C. canephora* x *C. heterocalyx*) x *C. canephora*]. *Theoretical and Applied Genetics*, **107**, 1148–1155.
- Cramer, P.J.S. (1957) *A Review of Literature of Coffee Research in Indonesia*. SIC Editorial, Inter American Institute of Agricultural Sciences, Turrialba, Costa Rica.
- Cubry, P., Musoli, P., Legnate, H., Pot, D., de Bellis, F., Poncet, V., Anthony, F., Dufour, M. & Leroy, T. (2008) Diversity in coffee using SSR markers structure of the *Coffea* genus and perspectives for breeding. *Genome*, **51**, 50–63.
- Davis, A.P., Govaerts, R., Bridson, D.M. & Stoffelen, P. (2006) An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). *Botanical Journal of the Linnean Society*, **152**, 465–512.
- Durka, W., Bossdorf, O., Prati, D. & Auge, H. (2005) Molecular evidence for multiple introductions of garlic mustard (*Alliaria petiolata*, Brassicaceae) to North America. *Molecular Ecology*, **14**, 1697–1706.
- Dussert, S., Lashermes, P., Anthony, F., Montagnon, C., Trouslot, P., Combes, M.C., Berthaud, J., Noirot, M. & Hamon, S. (1999) *Coffea canephora*. Diversité génétique des plantes tropicales cultivées (eds P. Hamon, M. Seguin, X. Perrier & J.-C. Glaszmann), pp. 175–794. CIRAD, Montpellier, France.
- Elliott, L.J., Mason, D.C., Wilkinson, M.J., Allainguillaume, J., Norris, C., Alexander, M. & Walters, R. (2004) The role of satellite image-processing for national-scale estimates of gene flow from genetically modified crops: rapeseed in the UK as a model. *Journal of Applied Ecology*, **41**, 1174–1184.
- Ferguson, M.E., Jarvis, A., Stalker, H.T., Williams, D.E., Guarino, L., Valls, J.F.M., Pittman, R.N., Simpson, C.E. & Bramel, P.J. (2005) Biogeography of wild *Anacis* (Leguminosae): distribution and environmental characterisation. *Biodiversity and Conservation*, **14**, 1777–1798.
- Fortune, P.M., Schierenbeck, K., Ayres, D., Bortolus, A., Catrice, O., Brown, S. & Ainouche, M.L. (2008) The enigmatic invasive *Spartina densiflora*: a history of hybridizations in a polyploidy context. *Molecular Ecology*, **17**, 4304–4316.
- Gamier, S., Alibert, P., Audiot, P., Prieur, B. & Rasplus, J.Y. (2004) Isolation by distance and sharp discontinuities in gene frequencies: implications for the phylogeography of an alpine insect species, *Carabus solieri*. *Molecular Ecology*, **13**, 1883–1897.
- Gavashelishvili, A. & Lukarevskiy, V. (2008) Modelling the habitat requirements of leopard *Panthera pardus* in west and central Asia. *Journal of Applied Ecology*, **45**, 579–588.
- Gerard, P.R., Klein, E.K., Austerlitz, F., Fernandez-Manjarres, J.F. & Frascaria-Lacoste, N. (2006) Assortative mating and differential male mating success in an ash hybrid zone population. *BMC Evolutionary Biology*, **6**, 96.
- Giordanengo, P. (1992) *Biologie, éco-éthologie et dynamique des populations du scolyte des grains de café, Hypothenemus hampei Ferr. (Coleoptera, Scolytidae)*, en Nouvelle-Calédonie. Thèse Doctorale, Université de Rennes 1, Rennes, France.
- Godoy, O., Richardson, D.M., Valladares, F. & Castro-Diez, P. (2009) Flowering phenology of invasive alien plant species compared with native species in three Mediterranean-type ecosystems. *Annals of Botany*, **103**, 485–494.
- Gomez, C., Dussert, S., Hamon, P., Hamon, S., de Kochko, A. & Poncet, V. (2009) Current genetic differentiation of *Coffea canephora* Pierre ex A. Froehn in the Guineo-Congolian African zone: cumulative impact of ancient climatic changes and recent human activities. *BMC Evolutionary Biology*, **9**, 167.
- Gow, J.L., Peichel, C.L. & Taylor, E.B. (2006) Contrasting hybridization rates between sympatric three-spined sticklebacks highlight the fragility of reproductive barriers between evolutionarily young species. *Molecular Ecology*, **15**, 739–752.
- Guarino, L., Jarvis, A., Hijmans, R.J. & Maxted, N. (2001) Geographic Information Systems (GIS) and the conservation and use of plant genetic resources. *International Conference on Science and Technology for Managing Plant Genetic Diversity in the 21st century (SAT21)*. Kuala Lumpur, Malaysia.
- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Jaffré, T. & Veillon, J.-M. (1995) Structural and floristic characteristics of a rain forest on schist in New Caledonia: a comparison with an ultramafic rain forest. *Bulletin du Muséum national d'Histoire naturelle. Section B, Adansonia*, **3–4**, 201–226.
- Jarvis, A., Williams, K., Williams, D., Guarino, L., Caballero, P.J. & Mottram, G. (2005) Use of GIS for optimizing a collecting mission for a rare wild pepper (*Capsicum flexuosum* Sendtn.) in Paraguay. *Genetic Resources and Crop Evolution*, **52**, 671–682.
- Kaczynski, P., Ganbaatar, O., von Wehrden, H. & Walzer, C. (2008) Resource selection by sympatric wild equids in the Mongolian Gobi. *Journal of Applied Ecology*, **45**, 1762–1769.

- Ky, C.L., Barre, P., Lorieux, M., Trouslot, P., Akaffou, S., Louarn, J., Charrier, A., Hamon, S. & Noirot, M. (2000) Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). *Theoretical and Applied Genetics*, **101**, 669–676.
- Lamont, B.B., He, T., Enright, N.J., Krauss, S.L. & Miller, B.P. (2003) Anthropogenic disturbance promotes hybridization between *Banksia* species by altering their biology. *Journal of Evolutionary Biology*, **16**, 551–557.
- Le Pierrès, D. (1995) *Etude des hybrides interspécifiques tétraploïdes de première génération entre Coffea arabica L. et les caféiers diploïdes*. Thèse Doctorale, Université Paris XI, Orsay, France.
- Le Pierrès, D. (1999) *Etude des hybrides naturels entre Coffea arabica et Coffea canephora de Nouvelle-Calédonie*. Field report (IRD) Nouméa.
- Lexer, C., Fay, M.F., Joseph, J.A., Nica, M.S. & Heinze, B. (2005) Barrier to gene flow between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. *Molecular Ecology*, **14**, 1045–1057.
- Liu, K. & Muse, S.V. (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, **21**, 2128–2129.
- Mahe, L., Le Pierrès, D., Combes, M.C. & Lashermes, P. (2007) Introgressive hybridization between the allotetraploid *Coffea arabica* and one of its diploid ancestors, *Coffea canephora*, in an exceptional sympatric zone in New Caledonia. *Genome*, **50**, 316–324.
- Mallet, J. (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, **20**, 229–237.
- Mallet, J., Beltran, M., Neukirchen, W. & Linares, M. (2007) Natural hybridization in heliconine butterflies: the species boundary as a continuum. *BMC Evolutionary Biology*, **7**, 28.
- Milne, R.I., Abbott, R.J., Wolff, K. & Chamberlain, D.F. (1999) Hybridization among sympatric species of *Rhododendron* (Ericaceae) in Turkey: morphological and molecular evidence. *American Journal of Botany*, **86**, 1776–1785.
- Mitas, L. & Mitasova, H. (1988) General variational approach to the interpolation problem. *Computers & Mathematics with Applications*, **16**, 983–992.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.
- N' Diaye, A., Poncet, V., Louarn, J., Hamon, S. & Noirot, M. (2005) Genetic differentiation between *Coffea liberica* var. *liberica* and *C. liberica* var. *devevrei* and comparison with *C. canephora*. *Plant Systematics and Evolution*, **253**, 95–104.
- N' Diaye, A., Noirot, M., Hamon, S. & Poncet, V. (2007) Genetic basis of species differentiation between *Coffea liberica* Hiern and *C. canephora* Pierre: analysis of an interspecific cross. *Genetic Resources and Crop Evolution*, **54**, 1011–1021.
- Novak, S.J., Mack, R.N. & Soltis, P.S. (1993) Genetic-variation in *Bromus*-*Tectorum* (Poaceae) – introduction dynamics in North-America. *Canadian Journal of Botany-Revue Canadienne De Botanique*, **71**, 1441–1448.
- Pearman, P.B., Guisan, A., Broennimann, O. & Randin, C.F. (2008) Niche dynamics in space and time. *Trends in Ecology & Evolution*, **23**, 149–158.
- Peltzer, D.A., Ferriss, S. & FitzJohn, R.G. (2008) Predicting weed distribution at the landscape scale: using naturalized *Brassica* as a model system. *Journal of Applied Ecology*, **45**, 467–475.
- Pesch, R., Pehlke, H., Jerosch, K., Schroder, W. & Schluter, M. (2008) Using decision trees to predict benthic communities within and near the German Exclusive Economic Zone (EEZ) of the North Sea. *Environmental Monitoring and Assessment*, **136**, 313–325.
- Piry, S., Alapetite, A., Cornuet, J.M., Paetkau, D., Baudouin, L. & Estoup, A. (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, **95**, 536–539.
- Poncet, V., Hamon, P., Minier, J., Carasco-Lacombe, C., Hamon, S. & Noirot, M. (2004) SSR cross-amplification and variation within coffee trees (*Coffea* spp.). *Genome*, **47**, 1071–1081.
- Poncet, V., Rondeau, M., Tranchant, C., Cayrel, A., Hamon, S., de Kochko, A. & Hamon, P. (2006) SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. *Molecular Genetics and Genomics*, **276**, 436–449.
- Poncet, V., Dufour, M., Hamon, P., Hamon, S., de Kochko, A. & Leroy, T. (2007) Development of genomic microsatellite markers in *Coffea canephora* and their transferability to other coffee species. *Genome*, **50**, 1156–1161.
- Prakash, N.S., Combes, M.-C., Dussert, S.P., Naveen, S. & Lashermes, P. (2005) Analysis of genetic diversity in Indian robusta coffee gene pool (*Coffea canephora*) in comparison with a representative core collection using SSRs and AFLPs. *Genetic Resources and Crop Evolution*, **52**, 333–343.
- Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Rannala, B. & Mountain, J.L. (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 9197–9201.
- Rieseberg, L.H. & Willis, J.H. (2007) Plant speciation. *Science*, **317**, 910–914.
- Riley, S.P.D., Shaffer, H.B., Voss, S.R. & Fitzpatrick, B.M. (2003) Hybridization between a rare, native tiger salamander (*Ambystoma californiense*) and its introduced congener. *Ecological Applications*, **13**, 1263–1275.
- Swenson, N.G. (2008) The past and future influence of geographic information systems on hybrid zone, phylogeographic and speciation research. *Journal of Evolutionary Biology*, **21**, 421–434.
- Thuiller, W., Albert, C., Araujo, M.B., Berry, P.M., Cabeza, M., Guisan, A., Hickler, T., Midgely, G.F., Paterson, J., Schurr, F.M., Sykes, M.T. & Zimmermann, N.E. (2008) Predicting global change impacts on plant species' distributions: future challenges. *Perspectives in Plant Ecology Evolution and Systematics*, **9**, 137–152.
- Tiranti, B. & Negri, V. (2007) Selective microenvironmental effects play a role in shaping genetic diversity and structure in a *Phaseolus vulgaris* L. landrace: implications for on-farm conservation. *Molecular Ecology*, **16**, 4942–4955.
- Valbuena-Carabana, M., Gonzalez-Martinez, S.C., Sork, V.L., Collada, C., Soto, A., Goicoechea, P.G. & Gil, L. (2005) Gene flow and hybridization in a mixed oak forest (*Quercus pyrenaica* Willd. and *Quercus petraea* (Matt.) Liebl.) in central Spain. *Heredity*, **95**, 457–465.
- Vayssières, M.P., Plant, R.E. & Allen-Diaz, B.H. (2000) Classification trees: an alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science*, **11**, 679–694.
- Wilkinson, M.J., Elliott, L.J., Allainguillaume, J., Shaw, M.W., Norris, C., Welters, R., Alexander, M., Sweet, J. & Mason, D.C. (2003) Hybridization between *Brassica napus* and *B. rapa* on a national scale in the United Kingdom. *Science*, **302**, 457–459.
- Zhang, B.S., Valentine, I. & Kemp, P.D. (2005) A decision tree approach modelling functional group abundance in a pasture ecosystem. *Agriculture Ecosystems & Environment*, **110**, 279–288.
- Zhang, B.S., Valentine, I., Kemp, P. & Lambert, G. (2006) Predictive modelling of hill-pasture productivity: integration of a decision tree and a geographical information system. *Agricultural Systems*, **87**, 1–17.

Received 13 March 2009; accepted 30 November 2009
Handling Editor: Rob Freckleton

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Coffee tree species introduction and hybridization in New Caledonia.

Fig. S1. Map of the study area in New Caledonia.

Fig. S2. Trispecific plot representing genetic admixtures.

Table S1. Diversity indices in the African reference samples.

Table S2. Decision tree model.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Appendix S1. Coffee tree species introduction and hybridization in New Caledonia

Following human migrations or colonisations, coffee trees were introduced from their native tropical areas to others. In some cases, plantations were established in marginal areas such as isolated islands like Timor, Hawaii or New Caledonia. The climate is modified on these islands due to the proximity of the sea, thus enabling co-cultivation of *C. canephora* and *C. arabica*, which are usually isolated from each other in Africa, and inducing deregulation of their phenologies. The first spontaneous inter-specific hybrid between *C. arabica* and *C. canephora* was reported on Timor Island by Cramer (1957). In New Caledonia, *Coffea* species have been introduced since 1856, mainly from Java to develop coffee production (Le Pierrès, 1999; Pellegrin *et al.* 1983). *C. arabica* was first introduced, but emergence of the *Hemilia vastatrix* pathogen (coffee leaf rust) around 1910 severely damaged many coffee plantations, promoting the introduction of *C. liberica* and *C. canephora*. In the 1920s, it was common to find shade plantations consisting of different species. Finally, because of economic constraints, plantations were abandoned for more than 70 years, thus leaving the trees to grow and evolve as natural populations. In New Caledonia, natural hybridizations between the two main coffee species has been noted in abandoned shade plantations (Mahe *et al.* 2007), but only in the central part of the main island. A tri-specific plantation was identified in Sarramea Valley (S 21°38.584' - E 165°51.733') (Le Pierrès, 1999).

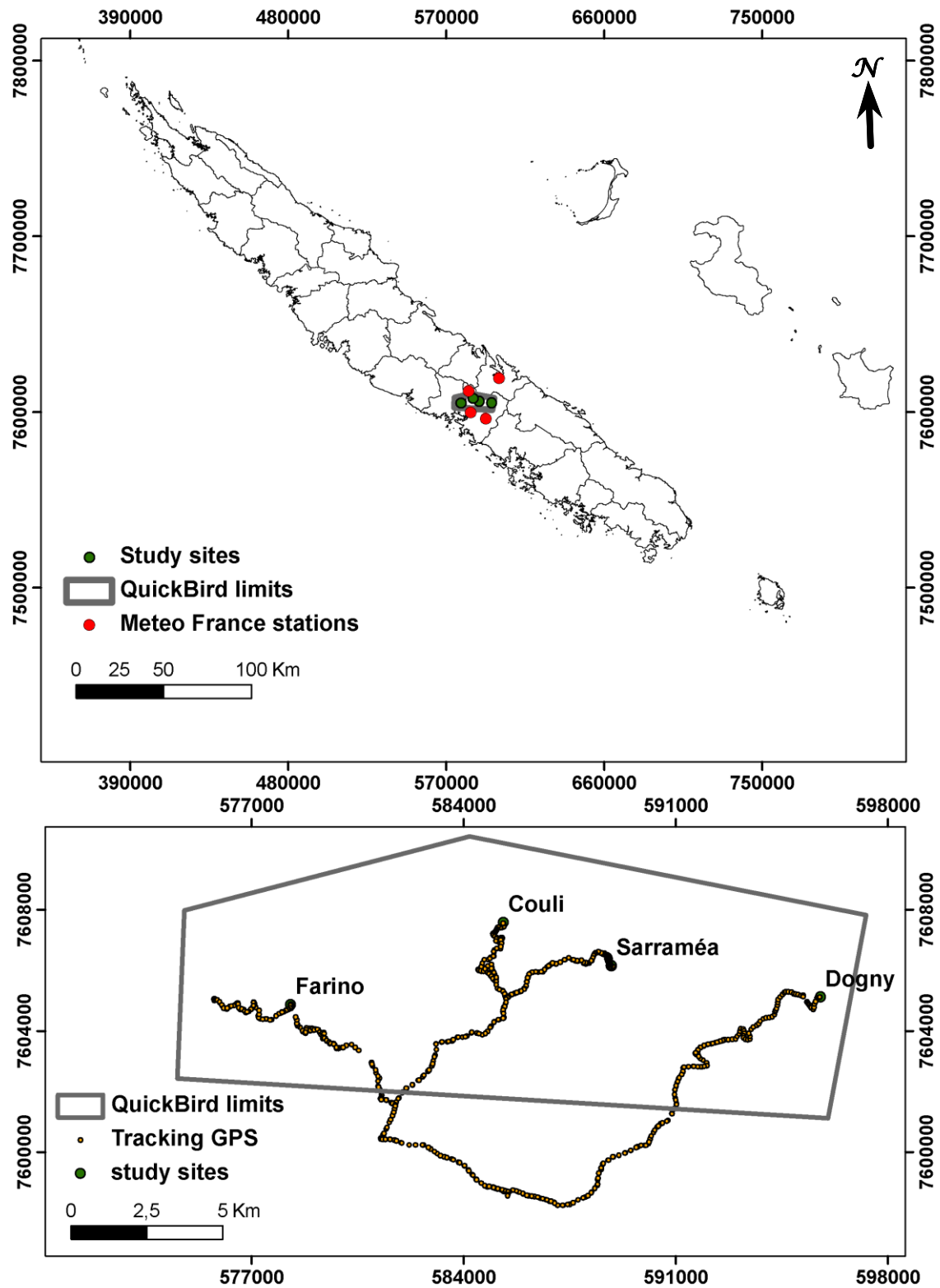


Fig. S1. Map of the study area in New Caledonia, the four reference sites and details of the study area delimited by the Quickbird satellite images georeferenced by the tracking GPS.

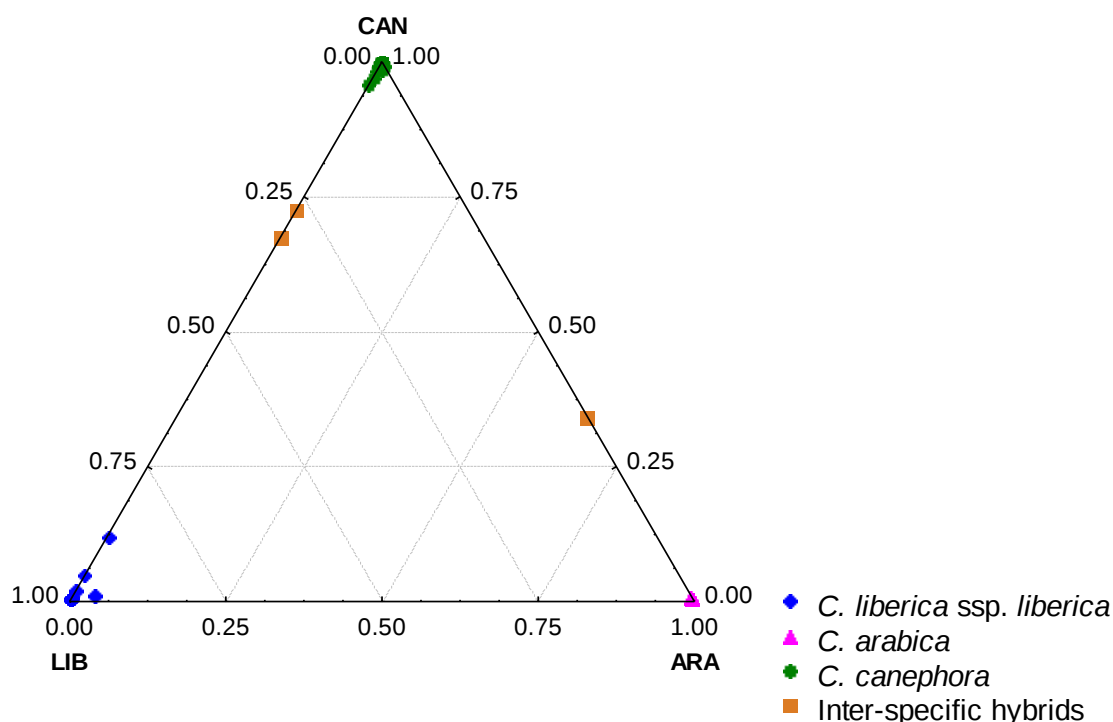


Fig. S2. Trispecific plot representing the genetic admixture of the sampled individuals studied. The three species groups are positioned at each triangle vertex (LIB: *C. liberica* ssp. *liberica*, ARA: *C. arabica* and CAN: *C. canephora*) and inter-specific hybrids at intermediate positions. Numbers correspond to the relative proportion of genome drawn from each group.

Table S1. Diversity indices for the 16 SSRs and 10 EST-SSR loci in the African reference samples (N=54). Linkage group locations for mapped markers are indicated.

N=54 SSR						EST-SSR					
Locus	C*H ¹	Na	Ho	H _E	PIC	Locus	L*C ²	Na	Ho	H _E	PIC
M257	B	7.00	0.32	0.70	0.67	Es18	-	4.00	0.18	0.54	0.46
M 259	N	10.00	0.63	0.71	0.68	Es24	-	5.00	0.25	0.65	0.61
M314	G	9.00	0.40	0.76	0.72	Es25	LG2	5.00	0.26	0.66	0.61
M367	-	14.00	0.32	0.86	0.85	Es43	-	8.00	0.47	0.77	0.73
M368	-	16.00	0.61	0.92	0.91	Es47	-	13.00	0.60	0.85	0.83
M387	J	10.00	0.42	0.78	0.76	Es62	-	17.00	0.46	0.85	0.84
M394	C	13.00	0.62	0.81	0.78	Es70	-	6.00	0.15	0.55	0.50
M495	-	8.00	0.38	0.76	0.72	Es73	-	3.00	0.16	0.48	0.38
M755	-	14.00	0.57	0.90	0.89	Es82	-	9.00	0.35	0.71	0.67
M764	-	9.00	0.51	0.79	0.77	Es90	-	7.00	0.48	0.76	0.73
M779	-	11.00	0.69	0.84	0.82						
M782	-	5.00	0.12	0.51	0.42						
M804	M	2.00	0.04	0.04	0.04						
M809	H	9.00	0.36	0.56	0.54						
M821	F	14.00	0.42	0.78	0.76						
M856	I	16.00	0.52	0.89	0.88						
total		167.00				total		77.00			
Mean		10.44	0.43	0.73	0.70	Mean		7.70	0.34	0.68	0.64

¹Linkage group of (*C. canephora* x *C. heterocalyx*) x *C. canephora* genetic map from Coulibaly et al. (2003)

²Linkage group of (*C. liberica* x *C. canephora*) x *C. canephora* genetic map from N'Diaye et al. (2007)

Na: number of alleles; Ho: the observed heterozygosity; H_E: Gene diversity or expected heterozygosity; PIC: polymorphism information content

FOREST COVER	CREEK PROXIMITY	SLOPE	GAP TEMPERATURE	PRECIPITATION SENSITIVITY	SENSITIVITY
FOREST (1)	< 100m (1)	> 30% (1)	> 7°C (1)	1	5
				2	6
				3	7
			< 7°C (2)	4	8
				5	9
				6	10
		10-30% (2)	> 7°C (1)	1	6
				2	7
				3	8
			< 7°C (2)	4	9
				5	10
				6	11
		< 10% (3)	> 7°C (1)	1	6
				2	7
				3	8
			< 7°C (2)	4	9
				5	10
				6	11
	100-200m (2)	> 30% (1)	> 7°C (1)	1	7
				2	8
				3	9
			< 7°C (2)	4	10
				5	11
				6	12
		10-30% (2)	> 7°C (1)	1	7
				2	8
				3	9
			< 7°C (2)	4	10
				5	11
				6	12
		< 10% (3)	> 7°C (1)	1	8
				2	9
				3	10
			< 7°C (2)	4	11
				5	12
				6	13
	> 200m (3)	> 30% (1)	> 7°C (1)	1	8
				2	9
				3	10
			< 7°C (2)	4	11
				5	12
				6	13
		10-30% (2)	> 7°C (1)	1	9
				2	10
				3	11
			< 7°C (2)	4	12
				5	13
				6	14
		< 10% (3)	> 7°C (1)	1	9
				2	10
				3	11
			< 7°C (2)	4	12
				5	13
				6	14
NO FOREST (2)				6	15

Table S2. Decision tree model used to calculate the inter-specific hybridization phenomenon sensitivity. The sensitivity is the sum of the five parameter modalities.sensitivity. The sensitivity is the sum of the five parameter modalities.

III – Conclusion

Les conditions locales semblent favorables à la co-habitation et à l'hybridation, comme le montre notre étude génétique qui confirme la présence de 3% d'hybrides inter-spécifiques dans un sous-échantillon de la population tri-spécifique témoin.

La comparaison avec le milieu d'origine des espèces concernées a permis non seulement de souligner les exigences environnementales incontournables pour les caféiers (selon des conditions environnementales rencontrées en Afrique et à Sarraméa) mais aussi de mettre en exergue les conditions environnementales originales rencontrées à Sarraméa. L'expertise environnementale mise en place a notamment révélé des différences très marquées ainsi que des similitudes quant au régime climatique annuel entre Sarraméa et les régions d'origine des espèces, la Côte d'Ivoire pour *C. canephora* et *C. liberica* et l'Ethiopie pour *C. arabica*. Les conclusions tirées de l'analyse fine des climogrammes tendent à montrer que le climat rencontré à Sarraméa permet un rapprochement des niches climatiques des espèces favorisant leur cohabitation tout en répondant à leurs exigences environnementales. L'introduction des caféiers en Nouvelle-Calédonie a donc engendré une modification de leurs niches écologiques (comme illustré Figure 1.1, p24), et plus particulièrement un rapprochement des niches des deux espèces les plus fréquentes (*C. arabica* et *C. canephora*). Les niches écologiques ont donc été décrites par l'analyse des séquences climatiques (au travers de climogrammes) mais aussi selon deux variables environnementales : l'humidité du sol (compromis entre humidité et drainage) et la lumière (présence de couvert forestier dense).

Nous avons ici mis en évidence la modification des niches écologiques permettant la sympatrie de ces espèces. Ceci est la première condition *sine qua non* à l'installation d'une nouvelle dynamique d'hybridation inter-spécifique naturelle entre ces espèces. La véritable description des niches est présentée dans le chapitre suivant.

Chapitre 5 - Prédiction de distribution des niches favorables aux hybridations inter-spécifiques entre caféiers : caractérisation fine du couvert forestier et modèle environnemental

Le travail présenté dans ce chapitre a fait l'objet d'une publication :

Gomez, C. et al (en révision) Predicting New-Caledonian distribution of shade coffee hybrid zones using high resolution satellite imagery in an integrated model. Remote Sensing of Environment



Le travail précédent d'expertise a mis en évidence les conditions environnementales favorables à la cohabitation des espèces de caféiers et à la levée de barrière à la reproduction entre elles.

A présent, afin de pouvoir parler d'une véritable description de niches environnementales, il est nécessaire de travailler à une échelle plus fine permettant de révéler toute la richesse de l'hétérogénéité des habitats.

Pour cela, les données climatiques n'ont pas été considérées dans la mesure où elles ne montrent pas de variations à cette échelle spatiale. Il convient donc de se baser sur des données d'imagerie à très haute résolution qui permettent de caractériser le couvert forestier (témoin des différents types d'habitats forestiers) à une échelle métrique. Les données d'imagerie satellitaires nous permettent aussi de caractériser finement la couverture forestière dense présente au dessus des anciennes plantations de caféiers. En effet, une expertise botanique a décrit le couvert végétal au dessus des anciennes plantations comme un mélange original et spécifique entre des espèces autochtones (forêt primitive) et des espèces introduites (Légumineuses habituellement associées aux plantations de caféiers sous couvert à travers le monde).

I – Contexte de l'étude et démarche abordée

Il s'agit ici de travailler à la prédiction de distribution des niches favorables aux hybridations inter-spécifiques entre espèces de caféiers, par l'analyse précise du couvert forestier et au travers d'un modèle environnemental intégré. Ce dernier est inspiré de l'expertise présentée précédemment mais inclut des variables prédictives issues de télédétection.

La démarche que nous avons abordée ici, a consisté à :

- 1) identifier les arbres de couvert dans l'objectif de mettre en évidence une signature particulière de la canopée par rapport au reste de la forêt qui n'abrite pas d'anciennes plantations de caféiers,
- 2) définir des variables permettant de distinguer ce couvert forestier particulier au travers d'analyses fines des images satellites à très haute résolution spatiale,
- 3) mettre en place un modèle mathématique environnemental prenant en compte les résultats issus de l'expertise environnementale précédente et les variables prédictives

issues de télédétection afin de définir et de prédire la présence de zones favorables à l'hybridation entre espèces de caféiers dans d'anciennes plantations.

1.1 - Cas particulier des plantations sous couvert

La présence d'un couvert forestier dense au dessus des populations de caféiers impose à ces derniers des conditions environnementales particulières, comme par exemple l'accès limité à la lumière qui modifie notamment leur croissance et leur rendement (Campanha *et al.*, 2005; Perfecto *et al.*, 2005). Ce couvert forestier peut également avoir des conséquences sur le taux d'humidité et les variations de températures journalières.

En prenant en compte les caractérisations botaniques de ce couvert au site de Sarraméa, il apparaît que les caféiers ont été plantés dans des forêts préexistantes mais conjointement avec certaines espèces habituellement associées aux plantations de caféiers.

Ceci correspond à un intermédiaire entre des plantations rustiques et des polycultures traditionnelles en termes de composition de la canopée selon l'expertise de Moguel et Toledo (1999) (Figure 5.1).

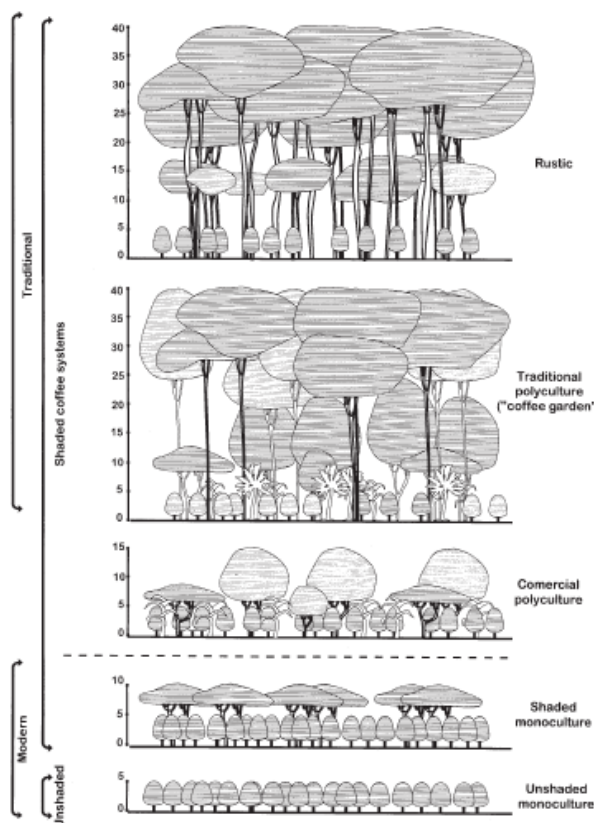


Figure 5.1 Représentation des différents types de plantations de caféiers existantes.

Le système de plantation **rustique** consiste à planter des pieds de caféiers sous des forêts autochtones. Ce système n'a d'effet que sur le sous-bois, mais la canopée originale est grandement conservée. La **polyculture traditionnelle** correspond à un stade plus avancé de modification de la forêt originelle avec l'ajout dans le couvert d'espèces introduites et parfois le retrait de certaines espèces autochtones pour optimiser l'ombrage. La **polyculture commerciale** consiste à éliminer complètement la forêt originelle et constituer un couvert d'arbres spécifiques appropriés aux différentes cultures. Enfin, les systèmes de **monocultures** sont plus intensifs avec ou sans l'ajout des seules légumineuses spécifiques.

Les espèces autochtones majoritairement présentes sont : le goudronnier (*Semocarpus atra*), le bancoulier (*Aleurites moluccana*), le cerisier bleu (*Elaeocarpus angustifolius*), ainsi que des espèces endémiques des genres *Ficus* sp. et *Cryptocarya* sp. Conjointement à ces espèces

indigènes, une majorité de bois noir d'Haïti (*Samanea saman*) et de bois noir (*Albizia lebbek*) ont été identifiés. Il s'agit de légumineuses connues pour leurs capacités à enrichir le sol en azote, pour leur port en parasol de grande dimension et leur ombrage modéré. Ces forêts sont donc potentiellement caractérisables par des attributs particuliers au niveau de la canopée.

1.2 - La caractérisation fine de la canopée

Initialement, la caractérisation botanique des espèces de couvert avait pour objectif de repérer ces espèces une à une sur les images satellitaires à Très Haute Résolution (THR) par un processus de segmentation (détection des objets en fonction de leurs caractéristiques spectrales et de forme) séparant les houppiers puis une classification par espèce. Il s'agissait d'utiliser la présence d'arbres marqueurs (les légumineuses introduites) pour caractériser le couvert forestier protégeant des anciennes plantations de caféiers abandonnées. Dans la mesure où nous travaillons sur des images satellites THR, il aurait été possible de délimiter chaque arbre et de les caractériser par des propriétés texturales ou spectrales particulières, comme cela a été fait pour des inventaires de diversité notamment (Couteron *et al.*, 2005 ; Johansen *et al.*, 2007 ; Palace *et al.*, 2008). Or les forêts de cette région sont particulièrement hétérogènes et la segmentation des images, dans le but d'obtenir un objet par houppier est irréalisable.

La principale difficulté a résidé dans l'aspect hétérogène des forêts de cette région du Pacifique sud, car les milieux tropicaux et insulaires sont très complexes d'un point de vue topographique, avec une couverture végétale à biodiversité élevée (Kier *et al.*, 2009).

Nous avons donc opté pour une méthode à l'échelle de la forêt en travaillant sur les aspects globaux de la canopée et plus particulièrement sur son hétérogénéité et la taille des arbres qui la composent. Les approches texturales (Haralick *et al.*, 1973) se sont avérées être les plus appropriées dans notre étude pour détecter des configurations de forêts particulières, contournant ainsi les difficultés liées à la délimitation des houppiers en milieu hétérogène. En prenant en considération les connaissances historiques et botaniques de ces anciennes plantations, nous avons travaillé à la détection des canopées hétérogènes (forêt primaire en association avec d'autres arbres introduits) et des canopées présentant majoritairement des houppiers larges (caractéristiques des légumineuses spécifiques introduites). Ces deux attributs constituent alors deux variables prédictives caractérisant la présence d'anciennes plantations de caféiers. Afin de prédire la distribution des zones favorables aux hybridations, il convient ensuite de combiner l'information de ces deux variables avec les variables environnementales qui favorisent les hybridations inter-spécifiques.

1.3 - Mise en place d'un modèle de prédiction

La prédiction des habitats favorables aux hybridations inter-spécifiques a été améliorée non seulement par l'intégration de variables prédictives issues de l'analyse d'images THR, mais aussi par la combinaison statistique de variables environnementales. Ces paramètres environnementaux insérés au modèle sont la pente, l'accumulation des flux (importance du drainage) et l'altitude.

Un aspect original de cette étude réside dans la comparaison de l'application de deux types de modèles, un arbre de décision et un réseau de neurones, pour la détection des habitats favorables. En effet, ces deux types de modèles sont très différents et complémentaires à la fois car :

- l'arbre de décision est un modèle assez simple à mettre en place, et fournit des résultats très clairs stipulant le cheminement exact des combinaisons de paramètres aboutissant à des zones favorables ou non favorables. Ces combinaisons de paramètres illustrent des situations écologiques bien particulières.
- le réseau de neurones est une méthode non-linéaire qui peut donc paraître moins intuitive. En effet, la nature des combinaisons non linéaires des paramètres mises en place pour caractériser les zones n'est pas accessible. Les résultats d'un tel modèle sont compris entre -1 et 1, du moins favorable au plus favorable. L'avantage de cette méthode est d'estimer le taux d'erreurs dans la classification associé à chaque combinaison, en parallèle avec l'apprentissage du modèle.

II – Résultats principaux

Le principal objectif de ce chapitre a été de proposer une carte de prédiction de distribution des zones favorables aux zones hybrides *via* la caractérisation précise du couvert forestier par télédétection.

Les résultats obtenus ont fait l'objet d'une publication, en révision :

Gomez, C. *et al* (en révision) Predicting New-Caledonian distribution of shade coffee hybrid zones using high resolution satellite imagery in an integrated model. ***Remote Sensing of Environment***

Predicting New-Caledonian distribution of shade coffee hybrid zones using high resolution satellite imagery in an integrated model

Gomez, C.¹, Mangeas, M.², Petit, M.⁴, Corbane, C.⁵, Hamon, P.³, Hamon, S.³, De Kochko, A.³, Le Pierres, D.³, Poncet, V.³, Despinoy M.²

Addresses

¹ IRD, UMR DIAPC, 101 Promenade Roger Laroque – Anse Vata, BP A5, 98848 Noumea Cedex, New-Caledonia

² IRD, US ESPACE, 101 Promenade Roger Laroque – Anse Vata, BP A5, 98848 Noumea Cedex, New-Caledonia

³ IRD, UMR DIAPC, 911 avenue Agropolis, BP 64501, 34394 Montpellier Cedex 5, France

⁴ IRD, US ESPACE, Maison de la Télédétection, 500 rue Jean-François Breton 34093 Montpellier Cedex 5, France

⁵ Joint Research Centre (JRC), Via E.Fermi 2749, I-21027 Ispra (VA) Italy, TP 267

Corresponding Author: marc.despinoy@ird.fr

Abstract

Key-words: Complex tropical environment, Decision tree, Multi-scale textural analysis, Neural Networks, New-Caledonia, Quickbird imagery, Rustic coffee-grown

In New Caledonia, shade grown coffee plantations were abandoned for economic reasons in the middle of the 20th century. Introduced coffee species (*Coffea arabica*, *C. canephora* and *C. liberica*) survived in the wild and have cross-hybridized. Coffee species were planted in native forest in association with certain leguminous trees (mostly introduced species) to improve their growth. Thus the canopy covers over rustic shade coffee plantations is heterogeneous with a majority of large crowns (attributed to leguminous trees).

The aim of this study was to detect suitable areas for coffee inter-specific hybridization in New Caledonia using environmental parameters and remotely sensed predictors. Due to the complex tropical environment, remote sensing imagery needs to be spatially accurate and to have the appropriate bands for monitoring vegetation cover. QuickBird imagery was selected with a spatial resolution of 0.6 meters and 4 bands in the visible and near infrared wavelength range.

The two most relevant remotely sensed indicators, canopy heterogeneity and tree crown size, were acquired by the sequential use of an artificial neural network (to detect the presence of tree crowns), image processing (such as textural analysis) and modelling. The final model has two remotely sensed indicators and three physical parameters based on the Digital Elevation Model: elevation, slope and flow accumulation.

Using these five predictive variables as input, two modelling methods, a decision tree and a neural network, were implemented. The decision tree, which showed an accuracy of 96.9% on the test set, revealed the involvement of ecological parameters in coffee hybridization. Hybrid zones could be characterized by different combinations of modalities, underlining the complexity of the environment concerned. For instance, in heterogeneous zones (heterogeneity class > 1.5), with a majority of “large” and “medium” crowns (crown size class > 2.5), the environment for hybrid zones was located either on a slope > 53.5% between an elevation of 194-429 m or on a slope < 53.5% between an elevation of 317-429 m.

The application of the neural network on the whole area gave a predictive map that distinguished the most suitable areas by means of a non-linear continuous indicator. This map provided a confidence level for each area. The most favourable areas were geographically localized, providing a clue for the detection and conservation of favourable areas for *Coffea* species neo-diversity.

Introduction

Coffee species (*Coffea* genus) are cultivated in inter-tropical regions using five different coffee-growing systems (Moguel, Toledo, 1999): rustic polyculture, traditional polyculture (“coffee garden”), commercial polyculture, shaded monoculture and sun monoculture. Philpott et al (2008) compared these five systems and two forest types using quantitative specifications for vegetation variables (tree richness, canopy cover, coffee density, number of shade strata) and for management techniques including variable human intervention.

Rustic polyculture was shown to be the system with the least human interventions in which coffee trees are planted under existing forest in association with selected leguminous trees and sometimes in association with other crops like banana, pineapple, orange or lemon. According to Soto-Pinto et al. (2007), 77% of the trees classified as the best for shade are leguminous trees. Leguminous trees are known to enhance coffee production by enriching the soil in nitrogen, providing suitable shade and growing rapidly. The canopy cover of such rustic polyculture has to be characterized by a special canopy signature.

It has been shown that coffee agroecosystems contribute significantly to biodiversity conservation (Bandeira *et al.*, 2005; Lopez-Gomez *et al.*, 2008; Mendez *et al.*, 2007; Perfecto *et al.*, 2003; Philpott *et al.*, 2008). Moreover, complex agroecosystems including dense canopies are able to maintain high levels of biodiversity as demonstrated in different taxa: ants (Philpott *et al.*, 2006), butterflies and birds (Mas, Dietsch, 2004) or orchids (Solis-Montero *et al.*, 2005). However, few studies have focused on vegetation structure assemblage in shade-grown coffee (Bandeira *et al.*, 2005; Soto-Pinto *et al.*, 2001; Soto-Pinto *et al.*, 2007). And existing studies focused on structure variables of shade vegetation in terms of floristic composition

(Bandeira *et al.*, 2005), number of vegetation strata (Soto-Pinto *et al.*, 2001), species, and the distribution frequency of rustic coffee cultivation systems (Soto-Pinto *et al.*, 2007). A quantitative classification of coffee agroecosystems was made based on structure variables, vertical vegetation density, tree richness and abundance, coffee plant density and the mean height of shade trees (Hernández-Martínez *et al.*, 2009). But no classification has yet been made from above using remote sensing despite its more generalizable suitability for characterizing the canopy cover of coffee agroecosystems.

Characterization of the forest cover is one of the main elements in forest resource management. Multispectral sensors and high spatial resolution satellite sensors (such as Quickbird or Ikonos) are both suitable for the detailed study of forest composition. As a result, very high-resolution remotely sensed canopy images are now being increasingly used to characterize forest structure with the aim of producing an inventory of diversity and forest resources (Couteron *et al.*, 2005; Johansen *et al.*, 2007; Palace *et al.*, 2008). Tropical forests are structurally very complex with high biological diversity and complex vegetation dynamics resulting in a spatially diverse array of forest structures (Palace *et al.*, 2008), particularly in forests and mountains areas (Whitmore, 1982). Such forests are a challenge for remote sensing, given the persistent cloud cover, topographic effects and spectral confusion between different types of woody cover (Cordero-Sancho, Sader, 2007). To map and characterize tropical land cover, especially tropical forests, several remote sensing systems and image analysis techniques have been used (Asner *et al.*, 2002; Cordero-Sancho, Sader, 2007; Johansen *et al.*, 2007; Mayaux *et al.*, 1999).

When working on forests, two methods can be used: segmentation analyses (classification of objects) or direct textural analyses of the image when segmentation is not possible. When segmentation is possible,

very powerful analyses can delineate individual tree crowns at a high spatial resolution to investigate forest diversity (Culvenor, 2002; Erikson, 2004; Orka *et al.*, 2009; Palace *et al.*, 2008; Pouliot *et al.*, 2002; Read *et al.*, 2003). However, most algorithms for the delineation of tree crowns were developed on homogenous forest (either mono-specific or temperate forest) (Culvenor, 2002; Erikson, 2004), while others were based on isolated trees (Read *et al.*, 2003), with specific objectives such as species inventory.

In the analysis of tropical forest patterns, the individual tree scale is difficult to achieve. Textural information is useful to process and analyze satellite image data whereas common techniques (e.g. standard per-pixel based classifiers or unique use of vegetation indices) are not suitable for high spatial resolution image data (Johansen *et al.*, 2007; Kayitakire *et al.*, 2006). Kayitakire *et al.* (2006) showed texture attributes derived from the Grey Level Co-occurrence Matrix (GLCM) relevance as proposed by Haralick (1973) for characterizing forest patterns. The density of the forest canopy is one of the attributes used to characterize a specific forest pattern.

Our study was conducted in New Caledonia, and focused on the characterization of canopy cover using remote sensing tools. In New Caledonia, three cultivated coffee species have been introduced since 1856: *Coffea arabica* L., *Coffea canephora* Pierre and *Coffea liberica* Hiern.

The leguminous trees that have been planted are mainly introduced species, like the two main species associated with rustic coffee plantation, *Samanea saman* or *Albizzia lebbek* (Muzinger, personal communication). Based on the coffee management systems described (Moguel, Toledo, 1999; Philpott *et al.*, 2008), coffee agroecosystems in the Sarramea region correspond to “traditional coffee polyculture” in the valley, and to “rustic coffee system” at higher elevations. The native forest canopy has been preserved from

human intervention and few leguminous trees have been added. Coffee species have been planted everywhere it appeared possible. As a consequence, the canopy cover above rustic coffee is a native forest with specific leguminous species, which are characterized by large crowns (*Samanea saman* and *Albizzia lebbek*). These rustic plantations are now abandoned but coffee species have been observed obviously well adapted to their environment since they survived without human intervention. In addition, a very original phenomenon has been observed, i.e. the appearance of inter-specific hybrids (Jagoret *et al.*, 1999; Le Pierrès, 1999b; Mahe *et al.*, 2007a).

To understand how coffee adapted in a complex environment whose structure was modified by the introduction of cover species, our study focused on the characterization of the forest cover. First, canopy specific bio-physical characteristics for coffee presence based on remotely sensed predictive variables have had to be computed to describe the structure of the forest in the coffee hybrid zone. Second, an integrated model had to include the canopy characterization and environmental parameters which could be important for defining suitable environments for coffee inter-specific hybridization. Indeed, physical parameters such as elevation based on Digital Elevation Model (DEM) and derived products (slope or flow accumulation) are critical variables for predicting habitats that are suitable for coffee (Lin, 2007). Finally, environmental niches that are favourable for coffee inter-specific hybridization and overall biodiversity evolution could be located in the region of Sarramea.

The paper is organized as follows: in the following Section 2 we describe the study area, the training data set and the satellite images used. Specifications concerning the method of selecting the attributes are presented in Section 3. Methods used for image analysis (neural networks and decision tree models based on textural

analysis) and statistical analyses are also described in Section 3. In Section 4 we present our experimental results on classifications including the analysis of the contribution of textural attributes according to the different models and the environmental characterization of suitable areas for coffee hybridization. The discussion and our conclusions are in Section 5.

2- Materials

2.1. Study area

New Caledonia is an island located in the southwest Pacific Ocean around 21°30'S, 165°30'E, approximately 1200 kilometres east of Australia and 1500 kilometres northwest of New Zealand. The study area is located in the central mountain region, around Sarramea (Figure 1). We focused on two study sites where different coffee species cohabit and hybridize and the cover trees had already been identified. Data on the study sites and on the region were collected during field campaigns.

2.2. Data collection, processing and technical issues

Due to the complex tropical environment in New Caledonia, which is associated with high species richness and biodiversity (Kier *et al.*, 2009), remote sensing imagery had to be spatially accurate to enable specific estimation of forest parameters and had to have suitable bands to monitor vegetation cover. QuickBird imagery has a spatial resolution of 0.6 metres and 4 bands in the visible and near infrared wavelength range. Two Quickbird satellite images were acquired (the eastern one on June 18, 2005 and the western one on the August 16, 2005). Panchromatic and multispectral images were then merged using a standard numerical multiplicative method specially adapted for high resolution imagery (Table 1).

As tropical islands like New Caledonia have a complex topography, a digital elevation model (DEM) interpolated at 0.6 m was used for ortho-correction of the images and to

determine the slope and flow accumulation. The slope is expressed in percents. To calculate flow accumulation, the flow direction was first computed from the DEM sink (eliminating the false low points). The result of flow accumulation is a raster of accumulated flow to each cell, determined by summing the weight for all the cells that flow into each down slope cell. The flow accumulation is expressed in number of pixels. The slope and the flow accumulation were calculated using the Hydrology Spatial Analyst Tools in ESRI® ArcGis™ 9.2.

Attributes were selected using Weka software (Witten, Frank, 2005) and the models were designed using MATLAB®. All the image pre-processing was implemented in ENVI/IDL framework.

2.3. Training and test sets

Image processing and the models were first tested on two subsets (Figure 1), *Farino* and *Sarramea* (according to their geographic position), as they were representative of the whole area in terms of diversity of land cover (inland water bodies, roads, bare soil, vegetation cover types) and forest patterns (dense, sparse, different sized tree crowns). The *Farino* and *Sarramea* subsets contained 1217*1714 and 705*795 pixels respectively. Both subsets were used for training the model (5% of the data subset) and testing the model (the whole subset). Over-fitting was avoided and the capacity of generalisation of the models was evaluated.

Considering the situation of shade grown coffee in native forest and in association with specific leguminous trees (characterized by their large crowns), the remote sensing analysis consisted in detecting canopy heterogeneity and the crown-size trends.

Training data sets were designed by hand using the subsets (*Farino* and *Sarramea*) according to the three classification types: (1) discrimination of tree crowns, (2) classification of four sizes of tree crowns and (3) canopy heterogeneity or homogeneity patterns (Figure 2). The four classes of tree crowns were defined by window size: large crowns comprised from 37 to 73 pixels,

medium crowns from 24 to 36 pixels, medium-small crowns from 15 to 23 pixels and small crowns 1 to 14 pixels. The larger crowns, corresponding to the specific leguminous trees planted in association with coffee, measured more than 22 metres but less than 44 metres (which correspond to 14 - 26 pixels). The upper limit was based on the largest tree crowns.

3- Methods

To obtain the predictive distribution of favourable habitats for coffee inter-specific hybridization the global scheme consisted in four steps detailed in figure 3:

- step 1: detection of tree crowns
- step 2: extraction of textural attributes
- step 3: calculation of remotely sensed indicators (heterogeneity and tree crown sizes)
- step 4: global model

3.1 Attributes selection

In our study, one of the key issues is to select attributes by ranking the predictive variables in order to reduce the complexity of the model without losing ecological significance. Indeed, when inputs are selected on the basis of ecological knowledge, the large number of their attributes can introduce complexity and considerably increase computational time, leading to the “curse of dimensionality” (Bishop, 1995) and disrupting the expected increase in class discrimination.

Attribute selection methods (implemented in Weka software) were computed two times in the course of the global scheme: at step 3, to select the most relevant attributes among the calculated textures, and at step 4, to evaluate the contribution of the five predictive variables.

In step 3, attribute selection methods were implemented independently for the two models (canopy heterogeneity and size of the tree crowns) based on the two corresponding training data sets. According to Kira and Rendell (1992), the “Relief Attribute

Evaluation” algorithm (Relief Ranking Filter) is best suited for numeric inputs. Indeed, the relief algorithm has interesting characteristics: it selects only statistically relevant attributes, it is not misled by interaction between attributes, and is noise-tolerant. All sampled attributes were included in this operation with equal influence; the nearest neighbour’s k was constrained to 10. The 50 attributes were ranked with the most relevant ones in the highest position. The sill was fixed according to a cross validation test which consisted in running the models with an increasing number of inputs from one to the number that allowed 90% of accuracy.

In step 4, all the predictive variables had to be computed in a global model. The contribution of the attributes was evaluated using “Info Gain Attribute Evaluation classifier”. This evaluator gave the significance of an attribute by measuring the information gain with respect to the class (<http://dataminingntua.files.wordpress.com/2008/04/weka-select-attributes.pdf>).

3.2. Detection of tree crowns

Tree crowns were detected on the image to design a fixed mask (Figure 3, step 1). To distinguish the forest cover (presence of tree crowns), a nonlinear method such as an artificial neural network can be used. In this method, a single source of information (Quickbird spectral bands) is used, and the approach combines the raw information nonlinearly to make the classification (Joshi *et al.*, 2006; Pacifici *et al.*, 2009). An artificial neural network has been shown to outperform multiple linear regression and maximum likelihood in predicting forest canopy density from remote sensing data (Landsat TM) without any bias in the prediction (Joshi *et al.*, 2006). A Multi-Layer Perceptron was used as described by Minsky and Paper (1969). The architecture we used is a feed forward neural network, with a hyperbolic tangent as activation function for hidden (10 neurons) and output layers (1 neuron). The parameters of the complex functions used in the network were

calibrated during training on the data set using the classification method (tree crowns/no tree crowns). The parameters were trained using one of two alternative optimization methods, the quadratic method (BFGS) or resilient back propagation (RP). The advantage of the back propagation method over the quadratic one is that it takes into consideration the error in classification and minimizes it by estimating the gradients and iteratively changing parameters.

Only two bands were used here: NIR (Near Infra Red) and RED, since they are the most sensitive to chlorophyll activity and accentuate contrasts. For instance, the NIR band is known to accentuate shade between tree crowns, leading to better delineation of crown boundaries (Culvenor, 2002).

The efficiency of the neural network was evaluated by comparison with results obtained with a decision tree implemented on the same training data set.

This technique produced the forest mask that was used in subsequent analyses.

3.3. Extraction of textural attributes

Eight textural attributes were calculated on two vegetation indices, within three different window sizes to describe the forest canopy structure (Figure 3, step 2). In this way, we obtained 48 textural attributes.

Within forest areas delineated by the forest mask created in step 1, textural attributes derived from GLCM were implemented on the most pertinent predicted vegetation indices (NDVI and VIN). The NDVI (Normalized Difference Vegetation Index) was proposed by Rouse et al. (1974) and is widely used in remote sensing vegetation studies as it reveals fine contrasts in vegetation, while the VIN (Vegetation Index Number) provides straightforward information as it is easy to calculate. Both are robust to the effects of slope on radiometry pixel values.

Let R and NIR be the Red and the Near Infra Red bands respectively. The NDVI and VIN indices can be computed as follows:

$$(1) \text{NDVI} = (NIR - R) / (NIR + R)$$

$$(2) \text{VIN} = NIR / R$$

The eight textural attributes we used were defined by Haralick et al. (1973): homogeneity, dissimilarity, mean, contrast, variance, second moment, correlation and entropy, and were calculated with three window sizes (15, 45 and 99 pixels) using the following formulae:

$$(3) \text{Homogeneity} = \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} \frac{p(i, j)}{1 + (i - j)^2}$$

$$(4) \text{Dissimilarity} = \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} p(i, j) \cdot |i - j|$$

$$(5) \text{Mean} = \sum_{i,j=0}^{N-1} i \cdot p(i, j)$$

$$(6) \text{Contrast} = \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} p(i, j) \cdot |i - j|^2$$

$$(7) \text{Variance} = \sum_i \sum_j p(i, j) \cdot (i - \mu_i)^2$$

$$(8) \text{SecondMoment} = \sum_i \sum_j p(i, j) \cdot (i - \mu_i)^2$$

$$(9) \text{Correlation} = \frac{\sum_i \sum_j p(i, j) \cdot (i - \mu_i) \cdot (j - \mu_j)}{\sigma_i \cdot \sigma_j}$$

$$(10) \text{Entropy} = - \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} p(i, j) \cdot \log_2 p(i, j)$$

Where:

- σ and μ denote the mean and standard deviation respectively,
- i, j are the grey tones in the window, which are also the coordinates of the co-occurrence matrix space,
- $p(i, j)$ denotes the normalized frequencies with which two neighbouring resolution cells separated by a fixed shift occur on the image,
- N stands for the co-occurrence matrix dimension.

The size of the calculation window depends to a great extent on the characteristics of the remotely sensed objects, mainly the size attribute (for instance in our case, the

maximum tree crown size was around 75 pixels).

3.4. Remotely sensed indicators (heterogeneity and tree crown sizes)

Two remotely sensed predictive variables based on the selected textural attributes were created (Figure 3, step 3). Then, based on relevant textural attributes, two decision trees were applied in order to discriminate i) four crown sizes (Small, Medium small, Medium, Large) and ii) heterogeneous versus homogeneous areas (Figure 3, step 3). At each node of the tree, the standard C4.5 algorithm chooses one attribute of the data that most effectively splits the samples into subsets enriched in one class or the other. The splitting criterion is the normalized information gain (difference in entropy). The attribute with the highest normalized information gain is chosen in order to make the decision. The C4.5 algorithm then recurses on the next smaller sub-list until it reaches the optimal classification (Breiman *et al.*, 1984; Quinlan, 1993).

Classification models were applied to the test subsets *Farino* and *Sarramea*. The performance criterion is given by confusion matrices. After validation and testing, the models were applied to the whole study area.

3.5. Statistical analysis and global model

The global model was elaborated using environmental data and the two remotely sensed predictive variables (Figure 3, step 4). The contribution of each variable included in the model was assessed using the Info Gain Attribute Evaluation method.

To model the coffee environment, other layers were added: elevation, slope, and flow accumulation. The whole area was screened during a field campaign to find positive sites (i.e. potential coffee hybrids zones) where coffee species are hybridizing and negative sites (i.e. areas where coffee species do not grow). The positive and negative areas were carefully delineated. Positive and negative masks represent 0.086% and 0.27% of the whole area respectively. These sites were

geo-referenced as validating points on a positive mask to compute statistics.

Statistics were computed according to the validating masks in order to identify the environmental attributes linked with the presence of coffee hybrid zones. Validating masks were used to train the models with 5% of the data.

Two models were then applied to differentiate the whole area into positive and negative areas: a decision tree and a neural network.

First, the relationship between positive or negative zones and the five environmental factors (including the two different forest pattern indicators) was examined using decision tree based on the positive and negative training data sets. The decision tree was pruned to extract a limited amount of environmental information for suitable environment patterns, and the pruned decision tree was subsequently applied to the whole study area. Second, a neural network was implemented with the same training data sets. Artificial neural networks generally provide more accurate prediction of species range dynamics than generalized linear models or classification tree analysis (Araujo *et al.*, 2005). The architecture used in this study was a standard feed forward neural network, with a hyperbolic tangent as activation function for the hidden layer of six neurons and a linear function for the single neuron output layer. The output map comprises two values: -1 for negative data (no hybridization) and 1 for positive data (hybridization). The linear function as output layer allows a smooth scale to be computed ranging from 'no suitable environment' to 'very appropriate environment' for hybridization.

4- Results

All the results are presented here, from the creation of the tree crown masks, to the integrated predictive environmental model, and the selection of textural attributes to compute the remotely sensed predictive variables.

4.1. Tree crown mask

Step 1 of the image treatments consisted in obtaining a mask of the presence of tree crowns as complete as possible without missing a tree because it would form the base of the two next models to classify tree crown patterns (heterogeneity criteria and size classes).

Two methods to obtain the tree crown mask were tested: a decision tree and an artificial neural network. As expected, the results of the neural network (79.15% of accuracy for the tree crown class) were significantly better than the results of the decision tree (61% of accuracy for the tree crown class) (see Figure 4.).

Moreover, two modalities revealed the advantage of the neural network method for implementation of the forest mask. First, the neural network can run with raw data and computes a nonlinear combination of them. On the other hand, a decision tree provides interesting analysis when fed with specific attributes such as input, but the choice of these attributes is somewhat arbitrary as it depends on the author's decision. Choosing the neural network allowed us to take into account only the Quickbird RED and NIR (Near Infra Red) bands, which best characterize the spectral and textural parameters of the vegetation, and let the model find the best combination. The neural network provides a vegetation index that fitted our case study. Second, one of the main differences between the results of the neural network and the results of the decision tree concerns the output types. The neural network provides a continuous result with a smooth scale of output that characterises the closeness to each class, while the decision tree simply provides binary results. We used the probability gradient provided by the neural network to find the best fit with the land cover pattern. In this way, the detection of tree crowns in the whole image can be optimized. As the two next models were implemented according to the extent of the tree crown mask, there was no confusion between forest

and other land cover throughout the study area.

4.2. Selected textural attributes

The attribute selection algorithm applied to the 48 textural attributes and the two vegetation indices (Relief Attribute Evaluator) provided ranked values for each input for both models (canopy heterogeneity and tree crown size). The cross validation test with the first four inputs showed that the canopy heterogeneity model reached 90% of performance for the Sarramea extraction (Figure 5). The same patterns were observed with the tree crown size model. The minimum number of attributes selected to train the model ensured a good performance on the validation data set. Both models were implemented with the first four classified attributes (see Figure 3, step 3).

Finally, the selected attributes were all based on texture calculated with a window of 99 pixels. For the canopy heterogeneity classification model, the variance and the contrast calculated on the VIN and the second moment and the correlation calculated on the NDVI were selected. The variance calculated on the VIN and the dissimilarity, the second moment and the correlation calculated on the NDVI were implemented for the tree crown size classification model.

4.3. Two remotely sensed predictive variables

To evaluate the accuracy of the models, two different confusion matrixes were obtained for each model. The first matrix gave the results of classification for each class globally on all the image pixels, while the second gives the results by class. The confusion matrices for the canopy heterogeneity model are given in Table 2 while the matrices for the tree crown size model are given in Table 1. The accuracy of the model was calculated with the first matrix by summing the correctly classified values (in grey in the tables). The results of the classification of canopy heterogeneity and tree crown size patterns obtained from

the two training subsets showed a high level of accuracy (Figure 6).

The accuracy of the canopy heterogeneity model was 87.8% and 92.7% for the Sarramea and the Farino extractions respectively. The second matrices revealed that the classification errors were higher for the heterogeneity class at 15.7% and 17.8% for Sarramea and Farino respectively, while for the homogeneity class the errors were negligible at 1.2% and 0.73% for Sarramea and Farino, respectively.

Similar matrices were obtained for the tree crown size model with five classes instead of two. The accuracy of the tree crown size model was 78.9% and 83.2% for Sarramea and Farino. For both extractions, the best accuracy was obtained for the large crown class (80.3% and 81.8% for Sarramea and Farino, respectively) while the lowest level of accuracy was obtained for the small crown class (73.76% and 76.45% for Sarramea and Farino respectively).

As expected, the accuracy of the canopy heterogeneity model was higher than that of the tree crown size model. Indeed, it was really difficult to discriminate different sets of tree crown sizes when each tree crown was not clearly separated.

High accuracy values were obtained for both models indicating that the models were well trained on the training data set. As only 5% of the data were used for the training step, over-fitting was avoided, promising good generalization of the models for the whole study area.

For these models, the decision trees appeared to be well suited for the characterization of specific tree crown attributes with specific expert knowledge.

4.4. Application of the models to the whole study area

The three model steps were applied to the whole area and provided three types of data directly linked to the presence of coffee:

1. the tree crown mask
2. canopy heterogeneity
3. the size of the tree crowns

It should be noted that the first step allows the immediate exclusion of 31% of the pixel set and thus considerably reduces the processing time of the two other steps.

4.5 Predicted suitable areas for hybridization

To assess the success of applying the models to the whole area, we focused on the two masks (positive and negative mask) representing the presence and absence of hybrid zones.

By filtering with the validating masks, statistics were calculated for the five layers: elevation, slope, flow accumulation, canopy heterogeneity and size of tree crowns. The results are presented in Figure 7. As expected, the positive areas presented a majority of heterogeneous canopy (78.4%) and of large tree crowns (65.8% of large and medium crowns). Conversely, the negative areas presented a majority of homogeneous canopy (96.2%) and of small and medium small tree crowns (88.9%).

The 'Info Gain Attribute Evaluator' implemented in Weka software ranked the five attributes according to all input data (positive and negative areas) in decreasing order of discrimination. The elevation and the canopy heterogeneity attributes were ranked first with 0.4995 and 0.3487 respectively. Crown size and slope were ranked second with 0.2348 and 0.1097 respectively. Flow accumulation appeared to be less informative as it was ranked third with 0.0465.

To classify the whole study area as suitable or not for coffee inter-specific hybridization based on these statistics, the five predictive variables were implemented in a decision tree. The decision tree was pruned down to 14 branches to obtain a clear picture of the implication of environmental parameters (Figure 8). The flow accumulation parameter did not appear in the pruned decision tree, confirming its lack of informative characteristics. The accuracy of the decision tree was calculated on the positive and negative masks of which 5% of the points had been used in the training step. Only

3.1% of errors were recorded. However, its application to the whole area did not give precise results as the output is binary.

As a consequence, a neural network was also applied to predict the confidence level of the presence of coffee hybrid zones. The neural network was accurate with only 9.79% of errors made on the test data set. The application of the neural network to the whole area resulted in a predictive map that distinguished the most suitable areas through a probabilistic gradient (Figure 9).

On the predictive map (figure 9), 50.77% of the pixels of whole area represent deforested area. For the remaining area, the distribution of the confidence level of the presence of coffee presence is shown in figure 10. Three peaks can be observed on this graph and the last one (values above 0.7, 25.5% of the total number of pixels) characterizes a combination of environmental factors that are favourable for coffee inter-specific hybridization.

5- Discussion and Conclusion

The first attempt to characterize environments favourable for coffee inter-specific hybridization focused on the genetic characterization of inter-specific hybrids and on the environmental expertise needed to understand which environmental parameters are involved in this type of dynamics (Gomez *et al.*, Submitted). The present study went further in environmental characterization to include a fine canopy description based on remote sensing and an integrated model based on an artificial neural network.

5.1 Contribution of remote sensing data to the predictive ecological distribution model

In the first preliminary predictive model of suitable areas for coffee hybridization in New Caledonia (Gomez *et al.*, Submitted), only physical, climatic data and a land cover map were taken into account. In the present study, we used remote sensing data to characterize more precisely the forest

patterns associated with the presence of coffee. The remote sensing data appear to be as informative as bioclimatic data and the accuracy of the implemented models as image treatments optimized the quality and efficiency of the integrated model.

Bioclimatic variables are very important for predictive distribution models at large spatial scales but adding remotely sensed predictors clearly improves the fit of individual species models (Thuiller *et al.*, 2004). Recently, Zimmermann (2007) studied the added value of remote sensing-based predictors in distribution models. This study showed that land cover patterns are highly correlated with bioclimatic gradients, thus both predictor sets are expected to provide similar accuracies. When working at high spatial resolution to detect habitat heterogeneity, the problematic is not the same. Data differs between bioclimatic variables which are often available at large spatial resolution, and remote sensing data, which can provide a very high spatial resolution (up to 0.6 meters for Quickbird imagery). High spatial resolution remote sensing data addresses the part of habitat information that helps to distinguish suitable and unsuitable sites more efficiently than climatic data (Pearson *et al.*, 2004).

In our study, the focus on the Sarramea region was based on climatic characterization (Gomez *et al.*, Submitted) since climate appeared to be particularly important for coffee hybridization (Pezzopane *et al.*, 2008). However, when focusing on the Sarramea region, the study area is less than 20 kilometres wide, and should not involve significant differences in climatic variables. Moreover when improving the spatial resolution to detect the spatial heterogeneity of habitats, the climatic resolution is insufficient, whereas remote sensing data provide high spatial resolution. As a consequence, high spatial resolution imagery was used in our study instead of climatic variables to detect suitable habitats.

The use of remote sensing data provides some information on habitat, and in our case this information (forest habitats) was essential. The high resolution satellite images were analysed using efficient image treatment methods such as textural attributes. The models were also optimized by comparing the ability of the decision tree with that of the neural network and choosing the best suited for our objectives. The neural network appeared to be best suited for distinguishing tree crowns using shade (Figure 4) as well as for delineating the most suitable areas for coffee hybridization. The advantage is the possibility to choose the best function to fit the data and to obtain a continuous result. Neural networks are widely recognized as being efficient tools for ecological modelling (Lek, Guegan, 1999; Pearson *et al.*, 2002). At the same time, we considered decision trees to be suitable for models to characterize forest patterns, as we were working with classes (Vayssières *et al.*, 2000b).

Introducing remote sensing data has led to a real improvement in environmental characterization by providing part of habitat information (congruent with the prediction of bioclimatic variables) and allowing the use of complementary methods.

5.2 Environmental patterns of under forest coffee trees

Our study is the first attempt to characterize the forest canopy of the rustic coffee system using remote sensing tools (observation from above), instead of the more common approach from below based on density, tree richness or shade strata.

Our results confirm the importance of characterizing the forest patterns under which coffee trees grow. As expected, canopy heterogeneity and the size of the tree crowns played a significant role in predicting favourable habitats. The combined use of remotely sensed predictive variables and environmental data revealed the areas that are favourable for inter-specific

hybridization and the biological demands of coffee trees.

The decision tree applied to the whole study area clearly showed the implication of parameters in the classification process (Figure 8). The Info Gain selection method also provided interesting information that was congruent with the decision tree. The heterogeneity appeared to be the best explanatory variable. Hybrid zones could be characterized by different combinations of modalities, underlining the complexity of the environment concerned. For instance, in heterogeneous zones (heterogeneity > 1.5), with a majority of “large” and “medium” crowns (crown size > 2.5), the environmental areas that favour hybridization are either located on a slope > 53.5% between 194-429 m., or on a slope < 53.5% between 317-429 m (see other favourable environmental patterns in Figure 8).

The importance of using forest cover for prediction of the distribution of favourable areas for coffee species hybridization is the originality of this work. Our two remotely sensed predictive variables are both described as useful forest structure parameters since they are based on textural analyses known to detect remotely sensed objects precisely (Berberoglu *et al.*, 2007). Considering the results of studies that dealt with the impact of forest cover on biodiversity (Moguel, Toledo, 1999; Philpott *et al.*, 2008) and on the reduction of stress in coffee trees by improving adverse climatic conditions and nutritional imbalance (Beer *et al.*, 1997), characterization of the canopy cover appeared to be particularly important.

5.3 Coffee species hybridization sanctuaries

The most suitable areas (red zones in Figure 9) delineate very interesting areas with suitable environmental conditions for coffee hybridization. Figure 10 shows the three confidence levels for the presence of environments favourable for coffee inter-specific hybridization. The three classes correspond to habitats that are more or less favourable for coffee hybridization. The third class is the one of interest (red zones in

Figure 9), where the environmental conditions are the best suited to the dynamics of coffee genetic resources.

These geographic zones are exceptional since natural coffee inter-specific hybridization has been observed only once and very sporadically in Timor island (Cramer, 1957a). Given the originality of the Sarramea region with respect to the presence of hybrid zones, this result confirms the interest of focusing on this geographical area since a concentration of suitable local environmental parameters was detected. Even though we were working on a small geographical region, the heterogeneity of the environment was underlined highlighting the distribution of environmentally favourable niches.

These areas can be called favourable environmental niches (as defined by Hutchinson (1957) and reviewed recently by Pearman (2008b)) for coffee species sympatry and gene flow. Environmental niches coupled with investigations on fundamental niches are the keys to understanding the distribution of genetic diversity, through spatial distribution modelling (Guisan, Thuiller, 2005; Pearman *et al.*, 2008c).

The areas delineated as environmental niches for coffee inter-specific hybridization could represent an early stage in a shift in niche for the three coffee species concerned and above all for their inter-specific hybrids (Pearman *et al.*, 2008b). For this reason, these former coffee plantations should be conserved to allow the dynamics between species to proceed where the diversity of these three species is evolving naturally. This map is crucial for the conservation of coffee diversity.

Our study has highlighted environmental niches for the adaptation and hybridization of coffee species. This environmental description (Figure 8) could also be used to give advice on more natural and environment friendly ways of exploiting coffee.

Acknowledgements

The authors gratefully acknowledge the financial support of the DDR (*Direction du Développement rural*), Province Sud of New Caledonia (Convention 531-PVF/DDR). We would like also to thank UR LEGOS (IRD) for the use of their computer platform.

References

- Araujo, M.B., Pearson, R.G., Thuiller, W., & Erhard, M. (2005). Validation of species-climate impact models under climate change. *Global Change Biology*, 11, 1504-1513
- Asner, G.P., Palace, M., Keller, M., Pereira, R., Silva, J.N.M., & Zweede, J.C. (2002). Estimating canopy structure in an Amazon Forest from laser range finder and IKONOS satellite observations. *Biotropica*, 34, 483-492
- Bandeira, F.P., Martorell, C., Meave, J.A., & Caballero, J. (2005). The role of rustic coffee plantations in the conservation of wild tree diversity in the Chinantec region of Mexico. *Biodiversity and Conservation*, 14, 1225-1240
- Beer, J., Muschler, R., Kass, D., & Somarriba, E. (1997). Shade management in coffee and cacao plantations. *Agroforestry Systems*, 38, 139-164
- Berberoglu, S., Curran, P.J., Lloyd, C.D., & Atkinson, P.M. (2007). Texture classification of Mediterranean land cover. *International Journal of Applied Earth Observation and Geoinformation*, 9, 322-334
- Bishop, C. (1995). *Neural networks for pattern recognition*. New-York: Oxford Univ, Press
- Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Boca Raton FL: CRC Press
- Cordero-Sancho, S., & Sader, S.A. (2007). Spectral analysis and classification accuracy of coffee crops using Landsat and a topographic-environmental model. *International*

- Journal of Remote Sensing, 28, 1577-1593
- Couteron, P., Pelissier, R., Nicolini, E.A., & Dominique, P. (2005). Predicting tropical forest stand structure parameters from Fourier transform of very high-resolution remotely sensed canopy images. *Journal of Applied Ecology*, 42, 1121-1128
- Cramer, P.J.S. (1957). A review of literature of coffee research in Indonesia. *Inter American Institute of Agricultural Sciences*, W. F. L. Ed., 262
- Culvenor, D.S. (2002). TIDA: an algorithm for the delineation of tree crowns in high spatial resolution remotely sensed imagery. *Computers & Geosciences*, 28, 33-44
- Erikson, M. (2004). Species classification of individually segmented tree crowns in high-resolution aerial images using radiometric and morphologic image measures. *Remote Sensing of Environment*, 91, 469-477
- Gomez, C., Batti, A., Le Pierres, D., Campa, C., Hamon, S., de Kochko, A., Hamon, P., Huynh, F., Despinoy, M., & Poncet, V. (Submitted). Favorable habitats for *Coffea* interspecific hybridization in central New Caledonia: combined genetic and spatial analyses. *Journal of Applied Ecology*
- Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8, 993-1009
- Haralick, R.M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Trans. Syst. man Cybernet.*, 610-621
- Hernández-Martínez, G., Manson, R.H., & Contreras Hernández, A. (2009). Quantitative classification of coffee agroecosystems spanning a range of production intensities in central Veracruz, Mexico. *Agriculture, Ecosystems and Environment*, in press
- Hutchinson, G.E. (1957). Concluding remarks. *Cold Spring Harbour Symp. Quant. Biol.*, 22, 415-427
- Jagoret, P., Cilas, C., Eskes, A.B., & Charmetant, P. (1999). Caractérisation et évaluation des hybrides naturels de caféiers de Nouvelle-Calédonie. *Plantations, recherche, développement*, Septembre-Octobre
- Johansen, K., Coops, N.C., Gergel, S.E., & Stange, Y. (2007). Application of high spatial resolution satellite imagery for riparian and forest ecosystem classification. *Remote Sensing of Environment*, 110, 29-44
- Joshi, C., De Leeuw, J., Skidmore, A.K., van Duren, I.C., & van Oosten, H. (2006). Remotely sensed estimation of forest canopy density: A comparison of the performance of four methods. *International Journal of Applied Earth Observation and Geoinformation*, 8, 84-95
- Kayitakire, F., Hamel, C., & Defourny, P. (2006). Retrieving forest structure variables based on image texture analysis and IKONOS-2 imagery. *Remote Sensing of Environment*, 102, 390-401
- Kenji, K., & Larry, A.R. (1992). A Practical Approach to Feature Selection. In, *Proceedings of the Ninth International Workshop on Machine Learning: Morgan Kaufmann Publishers Inc.*
- Kier, G., Kreft, H., Lee, T.M., Jetz, W., Ibisch, P.L., Nowicki, C., Mutke, J., & Barthlott, W. (2009). A global assessment of endemism and species richness across island and mainland regions. *Proc Natl Acad Sci U S A*, 106, 9322-9327
- Le Pierres, D. (1999). Etude des hybrides naturels entre *Coffea arabica* et *Coffea canephora* de Nouvelle-Calédonie. *Montpellier, I.R.D. Nouméa*
- Lek, S., & Guegan, J.F. (1999). Artificial neural networks as a tool in

- ecological modelling, an introduction. *Ecological Modelling*, 120, 65-73
- Lin, B.B. (2007). Agroforestry management as an adaptive strategy against potential microclimate extremes in coffee agriculture. *Agricultural and Forest Meteorology*, 144, 85-94
- Lopez-Gomez, A.A., Williams-Linera, G., & Manson, R.H. (2008). Tree species diversity and vegetation structure in shade coffee farms in Veracruz, Mexico. *Agriculture Ecosystems & Environment*, 124, 160-172
- Mahe, L., Le Pierres, D., Combes, M.C., & Lashermes, P. (2007). Introgressive hybridization between the allotetraploid *Coffea arabica* and one of its diploid ancestors, *Coffea canephora*, in an exceptional sympatric zone in New Caledonia. *Genome*, 50, 316-324
- Mas, A.H., & Dietsch, T.V. (2004). Linking shade coffee certification to biodiversity conservation: Butterflies and birds in Chiapas, Mexico. *Ecological Applications*, 14, 642-654
- Mayaux, P., Richards, T., & Janodet, E. (1999). A vegetation map of Central Africa derived from satellite imagery. *Journal of Biogeography*, 26, 353-366
- Mendez, V.E., Gliessman, S.R., & Gilbert, G.S. (2007). Tree biodiversity in farmer cooperatives of a shade coffee landscape in western El Salvador. *Agriculture Ecosystems & Environment*, 119, 145-159
- Minsky, M., & Papert, S. (1969). *Perceptrons*. Cambridge (E.U.)
- Moguel, P., & Toledo, V.M. (1999). Biodiversity conservation in traditional coffee systems of Mexico. *Conservation Biology*, 13, 11-21
- Orka, H.O., Naesset, E., & Bollandsas, O.M. (2009). Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. *Remote Sensing of Environment*, 113, 1163-1174
- Pacifici, F., Chini, M., & Emery, W.J. (2009). A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. *Remote Sensing of Environment*, 113, 1276-1292
- Palace, M., Keller, M., Asner, G.P., Hagen, S., & Braswell, B. (2008). Amazon forest structure from IKONOS satellite data and the automated characterization of forest canopy properties. *Biotropica*, 40, 141-150
- Pearman, P.B., Guisan, A., Broennimann, O., & Randin, C.F. (2008a). Niche dynamics in space and time. *Trends in Ecology & Evolution*, 23, 149-158
- Pearman, P.B., Randin, C.F., Broennimann, O., Vittoz, P., van der Knaap, W.O., Engler, R., Le Lay, G., Zimmermann, N.E., & Guisan, A. (2008b). Prediction of plant species distributions across six millennia. *Ecology Letters*, 11, 357-369
- Pearson, R.G., Dawson, T.P., Berry, P.M., & Harrison, P.A. (2002). SPECIES: A Spatial Evaluation of Climate Impact on the Envelope of Species. *Ecological Modelling*, 154, 289-300
- Pearson, R.G., Dawson, T.P., & Liu, C. (2004). Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, 27, 285-298
- Perfecto, I., Mas, A., Dietsch, T., & Vandermeer, J. (2003). Conservation of biodiversity in coffee agroecosystems: a tri-taxa comparison in southern Mexico. *Biodiversity and Conservation*, 12, 1239-1252
- Pezzopane, J.R.M., Pedro, M.J., de Camargo, M.B.P., & Faztioli, L.C. (2008). Heat requirements of Mundo Novo coffee for the flowering-harvest phenological stage. *Ciencia E Agrotecnologia*, 32, 1781-1786

- Philpott, S.M., Arendt, W.J., Armbrrecht, I., Bichier, P., Diestch, T.V., Gordon, C., Greenberg, R., Perfecto, I., Reynoso-Santos, R., Soto-Pinto, L., Tejeda-Cruz, C., Williams-Linera, G., Valenzuela, J., & Zolotoff, J.M. (2008). Biodiversity Loss in Latin American Coffee Landscapes: Review of the Evidence on Ants, Birds, and Trees. *Conservation Biology*, 22, 1093-1105
- Philpott, S.M., Perfecto, I., & Vandermeer, J. (2006). Effects of management intensity and season on arboreal ant diversity and abundance in coffee agroecosystems. *Biodiversity and Conservation*, 15, 139-155
- Pouliot, D.A., King, D.J., Bell, F.W., & Pitt, D.G. (2002). Automated tree crown detection and delineation in high-resolution digital camera imagery of coniferous forest regeneration. *Remote Sensing of Environment*, 82, 322-334
- Quinlan, J.R. (1993). C4.5.: Programs for machine learning. Morgan Kaufmann Publishers
- Read, J.M., Clark, D.B., Venticinque, E.M., & Moreira, M.P. (2003). Application of merged 1-m and 4-m resolution satellite data to research and management in tropical forests. *Journal of Applied Ecology*, 40, 592-600
- Rouse, J., Haas, R., Schell, J., Deering, D., & Harlan, J. (1974). Monitoring the vernal advancement of retrogradation of natural vegetation. In T.I. NASA / GSFC, Final Report (Ed.) (p. 371). Greenbelt, MD, US
- Solis-Montero, L., Flores-Palacios, A., & Cruz-Angon, A. (2005). Shade-coffee plantations as refuges for tropical wild orchids in central Veracruz, Mexico. *Conservation Biology*, 19, 908-916
- Soto-Pinto, L., Romero-Alvarado, Y., Caballero-Nieto, J., & Segura Warnholtz, G. (2001). Woody plant diversity and structure of shade-grown-coffee plantations in northern Chiapas, Mexico. *Rev Biol Trop*, 49, 977-987
- Soto-Pinto, L., Villalvazo-Lopez, V., Jimenez-Ferrer, G., Ramirez-Marcial, N., Montoya, G., & Sinclair, F.L. (2007). The role of local knowledge in determining shade composition of multistrata coffee systems in Chiapas, Mexico. *Biodiversity and Conservation*, 16, 419-436
- Thuiller, W., Araujo, M.B., & Lavorel, S. (2004). Do we need land-cover data to model species distributions in Europe? *Journal of Biogeography*, 31, 353-361
- Vayssières, M.P., Plant, R.E., & Allen-Diaz, B.H. (2000). Classification trees: an alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science*, 11, 679-694
- Whitmore, T., C. (1982). On pattern and process in forests. In E.I. Newman (Ed.), *The plant community as a working mechanism* (pp. 45-59). Oxford, UK: Black-well Scientific Publications
- Witten, I., H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. San Francisco: Morgan Kaufman, 2005
- Zimmermann, N.E., Edwards, T.C., Moisen, G.G., Frescino, T.S., & Blackard, J.A. (2007). Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *Journal of Applied Ecology*, 44, 1057-1067

Table 1. Remote sensing database

Images	Description	Source
Quickbird Image 1 West	18/06/05 Lat/Long: 21.68143' / 165.91898' Spatial resolution: 0.66 Cloud Cover: 0% Off Nadir Angle: 22° Target Azimuth: 315° Environment Quality: 99 Size in pixels: 20176*15101	DigitalGlobe http://digitalglobe.com
Quickbird Image 2 East	16/08/05 Lat/Long: 21.62288' / 165.76205' Spatial resolution: 0.66 Cloud Cover: 0% Off Nadir Angle: 7° Target Azimuth: 7° Environment Quality: 99 Size in pixels: 19747*14448	DigitalGlobe http://digitalglobe.com

Table 2. Confusion matrices (percentage of accuracy) for the canopy heterogeneity model and the two subsets: Farino and Sarramea. For both subsets, the first matrix values were calculated on all pixels (1- All Pixels). The accuracy of the model was calculated by summing the well-classified values (in grey cells) of the first matrix. The second matrix gives the score by classes (2-By classes).

FARINO		OBSERVED CLASSES	
CLASSIFIED CLASSES	1- All pixels	Heterogeneous	Homogeneous
	Heterogeneous	61.06	0.45
	Homogeneous	6.84	31.66
	2- By classes	Heterogeneous	Homogeneous
	Heterogeneous	99.27	0.73
	Homogeneous	17.76	82.24

SARRAMEA		OBSERVED CLASSES	
CLASSIFIED CLASSES	1- All pixels	Heterogeneous	Homogeneous
	Heterogeneous	24.13	29.66
	Homogeneous	11.85	63.72
	2- By classes	Heterogeneous	Homogeneous
	Heterogeneous	98.79	1.21
	Homogeneous	15.69	84.31

Table 3. Confusion matrices (percentage of accuracy) for the tree crown size model and the two subsets Farino and Sarramea. For both subsets, the first matrix values were calculated on all pixels (1- All Pixels). The accuracy of the model was calculated by summing the well-classified values (in grey) of the first matrix. The second matrix gives the score by classes (2-By classes).

FARINO		OBSERVED CLASSES				
CLASSIFIED CLASSES	1- All pixels	No crowns	Small	Med-small	Medium	Large
	No crowns	24.41	0.35	0.38	0.36	0.17
	Small	0.88	9.46	1.20	0.65	0.19
	Med-small	1.14	2.12	20.78	1.85	0.76
	Medium	0.91	1.35	1.80	19.17	0.65
	Large	0.33	0.38	0.73	0.63	9.34
	2- By classes	No crowns	Small	Med-small	Medium	Large
	No crowns	95.10	1.38	1.47	1.38	0.67
	Small	7.08	76.45	9.67	5.26	1.53
	Med-small	4.26	7.95	77.98	6.96	2.85
	Medium	3.81	5.65	7.55	80.26	2.74
	Large	2.92	3.35	6.35	5.55	81.82

SARRAMEA		OBSERVED CLASSES				
CLASSIFIED CLASSES	1- All pixels	No crowns	Small	Med-small	Medium	Large
	No crowns	10.63	0.40	0.50	0.30	0.10
	Small	0.66	10.38	1.74	0.94	0.35
	Med-small	1.09	3.07	26.41	2.38	1.08
	Medium	0.60	1.94	2.50	20.41	0.75
	Large	0.26	0.84	1.02	0.59	11.06
	2- By classes	No crowns	Small	Med-small	Medium	Large
	No crowns	89.10	3.36	4.23	2.49	0.82
	Small	4.70	73.76	12.39	6.66	2.50
	Med-small	3.20	9.03	77.60	6.99	3.18
	Medium	2.30	7.40	9.54	77.89	2.88
	Large	1.86	6.11	7.39	4.32	80.31

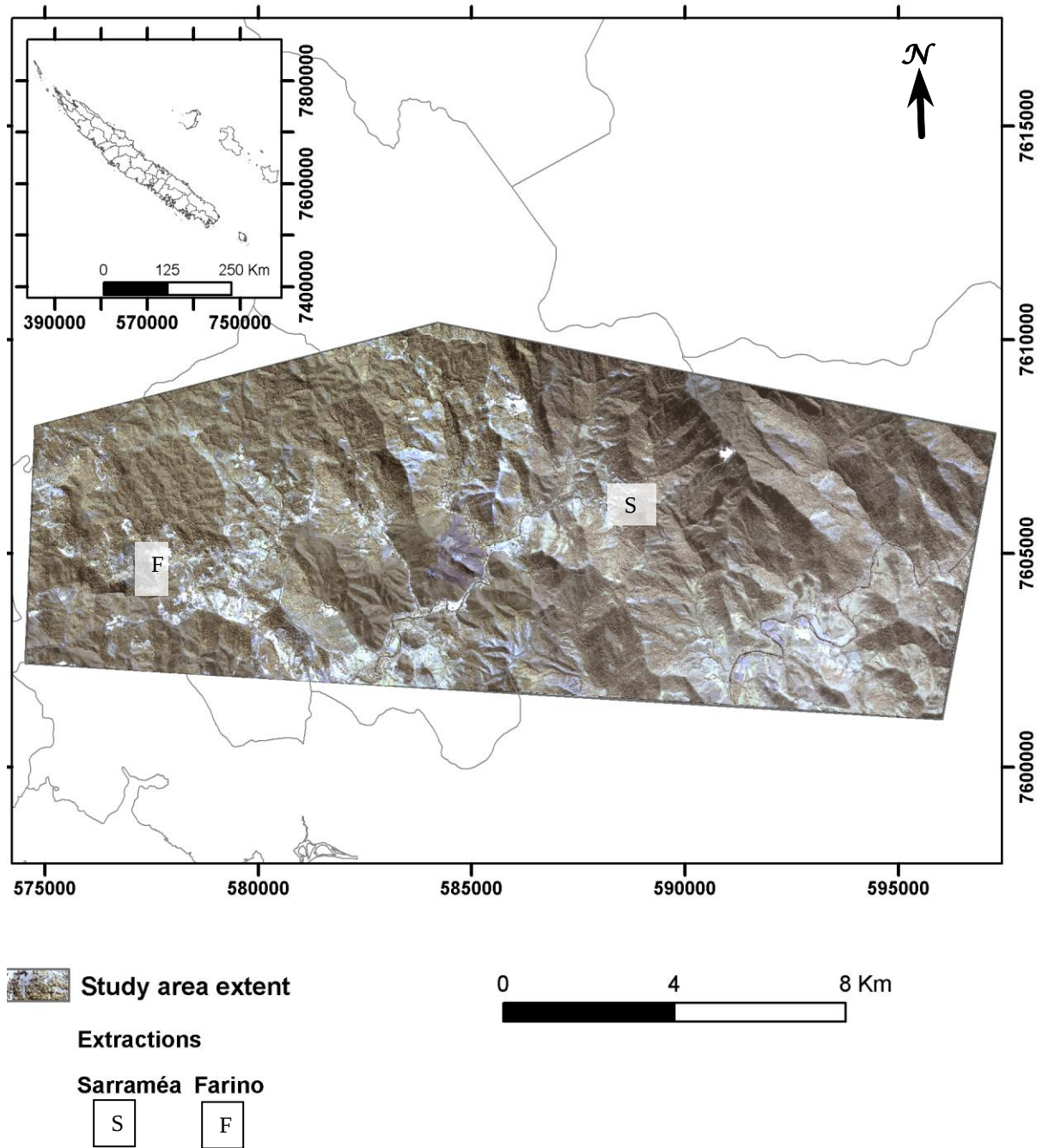


Figure 1. Map of the study area in New Caledonia and details of the study area delimited by the Quickbird satellite images geo-referenced by the tracking GPS. The extents of the two extractions (Farino and Sarramea) used for model calibration are shown.

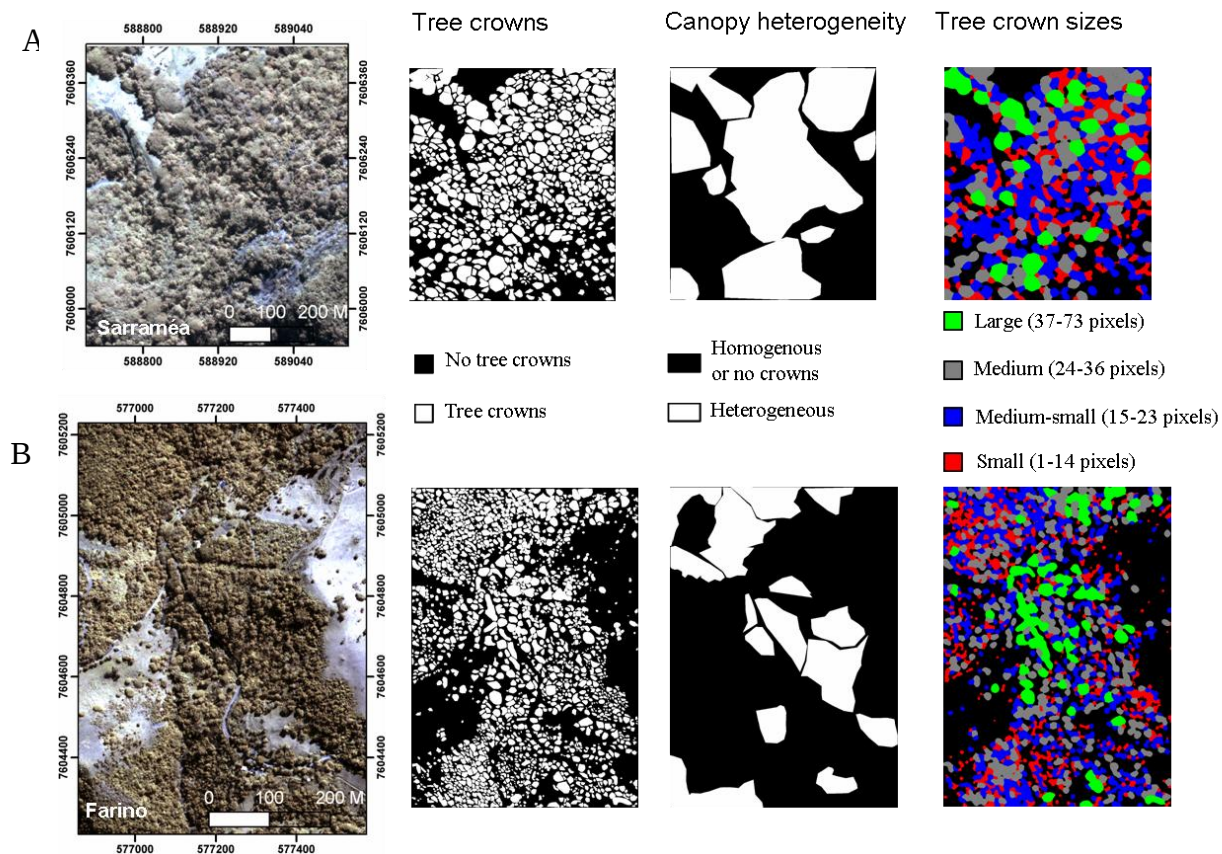


Figure 2. The three learning data sets were visually determined on both the training subsets based on Sarramea and Farino. The tree crowns and the heterogeneous zones were delineated by hand. The tree crown size learning data set was based on delineation of the tree crowns, and the four size classes were determined manually.

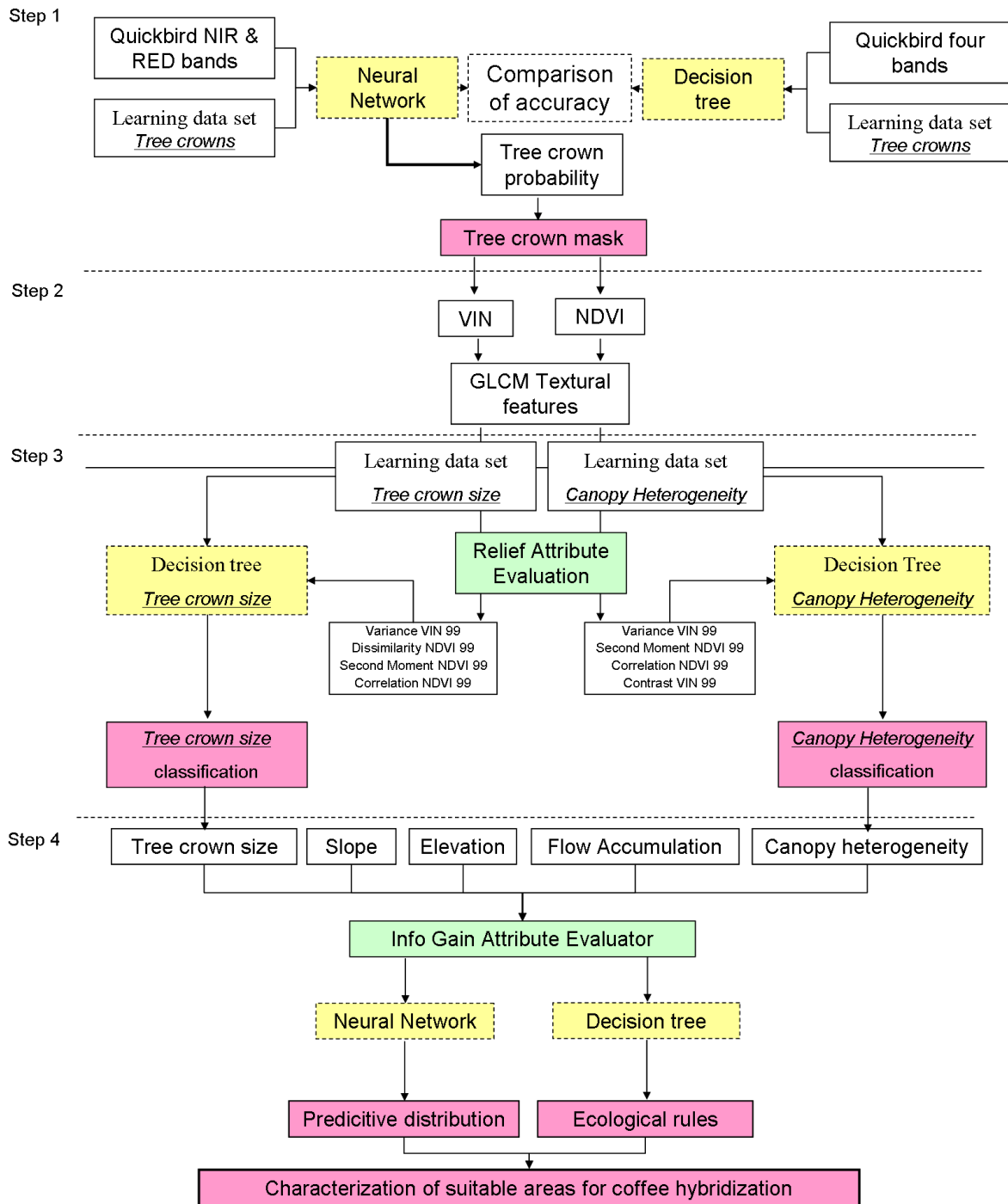


Figure 3. Flow chart of major steps in the classification of results to obtain the two remotely sensed predictors. The yellow boxes with a dotted line are the mathematical models used. The two green boxes show the attribute selection method. The pink boxes represent the main results while the other boxes are either input or learning data sets as annotated.

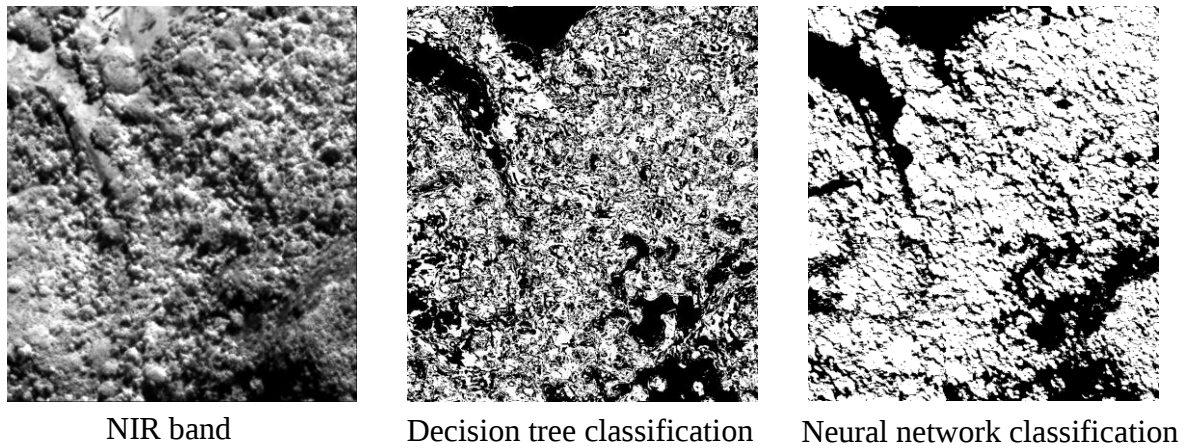


Figure 4. Comparison of the tree crown mask obtained with two different models: a decision tree and a neural network. The neural network (79.15% of accuracy) was more realistic than the decision tree (61% of accuracy).

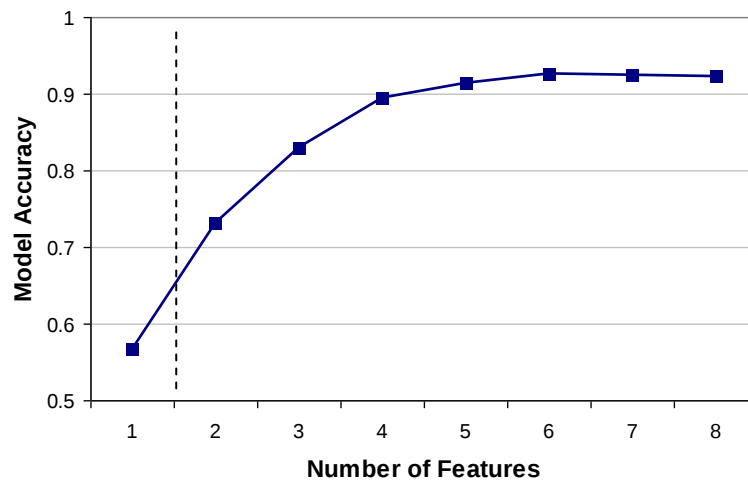


Figure 5. Model accuracy according to the number of features used as inputs of the model. This test was done for the canopy heterogeneity model and the Sarramea Extraction. 90% of accuracy was reached for four features.

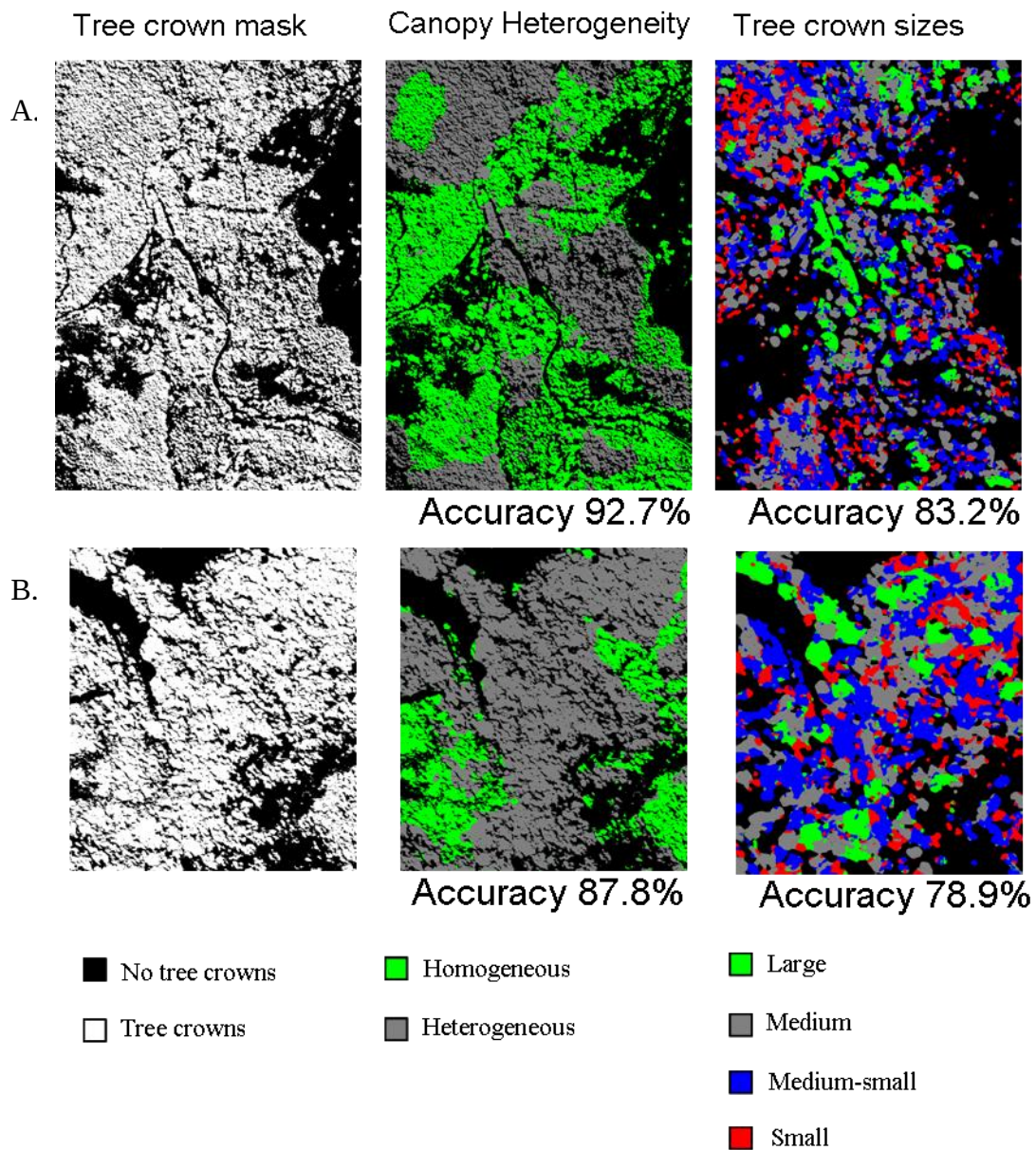


Figure 6. Classification results for the three models (tree crown discrimination, canopy heterogeneity and crown sizes) and accuracy obtained for both the extractions of Farino (A.) and Sarramea (B.).

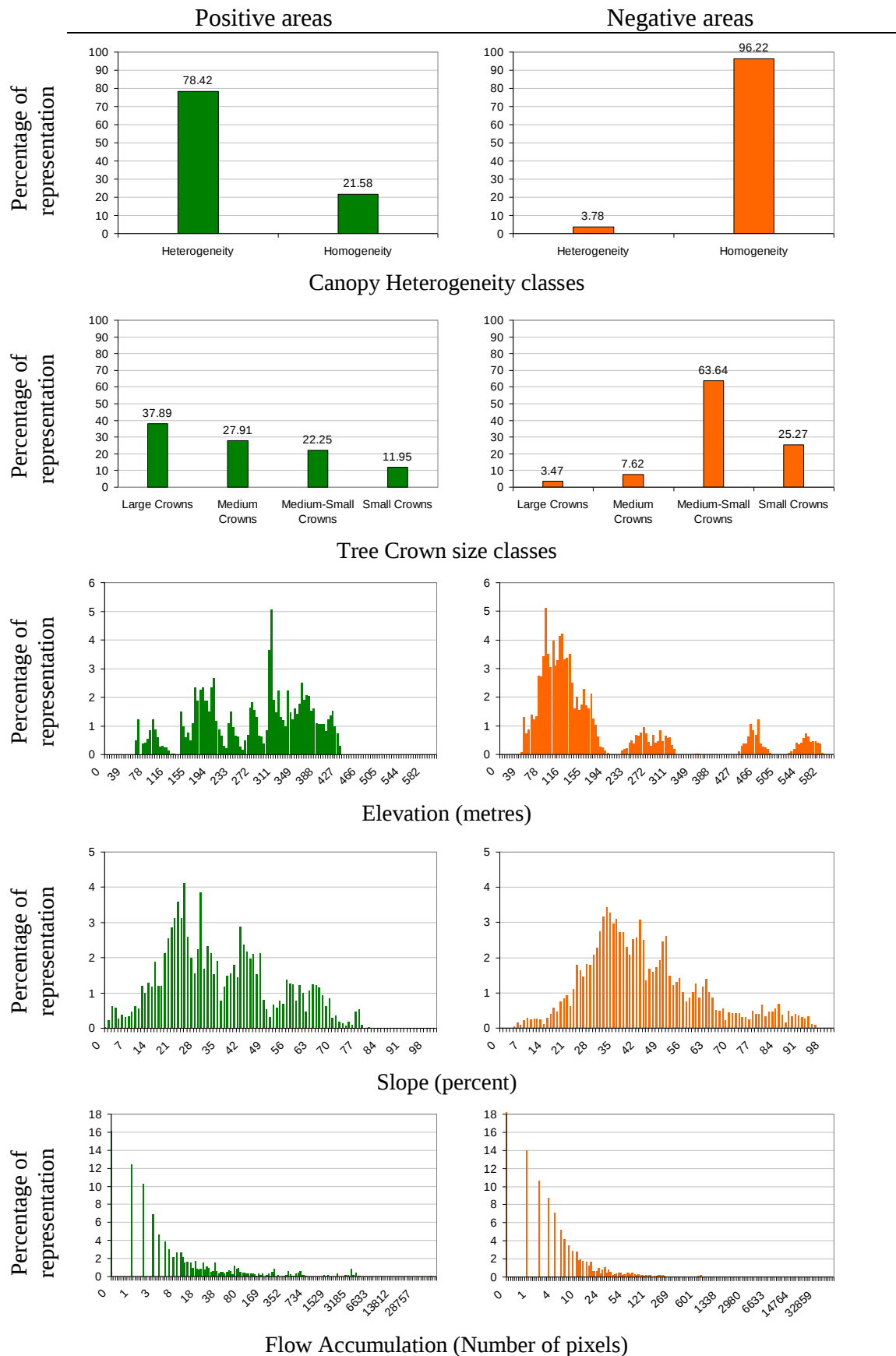


Figure 7. Range of values of each layer on the positive and the negative areas. In all graphs, the y axis is the percentage of representation of each x value. The x axis shows the layer values: in meters for the DEM, in percent for the slope, and in number of pixels for the flow accumulation.

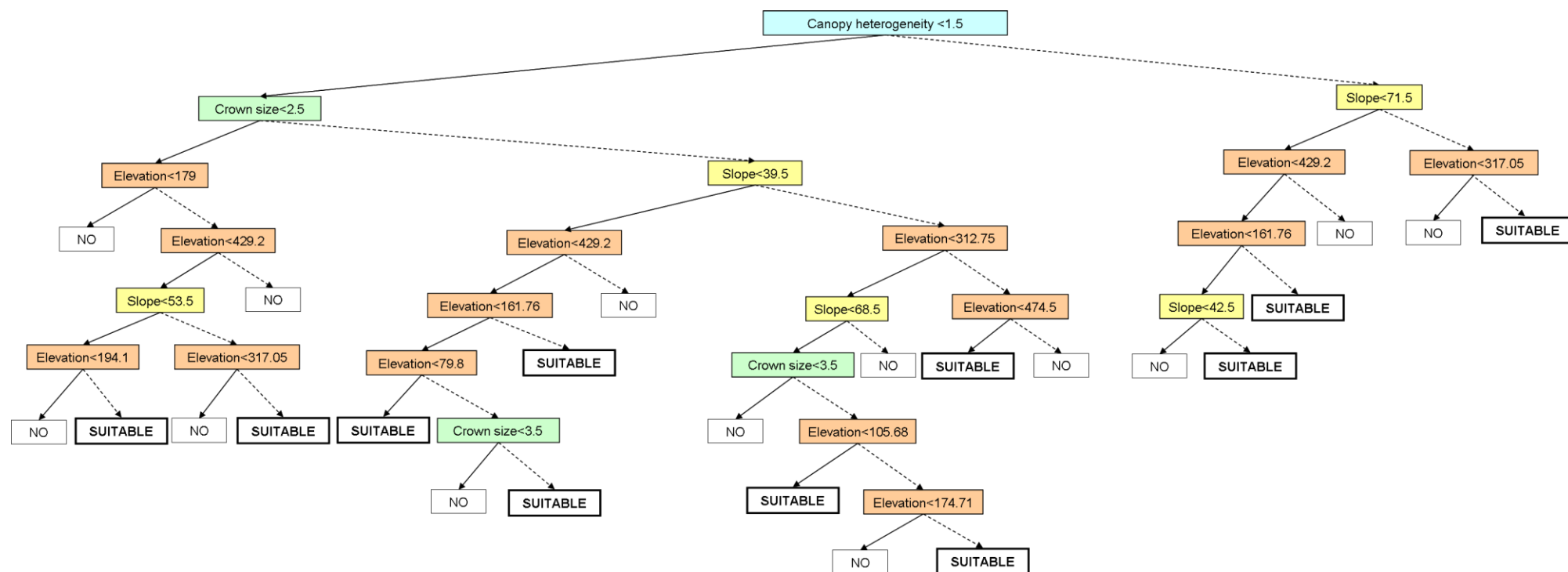


Figure 8. Representation of the pruned decision tree implemented to detect areas suitable for inter-specific hybridization between coffee species. Each predictive variable is associated with a colour. If the condition at a node is respected, the result or next node is shown by a continuous line (on the left), if not, it is shown by a dotted line (on the right).

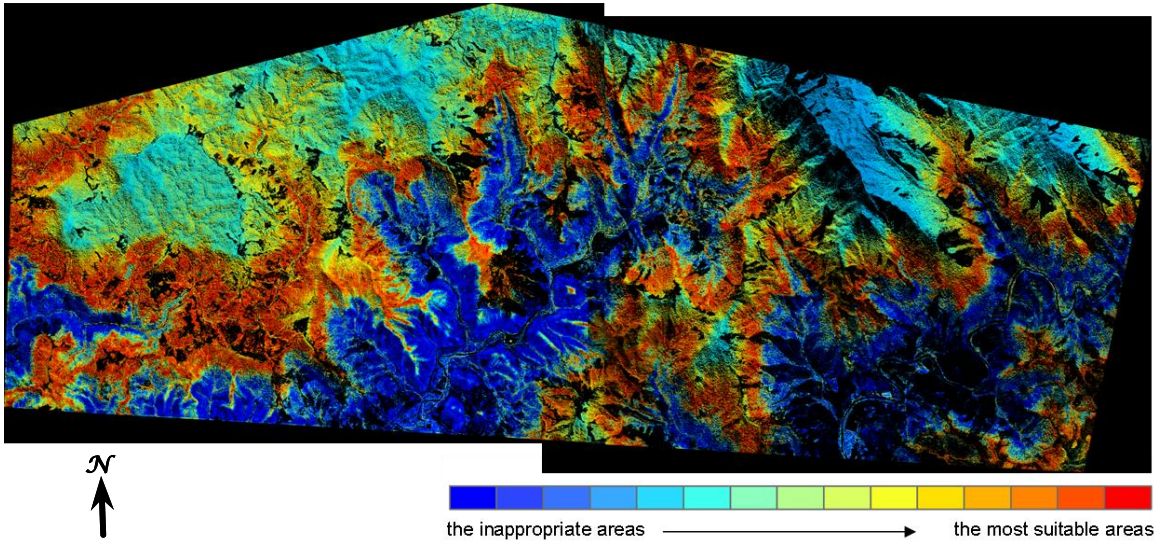


Figure 9. Predictive distribution map of areas suitable for coffee inter-specific hybridization using five features (Canopy Heterogeneity, Crown sizes, Elevation, Slope and Flow Accumulation) and a neural network model. The colour gradient goes from blue (unsuitable areas) to red (most suitable areas).

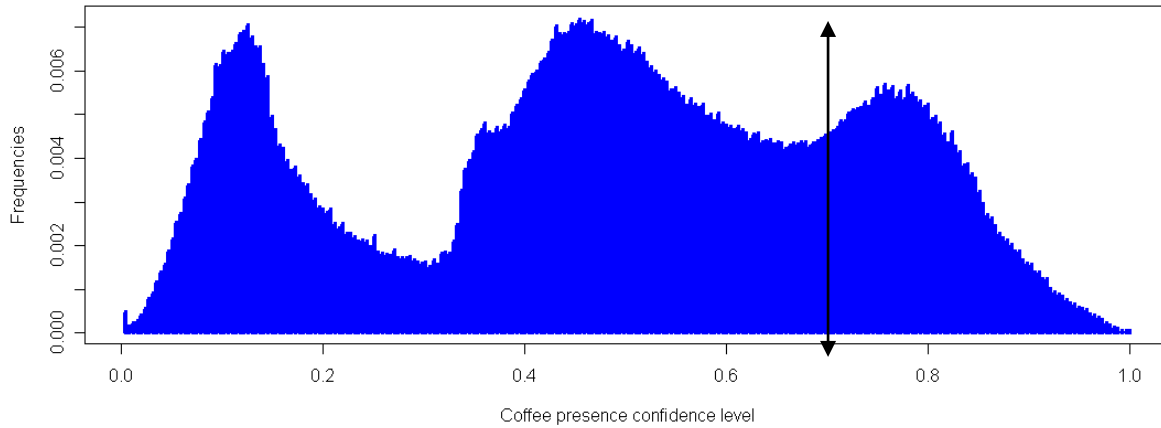


Figure 10. Distribution of the confidence level of the presence of coffee as detected in Figure 9 for forest areas only (49.23% of the pixels in the whole image). Very favourable areas for coffee interspecific hybridization are represented by a confidence level above 0.7 (the third peak on the curve).

III – Conclusion

Le principal objectif de ce chapitre a été de proposer une carte de prédiction de distribution des zones favorables aux zones hybrides *via* la caractérisation précise du couvert forestier par télédétection. Une attention particulière a été portée sur le choix des attributs en entrée de chacun des modèles.

En introduisant la caractérisation fine du couvert forestier dans le modèle de prédiction, toutes les composantes environnementales contribuant à la présence de micro-habitats favorables (niches environnementales) aux flux de gènes entre caféiers (par la levée de barrières à la reproduction principalement) sont prises en compte. L'arbre de décision construit et validé à 96,9%, permet d'illustrer les conditions favorables ou non aux hybridations inter-spécifiques. Cet arbre de décision témoigne de la complexité des combinaisons de paramètres et de la forte hétérogénéité des habitats. Les niches favorables sont définies par des compromis entre exigences environnementales et conditions favorables.

Ce travail a permis de déterminer finement l'ampleur du phénomène observé. Les résultats du modèle de prédiction des niches environnementales favorables montrent que ce sont des microenvironnements particuliers au milieu de la forêt dense humide de la région centrale de Nouvelle-Calédonie. Ces microenvironnements sont très favorables aux hybridations inter-spécifiques et à la culture du caféier de manière générale, pour les deux, voir les trois espèces.

Chapitre 6 – Conclusion générale



Les travaux menés pendant cette thèse sont partis du constat tout à fait original de l'existence d'hybridations inter-spécifiques spontanées entre trois espèces du genre *Coffea* (*C. arabica*, *C. canephora*, *C. liberica*) suite à leur introduction en Nouvelle-Calédonie. Afin d'étudier cette situation aussi particulière que rare, des travaux de recherche, explorant trois approches scientifiques différentes et complémentaires, ont été menés : l'étude de la dynamique génétique des hybrides, l'étude des facteurs environnementaux favorables à de telles hybridations et l'étude de leur répartition spatiale. Ces études ont permis d'apporter des éléments de réponse aux questions fondamentales abordées dans l'introduction et l'état de l'art.

I – Synthèse et discussion

A notre connaissance, aucune étude n'avait été faite auparavant sur le devenir de caféiers après l'abandon des cultures et en particulier sur le devenir d'espèces de caféiers évoluant naturellement en situation de sympatrie. Sur le site d'étude de Sarraméa, les conditions environnementales ont autorisé des flux de gènes spontanés entre espèces, phénomène original chez les caféiers.

D'un point de vue génétique, l'ampleur des flux de gènes et du brassage génétique peut être plus ou moins limitée. Quel est le niveau de diversité introduit par rapport à la diversité connue de l'espèce dans son milieu naturel ? Est-ce que les flux de gènes vont aboutir à des hybrides viables et fertiles ? Quelle est l'ampleur de ces hybridations en termes de mélange des génomes ?

Dans le cas de notre étude, la caractérisation de la diversité génétique globale des trois espèces étudiées a été réalisée afin d'évaluer la part de la diversité génétique introduite dans la population de Sarraméa en Nouvelle-Calédonie et d'y estimer les flux de gènes.

Afin d'établir une comparaison rigoureuse des niveaux de diversité, le choix de l'échantillon de référence est important. L'étude de la diversité africaine des trois espèces a donc été menée, à l'aide de marqueurs microsatellites, afin de mettre au point de véritables collections de référence représentatives du polymorphisme sauvage de l'espèce. Un important travail d'analyse de diversité a été mené pour l'espèce *C. canephora*, dont la diversité naturelle n'avait pas encore été évaluée avec des marqueurs microsatellites. Ce travail a permis de comprendre la structuration de

la diversité de manière précise. A partir de ces connaissances, nous avons construit une collection de référence qui inclut le maximum de diversité de l'échantillon global dans un échantillon de taille minimale, en évitant les redondances. La diversité introduite à Sarraméa a été ensuite évaluée par des analyses d'assignation génétiques bayésiennes. Il apparaît que la diversité introduite est élevée avec une valeur de PIC (Polymorphism Information Content) de 0,61 en moyenne pour les 92 individus de Sarraméa génétiquement caractérisés comme appartenant à une des trois espèces (non hybrides).

Les caféiers ont été introduits dans toute la zone intertropicale de la planète, où les conditions environnementales sont *a priori* favorables à leur culture. Pourtant, les niveaux de diversité introduits, ne sont pas toujours aussi élevés qu'à Sarraméa. Par exemple, la diversité génétique de l'espèce *C. canephora* introduite en Inde a été étudiée (Prakash *et al.*, 2005). Les résultats montrent que la diversité présente dans la collection de référence utilisée (62 allèles révélés) est supérieure à celle des échantillons indiens testés (51 et 44 allèles révélés).

Lors de l'introduction d'une espèce dans un nouvel écosystème, plusieurs facteurs peuvent influencer la perte ou le maintien de la diversité entre les populations naturelles et les populations introduites. Novak et Mack (1993) ont ainsi observé une diminution de la variabilité génétique sur l'ensemble des populations introduites d'une herbacée *Bromus tectorum* (en Amérique du nord) mais également une augmentation de la variabilité au sein de chaque population par rapport aux populations naturelles (en Eurasie et en Afrique). Ces résultats suggèrent à la fois un fort effet de fondation dû à un goulot d'étranglement mais aussi des introductions multiples qui ont renforcé la diversité intra-population.

A Sarraméa, le haut niveau de diversité génétique est peut être dû à des introductions multiples. En effet, l'histoire de la caféiculture dans la région semble indiquer qu'il y a eut plusieurs vagues d'introduction de plants (Atlas de la Nouvelle-Calédonie, 1981). A l'issue de l'analyse de la diversité introduite, un potentiel génétique apparaît intéressant pour les hybridations inter-spécifiques spontanées. L'ampleur génétique des hybridations spontanées est évaluée au travers du taux d'hybridation dans la population à un instant donné et sur un échantillon de 95 individus viables et fertiles (qui fleurissent et fructifient). Le taux d'hybridation révélé est de 3%. Deux hybrides inter-spécifiques *C. canephora* X *C. liberica*, retro-croisés avec *C. canephora* ont été mis en évidence avec 32% et 28% de leur génome issu de *C. liberica*, respectivement. Un hybride *C. arabica* X *C. canephora* retro-croisé avec *C. arabica* a aussi été mis en évidence avec 34% de

son génome issu de *C. canephora*. Aucun hybride entre *C. arabica* et *C. liberica* n'a été caractérisé. Or quand on observe de plus près les séquences de floraisons observées sur l'année 2008 et décrites dans le tableau 4.1 et sur la figure 4.5, il n'apparaît aucune superposition des stades fertiles de ces dernières (stade 4 et stade 5, émetteur/récepteur et récepteur de pollen respectivement). Ceci expliquerait l'absence d'échange de pollen effectif entre ces espèces, dans nos observations.

L'existence des croisements spontanés entre espèces est liée aux phénologies florales et par conséquent aux conditions climatiques de la région. Pourtant, la nature et le sens des croisements ne dépendent pas seulement de la dynamique des phénologies mais également des autres barrières à la reproduction telles que la stérilité des hybrides. Aucun hybride *C. canephora* X *C. arabica* retro-croisé sur *C. canephora*, n'a été observé ni dans cette étude, ni auparavant, ce qui pourrait être dû en particulier à l'existence de fortes barrières post-zygotiques. Ces dernières n'ont pas été étudiées dans cette thèse, mais l'impact de l'environnement en général sur le phénomène d'hybridation spontanée observé est mis en lumière.

D'un point de vue écologique, lors d'une introduction, une modification des niches écologiques se produit. Quelles sont les conséquences de cette modification sur les traits d'histoire de vie des espèces ou sur leur potentiel d'adaptation au nouvel environnement ? Comment l'environnement agit-il sur la levée de barrières à la reproduction entre espèces originellement isolées ?

Dans notre étude, les différences environnementales entre les milieux d'origine des espèces (en Afrique) et la région de Sarraméa en Nouvelle-Calédonie ont été étudiées. L'importance des caractéristiques environnementales de la région d'intérêt a été soulignée en décrivant les niches favorables aux flux de gènes.

Il s'agit de comprendre quels sont les paramètres environnementaux favorables aux flux de gènes. Or, l'étude des flux de gènes inter-spécifiques permet une meilleure compréhension de l'implication de facteurs écologiques sur les niveaux d'hybridation. Par exemple, Gow et al. (2006) ont montré des différences de taux d'hybridation entre plusieurs espèces d'épinoches *Gasterosteus* sp. en fonction du lieu d'échantillonnage et de l'adaptation phénotypique. Par ailleurs, il a été montré que les pressions anthropiques sur les habitats naturels ne cessent

d'augmenter les possibilités d'hybridation entre espèces originellement isolées (Milne *et al.*, 1999) et notamment par des modifications des phénologies (Lamont *et al.*, 2003).

Dans notre cas, les caféiers ont été introduits pour leur exploitation, mais les cultures traditionnelles sous couvert ont été progressivement abandonnées. Le contact entre les différentes espèces a été provoqué initialement par les activités de l'homme, mais les conditions environnementales particulières de la région de Sarraméa ont permis la mise en place d'une dynamique d'hybridation au niveau des niches favorables. Lors de l'introduction d'une espèce dans un nouvel environnement, la niche réalisée est inévitablement différente de celle de son milieu d'origine (Maron *et al.*, 2004 ; Pearman *et al.*, 2008a). Ainsi, l'introduction des trois espèces de caféiers en Nouvelle-Calédonie a induit une modification des niches environnementales auxquelles elles étaient originellement confrontées. La situation à Sarraméa illustre comment les espèces, alors contraintes à cohabiter dans le même milieu, ont survécu.

Dans notre étude, la comparaison avec le milieu d'origine des espèces concernées a permis de mettre en exergue les paramètres environnementaux importants. En effet, l'expertise environnementale mise en place à l'échelle de la région d'étude, a révélé des différences très marquées quant au régime climatique annuel entre Sarraméa et les régions d'origine des espèces (Côte d'Ivoire pour *C. canephora* et *C. liberica* et l'Ethiopie pour *C. arabica*). Cependant, la comparaison des climogrammes révèle que celui de Sarraméa chevauche ceux de Divo (Côte d'Ivoire) et Jimma (Ethiopie), montrant ainsi que les conditions climatiques générales de Sarraméa permettent de répondre aux exigences climatiques des trois espèces.

Des paramètres déterminants, tels que l'altitude et la latitude, sont très différents pour les conditions optimales de croissance des espèces de caféiers. *C. arabica* nécessite des températures froides la nuit, elle ne pousse pas naturellement en dessous de 1000 mètres d'altitude en Afrique. Pourtant, on la trouve en Nouvelle-Calédonie à Sarraméa, sans intervention particulière de l'homme (intrants, soins), à des altitudes inférieures à 500 mètres. Il semblerait que la différence de latitude compense la différence d'altitude. En effet, les températures enregistrées à Sarraméa peuvent être très faibles la nuit (souvent inférieures à 10°C), comme cela se produit près de l'équateur à haute altitude.

L'équilibre entre drainage et humidité (pente et réseau hydrographique) est un paramètre qui est apparu très important, illustré au travers de la distance aux creeks (réseaux hydrographiques temporaires) et l'accumulation des flux dans les différents modèles. Par ailleurs, l'influence d'un

couvert forestier dense a des impacts directs sur la lumière, sur la mise en place d'un microclimat protecteur et sur les fluctuations du taux d'humidité relative (Lin, 2007).

L'impact de l'environnement et en particulier de la pluviométrie sur les phénologies florales des trois espèces a été observé par l'analyse d'une saison de floraison. La modification des niches, caractérisée par le détail des influences environnementales, a favorisé cette levée de barrières à la reproduction. Afin de pouvoir parler d'une véritable description de niches environnementales, il a été nécessaire de travailler à une échelle plus fine permettant de révéler toute la richesse de l'hétérogénéité des habitats et prédire finement la distribution des zones favorables.

***D'un point de vue spatial** les hybridations peuvent être limitées à certaines niches particulières, ou plus ou moins généralisées à une région entière. Quelle est la distribution des espèces dans leur nouvel environnement ? Quelle est la distribution des niches favorables à la levée de barrières aux flux de gènes ? Quelle est l'ampleur géographique de ces hybridations ?*

***Dans notre étude**, afin d'avoir une vision globale, il a été important d'analyser cette distribution à l'échelle de la région d'intérêt entière. Cette prédiction de distribution a été réalisée par une description fine du couvert forestier sous lequel poussent les caféiers.*

Les paramètres environnementaux les plus utilisés pour prédire la distribution des espèces ont jusqu'à présent été bioclimatiques (Pearson et Dawson, 2003; Zimmermann *et al.*, 2007). En effet, le climat est très corrélé à la situation géographique de la région étudiée et explique une grande part de la distribution des espèces (Araujo et Pearson, 2005). Or, aujourd'hui, l'amélioration des résolutions des images satellitaires, ainsi que des méthodes de traitement d'images permettent une cartographie fine des habitats et de leur hétérogénéité. Ainsi la télédétection peut dans certains cas où la variabilité climatique n'est pas significative, remplacer les prédictions bioclimatiques (Pearson et Dawson, 2003; Zimmermann *et al.*, 2007).

Dans notre cas, les paramètres climatiques ont été exploités uniquement à large échelle spatiale pour mettre en évidence les possibilités de flux de pollen entre espèces *via* les floraisons et confirmer la région d'étude de Sarraméa (allant de Farino à Dogny) (Fig.1, chapitre 4) comme foyer principal d'intérêt pour les hybridations inter-spécifiques naturelles entre caféiers. Par la

suite, un modèle statistique complexe de prédiction de distribution des niches environnementales favorables a été élaboré, basé sur des analyses fines d'images satellites à Très Haute Résolution (THR) spatiale (images Quickbird à 0,6 m). L'expertise botanique a décrit le couvert forestier selon des attributs particuliers et nous a permis de travailler sur les images satellites THR afin de mettre au point des variables prédictives caractérisant ces attributs. La couverture forestière, est d'une importance particulière car dans notre cas, les anciennes plantations de caféiers étudiées sont sous couvert et ne sont donc pas détectables directement par télédétection. Ainsi, la caractérisation fine de la composition spécifique du couvert forestier au dessus des anciennes plantations de caféiers est un élément important à prendre en compte.

La composition d'une canopée est un des éléments principaux à connaître dans la gestion des ressources forestières. Deux méthodes sont principalement utilisées pour caractériser la canopée à partir d'images satellitaires : la segmentation de l'image afin de délimiter les houppiers de chaque arbre (Culvenor, 2002 ; Erikson, 2004, Palace *et al.*, 2008) et l'utilisation d'approches texturales directes afin de caractériser des ensembles forestiers selon des attributs particuliers (Johansen *et al.*, 2007 ; Kayitakire *et al.*, 2006). Ainsi, les images satellitaires THR sont utilisées pour des études détaillées de la structure des forêts dans l'objectif de produire un inventaire des espèces présentes ou des ensembles forestiers afin de mieux gérer les ressources forestières (Couteron *et al.*, 2005 ; Johansen *et al.*, 2007, Palace *et al.*, 2008).

L'étude environnementale réalisée a ainsi permis d'orienter les traitements d'images afin de caractériser au mieux cette couverture végétale spécifique. En effet, cette étude a conduit à la détermination de deux variables prédictives, à savoir l'hétérogénéité du couvert et la taille des houppiers. Ces deux variables ont été sélectionnées pour caractériser les deux attributs principaux permettant de discriminer le couvert forestier que l'on retrouve au dessus des anciennes plantations de caféiers, c'est-à-dire les forêts primaires très denses avec ajout de légumineuses spécifiques caractérisées par leur large port en parasol. Les classifications basées sur des approches texturales, ont été validées à 98,25% et 81,05% pour l'hétérogénéité de la canopée et la taille des houppiers respectivement.

Il existe un grand nombre de modèles mathématiques permettant de prédire la distribution d'espèces avec plusieurs paramètres environnementaux. L'efficacité des méthodes non linéaires du type réseaux de neurones dans les prédictions écologiques a été plusieurs fois démontrée (Lek, Guegan, 1999 ; Pearson *et al.*, 2002). En parallèle, les arbres de décisions possèdent des qualités

intéressantes en termes de développement et d'interprétation (Vayssières *et al.*, 2000) que d'autres modèles tels que des régressions linéaires ou non linéaires n'ont pas.

Un deuxième aspect original de notre travail est l'étude statistique des relations entre variables environnementales et hybridation des caféiers et la mise au point de deux modèles différents et complémentaires basés sur celle-ci, un arbre de décision et un réseau de neurones, permettant la prédiction de distribution des zones favorables à l'hybridation. L'utilisation de ces deux modèles, incluant notamment des données issues d'imagerie THR nous a permis de révéler l'hétérogénéité des habitats dans une région favorable dans son ensemble. Nous avons utilisé deux types de modèles complémentaires qui ont été validés à 96,97% et à 90,21% pour l'arbre de décision et le réseau de neurones respectivement. Nous avons obtenu, d'un côté, le détail des combinaisons environnementales précises favorables à la présence d'anciennes plantations de caféiers (selon les variables prédictives concernant le couvert forestier) et à l'hybridation inter-spécifiques (variables prédictives environnementales) (Figure 8, chapitre 5) et, d'un autre côté, la prédiction de distribution fine (probabilité) des zones favorables (Figure 9, chapitre 5).

Nous avons ainsi pu mettre en évidence l'hétérogénéité des habitats au travers de la prédiction de distribution des zones favorables à l'hybridation spontanée, localisant des niches très particulières. La reconnaissance de ces niches a des conséquences directes en termes d'opportunités de gestion ou de relance de la caféiculture et de développement rural dans cette région.

II - Applications

D'un point de vue appliqué, ces hybridations peuvent-elles avoir des conséquences sur les ressources génétiques des espèces concernées et sur la mise en place de méthodes de gestion adaptées au contexte socio-économique et environnemental du café en Nouvelle-Calédonie ?

Notre travail met ainsi en évidence des hybridations inter-spécifiques qui ont lieu dans des zones hybrides restreintes à des niches particulières. Nous nous retrouvons dans un cas de figure où les perspectives d'application peuvent prendre trois tournures distinctes mais complémentaires.

La première application concerne l'effort de conservation de ces zones hybrides au sein desquelles la diversité génétique évolue et se remanie. Les zones hybrides sont considérées comme des laboratoires où la diversité évolue (Barton et Hewitt, 1985). Ces laboratoires naturels permettent d'étudier les barrières aux flux de gènes, comme cela a été fait pour deux espèces du genre *Populus* (Lexer *et al.*, 2005). Plus largement, ils sont un moyen de comprendre tous les phénomènes liés à la création d'espèce, à l'isolement reproducteur entre espèces et les liens entre introgression et adaptation. Pour toutes ces raisons, il est important de conserver ces laboratoires naturels (Allendorf *et al.*, 2001). Afin d'orienter la conservation d'une zone hybride selon ses caractéristiques, Allendorf *et al.* (2001) ont proposé une classification de six types de zones hybrides scindés en deux groupes : les zones hybrides naturelles et les zones hybrides anthropogéniques.

Dans notre étude, nous sommes dans le cas de zones hybrides anthropogéniques où les introgressions existent mais où les hybrides n'ont pas pris le pas sur les espèces pures. Les recommandations de conservation dans ce type de zones hybrides proposées par Allendorf *et al.* (2001) portent surtout sur l'importance de conserver les espèces pures afin que ces zones hybrides ne deviennent pas des mélanges d'hybrides généralisés. En effet, conserver ces laboratoires naturels en Nouvelle-Calédonie, permet de conserver la diversité génétique des espèces pures, pour laisser évoluer cette « néo-diversité ». Cependant, cela reste contestable dans le cas des caféiers car le risque que ces espèces deviennent envahissantes existe.

La deuxième application consiste à identifier des hybrides ayant des caractéristiques agronomiques particulièrement intéressantes pour une culture plus aisée et plus naturelle. Un

aspect important des zones hybrides est qu'elles permettent potentiellement l'évaluation directe des géotypes hybrides adultes dans les conditions naturelles du milieu (Lexer *et al.*, 2003, 2004). Ces zones d'hybridation spontanée permettent la sélection des hybrides à la meilleure valeur adaptative et les mieux adaptés aux conditions environnementales locales. Ceci peut constituer un nouveau mode d'amélioration génétique original combinant hybridation et sélection naturelle. Au travers de l'hybridation, trois phénomènes favorables à la valeur adaptative des hybrides sont mis en évidence par Rieserberg *et al.* (2007) : l'hétérosis, l'augmentation de la variabilité génétique et la diminution du fardeau génétique.

Le redéploiement de la caféiculture en Nouvelle-Calédonie peut donc passer par l'évaluation agronomique des hybrides inter-spécifiques, afin de mettre au point une nouvelle variété de café, née à Sarraméa. Les hybrides inter-spécifiques peuvent représenter des opportunités de relance de la caféiculture au travers des combinaisons génétiques qui les caractérisent. L'intérêt agronomique de ces hybrides doit être évalué en champs expérimentaux. Des études sont en cours pour valider une nouvelle variété de caféier issue d'un hybride *C. arabica* ayant introgressé de *C. canephora* des facteurs de résistance aux maladies (Mahe *et al.*, 2007).

La troisième application consiste à utiliser les conditions environnementales rencontrées dans ces microenvironnements favorables caractérisés comme base de connaissances des conditions favorables à la culture du café en Nouvelle-Calédonie. Si l'environnement au sein duquel une plantation de caféiers évolue, est proche de celui des microenvironnements détectés, la culture sera alors plus naturelle et nécessitera moins d'eau et d'intrants comme les produits phytosanitaires. Certaines de ces anciennes plantations pourraient ainsi être ré-exploitées. Par exemple à Petit Couli (une des trois tribus de la commune de Sarraméa), une partie des terres a été libérée (anciens séquestres japonais) au début du siècle dernier. Ces terres avaient alors été revendiquées par les mélanésiens de la tribu concernée et du café y a été planté. La justification économique reste encore aujourd'hui l'unique moyen de revendication locale des terres par les mélanésiens. Les Groupements de Droit Particulier Local (GDPL) consistent à un regroupement de plusieurs familles qui s'engagent dans une activité économique commune. Il y a un potentiel de développement de la caféiculture très important dans la région de Sarraméa. Si ce redéploiement de la caféiculture avait lieu, cela pourrait être une opportunité de créer un label dans cette région, complètement justifié par l'incroyable conjoncture environnementale rencontrée dans cette région de la Nouvelle-Calédonie. Un effort de développement peut être mis en place, dans le respect des traditions et modèles sociaux mélanésiens qui, comme le précisait

Saussol (1979), est indispensable à la pérennité d'un projet de développement économique intégrant les populations locales. Grâce aux activités d'aide au développement des instances de la Nouvelle-Calédonie (notamment la Direction du Développement Rural de la Province Sud), des projets peuvent être mis en place pour l'exploitation des ressources génétiques exceptionnelles des caféiers dans cette région et la conservation d'un patrimoine naturel en perpétuelle diversification.

III- Perspectives de recherche

Notre travail aboutit à d'importantes conclusions qui constituent une contribution intéressante à la discipline émergente de la génétique des populations (Manel *et al.*, 2003). En effet, nous avons développé au cours de ce travail un modèle fin de prédiction de distribution des niches favorables aux hybridations inter-spécifiques sur l'ensemble d'une région. Nous avons démontré l'importance de la composante environnementale dans la dynamique du phénomène d'hybridation spontanée étudié.

Cependant, le modèle de prédiction pourrait être affiné en rajoutant de nouvelles variables environnementales. Par exemple, il serait intéressant d'insérer une variable bioclimatique dans le modèle. Pour cela, il faudrait placer de nouvelles stations météorologiques au sein de la zone d'étude, afin de détecter des variations climatiques fines.

Concernant les analyses génétiques fines, le travail s'est porté uniquement sur la population de Sarraméa afin de mettre en évidence l'ampleur de la dynamique d'hybridation à un endroit donné. Les autres sites témoins ont été étudiés ponctuellement mais nous n'avons pas caractérisé la diversité génétique présente, ni quantifier le taux d'hybridation. Afin d'obtenir une cartographie générale de la structure spatiale de la diversité des trois espèces de caféiers et des hybrides inter-spécifiques en fonction de l'hétérogénéité spatiale des habitats favorables, il faudrait élargir les analyses génétiques à toute la région d'étude. Pour cela, les trois autres sites témoins (Couli, Dogny et Farino), ainsi qu'une partie des sites utilisés pour la validation des modèles de prédiction, pourraient être utilisés. Ceci constituerait une analyse de génétique du paysage plus complète et apporterait des informations plus précises quant à l'impact de l'environnement sur la dynamique des hybridations.

Par ailleurs, nous n'avons présenté dans cette thèse que les résultats des analyses génétiques de 95 individus de la population de Sarraméa. Or, un échantillonnage complet a été effectué, avec le

positionnement et le génotypage des 400 individus adultes de la population. Les analyses génétiques de ces individus pourraient préciser la dynamique des flux de gènes dans la population tri-spécifique. La structuration spatiale de la diversité génétique pourrait être évaluée, et ses causes éventuelles pourraient être explicitées.

Plus globalement, ce type d'approches couplées (génétique, écologique et spatial) pourrait être appliqué à d'autres espèces introduites en Nouvelle-Calédonie ou ailleurs, afin de mettre en évidence une modification de leur niche écologique et de mieux appréhender leur comportement. Ce serait en l'occurrence très intéressant d'appliquer un tel modèle sur des espèces envahissantes, afin de se doter de moyens de lutte efficaces.

Références Bibliographiques

- Abbott RJ** (1992) Plant Invasions, Interspecific Hybridization and the Evolution of New Plant Taxa. *Trends in Ecology & Evolution* 7, 401-405.
- Abbott RJ, James JK, Milne RI, Gillies ACM** (2003) Plant introductions, hybridization and gene flow. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 358, 1123-1132.
- Adams JR, Lucash C, Schutte L, Waits LP** (2007) Locating hybrid individuals in the red wolf (*Canis rufus*) experimental population area using a spatially targeted sampling strategy and faecal DNA genotyping. *Mol Ecol* 16, 1823-1834.
- Aggarwal RK, Hendre PS, Varshney RK, et al.** (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. *Theoretical and Applied Genetics* 114, 359-372.
- Allendorf FW, Leary RF, Spruell P, Wenburg JK** (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution* 16, 613-622.
- Allendorf FW, Lundquist LL** (2003) Introduction: Population biology, evolution, and control of invasive species. *Conservation Biology* 17, 24-30.
- Alvarez-Moctezuma JG, Ochoa-Gaona S, de Jong BHJ, Soto-Pinto ML** (1999) Habitat and distribution of five *Quercus* (Fagaceae) species in the Chiapas Central Plateau, Mexico. *Revista De Biología Tropical* 47, 351-358.
- Anderson RP, Gomez-Laverde M, Peterson AT** (2002) Geographical distributions of spiny pocket mice in South America: insights from predictive models. *Global Ecology and Biogeography* 11, 131-141.
- Anthony F** (1992) Les ressources génétiques des caféiers. Collecte, gestion d'un conservatoire de la diversité génétique. In: *Travaux et Documents* (ed. ORSTOM), p. 320, Paris.
- Anthony F, Bertrand B, Quiros O, et al.** (2001) Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. *Euphytica* 118, 53-65.
- Anthony F, Combes MC, Astorga C, et al.** (2002) The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. *Theoretical and Applied Genetics* 104, 894-900.
- Anthony F, Couturon E, Namur de C** (1985) Les caféiers sauvages du Cameroun. Résultats d'une mission de prospection effectuée par l'ORSTOM en 1983, 495-501.
- Anthony NM, Johnson-Bawe M, Jeffery K, et al.** (2007) The role of Pleistocene refugia and rivers in shaping gorilla genetic diversity in central Africa. *Proceedings of the National Academy of Sciences of the United States of America* 104, 20432-20436.
- Antonovics J, Thrall P, Jarosz A, Stratton D** (1994) ecological genetics of metapopulations: the *Silene-Ustilago* plant-pathogen system. In: *Ecological genetics*, pp. 146-170. Princeton University Press, Princeton; NJ.
- Araujo MB, Pearson RG** (2005) Equilibrium of species' distributions with climate. *Ecography* 28, 693-695.
- Araujo MB, Pearson RG, Thuiller W, Erhard M** (2005) Validation of species-climate impact models under climate change. *Global Change Biology* 11, 1504-1513.
- Arcila-Pulgarin J, Buhr L, Bleiholder H, et al.** (2002) Application of the extended BBCH scale for the description of the growth stages of coffee (*Coffea* spp.). *Annals of Applied Biology* 141, 19-27.
- Arnold ML** (1997) Natural hybridization and evolution. *Oxford University Press, Oxford*.

- Arnold ML** (2006) Evolution through genetic exchange. *oxford University Press, Oxford, UK*.
- Asner GP, Palace M, Keller M, et al.** (2002) Estimating canopy structure in an Amazon Forest from laser range finder and IKONOS satellite observations. *Biotropica* 34, 483-492.
- Atlas N** (1981) *Atlas de la Nouvelle-Calédonie et dépendances* Orstom, Paris.
- Austin MP, Nicholls AO, Margules CR** (1990) Measurement of the Realized Qualitative Niche - Environmental Niches of 5 Eucalyptus Species. *Ecological Monographs* 60, 161-177.
- Avise JC** (2004) Molecular markers, natural history and evolution. *Second edition. Sinauer Associates, Inc, Sunderland MA*.
- Ayres DR, Garcia-Rossi D, Davis HG, Strong DR** (1999) Extent and degree of hybridization between exotic (*Spartina alterniflora*) and native (*S. foliosa*) cordgrass (Poaceae) in California, USA determined by random amplified polymorphic DNA (RAPDs). *Molecular Ecology* 8, 1179-1186.
- Ayres DR, Zaremba K, Strong DR** (2004) Extinction of a common native species by hybridization with an invasive congener. *Weed Technology* 18, 1288-1291.
- Bandeira FP, Martorell C, Meave JA, Caballero J** (2005) The role of rustic coffee plantations in the conservation of wild tree diversity in the Chinantec region of Mexico. *Biodiversity and Conservation* 14, 1225-1240.
- Baranski O** (1996) Etude des caractères morphologiques, phénologiques et de la fertilité de 29 taxons de caféiers sauvages africains. *Rapport VSN, ORSTOM*, pp 38.
- Barton NH** (2001) The role of hybridization in evolution. *Molecular Ecology* 10, 551-568.
- Barton NH, Hewitt GM** (1985) Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics* 16, 113-148.
- Baruah A, Naik P, Hendre S, et al.** (2003) Isolation and characterization of nine microsatellite markers from *Coffea arabica* L., showing wide cross-species amplifications. *Molecular Ecology Notes* 3, 647-650.
- Beer J, Muschler R, Kass D, Somarriba E** (1997) Shade management in coffee and cacao plantations. *Agroforestry Systems* 38, 139-164.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F** (1996-2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France).
- Benkman CW, Siepielski AM, Parchman TL** (2008) The local introduction of strongly interacting species and the loss of geographic variation in species and species interactions. *Mol Ecol* 17, 395-404.
- Berberoglu S, Curran PJ, Lloyd CD, Atkinson PM** (2007) Texture classification of Mediterranean land cover. *International Journal of Applied Earth Observation and Geoinformation* 9, 322-334.
- Berlacher SH** (1998) Can sympatric speciation via host or habitat shift be proven from phylogenetic and biogeographic evidence? In: *Endless forms: Species and speciation*. Oxford University Press; Oxford.
- Berthaud J** (1986) *Les ressources génétiques pour l'amélioration des caféiers africains diploïdes*, Collection Travaux et documents edn. Orstom, Montpellier, France.
- Berthaud J, Anthony F, Charrier A, et al.** (1989) Les caféiers en Afrique: diversité génétique et amélioration des plantes. *Bull. Soc. bot. Fr*, 239-250.
- Berthaud J, Charrier A** (1988) Genetic resources of *Coffea*. In: *Coffee. Vol 4 : Agronomy* (eds. Clarke RJ, Macrae R), pp. 1-42. Elsevier Applied Science, Londres.
- Berthaud J, Guillaumet JL** (1978) Les caféiers sauvages en Centrafrique. Résultats d'une mission de prospection (Janvier-février 1975). *Café Cacao Thé* 22, 171-186.

- Bertrand B** (2002) *L'amélioration génétique de Coffea arabica L. en Amérique Centrale par la voie hybride F1* Thèse de doctorat Montpellier ENSAM, Sciences agronomiques (Amélioration des plantes).
- Bettencourt AJ** (1973) Considerações gerais sobre o “Híbrido de Timor”. *Campinas, Brasil, Instituto Agronomico de Campinas.*, 20 p.
- Bhat PR, Krishnakumar V, Hendre PS, et al.** (2005) Identification and characterization of expressed sequence tags-derived simple sequence repeats, markers from robusta coffee variety 'CxR' (an interspecific hybrid of *Coffea canephora* x *Coffea congensis*). *Molecular Ecology Notes* 5, 80-83.
- Bishop C** (1995) *Neural networks for pattern recognition* Oxford Univ, Press, New-York.
- Blackburn TM, Cassey P, Duncan RP, Evans KL, Gaston KJ** (2004) Avian extinction and mammalian introductions on oceanic islands. *Science* 305, 1955-1958.
- Booth TH, Ekeleme F** (2002) Identifying climatically suitable areas for growing particular trees in Africa: An example using *Grevillea robusta*. *Agroforestry Systems* 54, 41-49.
- Booth TH, Jovanovic T, Old KM, Dudzinski MJ** (2000) Climatic mapping to identify high-risk areas for *Cylindrocladium quinqueseptatum* leaf blight on eucalypts in mainland South East Asia and around the world. *Environmental Pollution* 108, 365-372.
- Bossdorf O, Auge H, Lafuma L, et al.** (2005) Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia* 144, 1-11.
- Breiman L, Friedman J, Olshen R, Stone C** (1984) Classification and regression trees. *Boca Raton FL: CRC Press*.
- Bridson D, Verdcourt B** (1988) *Coffea*. In: *Flora of Tropical East Africa; Rubiaceae (Part 2)* (eds. Polhill RM, Baljema AA), pp. 703-723, Rotterdam.
- Bridson DM** (1994) Additional notes on *Coffea* (Rubiaceae) from tropical East Africa. *Kew bulletin* 49, 331-342.
- Bridson DM, Verdcourt B** (1988) Rubiaceae (Part 2). In: *Flora of Tropical East Africa*. R.M. Polhill, eds, Rotterdam, Balkema, 727 p.
- Broennimann O, Treier UA, Muller-Scharer H, et al.** (2007) Evidence of climatic niche shift during biological invasion. *Ecology Letters* 10, 701-709.
- Brouat C, McKey D, Douzery EJ** (2004) Differentiation in a geographical mosaic of plants coevolving with ants: phylogeny of the *Leonardoxa africana* complex (Fabaceae: Caesalpinioideae) using amplified fragment length polymorphism markers. *Mol Ecol* 13, 1157-1171.
- Buckley LB, Jetz W** (2007) Insularity and the determinants of lizard population density. *Ecol Lett* 10, 481-489.
- Buerkle CA, Morris RJ, Asmussen MA, Rieseberg LH** (2000) The likelihood of homoploid hybrid speciation. *Heredity* 84, 441-451.
- Campanha MM, Santos RHS, de Freitas GB, et al.** (2005) Growth and yield of coffee plants in agroforestry and monoculture systems in Minas Gerais, Brazil. *Agroforestry Systems* 63, 75-82.
- Carney SE, Arnold ML** (1997) Differences in pollen-tube growth rate and reproductive isolation between Louisiana irises. *Journal of Heredity* 88, 545-549.
- Carney SE, Cruzan MB, Arnold ML** (1994) Reproductive Interactions between Hybridizing Irises - Analyses of Pollen-Tube Growth and Fertilization Success. *American Journal of Botany* 81, 1169-1175.
- Carney SE, Hodges SA, Arnold ML** (1996) Effects of differential pollen-tube growth on hybridization in the Louisiana irises. *Evolution* 50, 1871-1878.
- Charles J** (1932) Le caféier. *Revue Agricole de la Chambre d'Agriculture de la Nouvelle-Calédonie* 32, 794 - 837.

- Charmetant P, Le Pierrès D** (1991) Rapport de mission en Nouvelle-Calédonie: prospection et collecte d'hybrides naturels de caféiers cultivés. *IRCC-CIRAD/ORSTOM* Paris, 21 p.
- Charrier A** (1978) *La structure génétiques des caféiers spontanés de la région malgache (Mascaro Coffea). Leurs relations avec les caféiers d'origine africaine (Eu Coffea)*. Thèse d'état, Université Paris-Sud.
- Charrier A, Berthaud J** (1985) Botanical classification of coffee. In: *Coffee: botany, biochemistry and production of beans and beverage*, pp. 13-47. London, UK, Croom Helm.
- Chevalier A** (1946) Ecologie et distribution géographique des caféiers sauvages et cultivés. *Revue de Botanique Appliquée d'Agriculture Tropicale*, 81-94.
- Combes MC, Andrzejewski S, Anthony F, et al.** (2000) Characterization of microsatellite loci in *Coffea arabica* and related coffee species. *Molecular Ecology* 9, 1178-1180.
- Cordero-Sancho S, Sader SA** (2007) Spectral analysis and classification accuracy of coffee crops using Landsat and a topographic-environmental model. *International Journal of Remote Sensing* 28, 1577-1593.
- Cornuet JM, Piry S, Luikart G, Estoup A, Solignac M** (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* 153, 1989-2000.
- Cottrell JE, Krystufek V, Tabbener HE, et al.** (2005) Postglacial migration of *Populus nigra* L.: lessons learnt from chloroplast DNA. *Forest Ecology and Management* 206, 71-90.
- Coulibaly I, Noirot M, Lorieux M, et al.** (2002) Introgression of self compatibility from *Coffea heterocalyx* to the cultivated species *Coffea canephora*. *Theoretical-and-Applied-Genetics*, 994-999.
- Coulibaly I, Revol B, Noirot M, et al.** (2003) AFLP and SSR polymorphism in a *Coffea* interspecific backcross progeny [(*C. canephora* x *C. heterocalyx*) x *C. canephora*]. *Theoretical and Applied Genetics* 107, 1148-1155.
- Couteron P, Pelissier R, Nicolini EA, Dominique P** (2005) Predicting tropical forest stand structure parameters from Fourier transform of very high-resolution remotely sensed canopy images. *Journal of Applied Ecology* 42, 1121-1128.
- Cox GW** (1999) *Alien species in North America and hawaii: impacts on natural ecosystems* Island Press, Washington D. C., USA.
- Coyne JA, Orr HA** (2004) *Speciation* Sinauer Associates, Sunderland, Massachussets.
- Cracraft J** (1989) Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In: *Speciation and its consequences*, pp. 28-59. Sinauer Associates, Sunderland, MA.
- Cramer PJS** (1957) A review of literature of coffee research in Indonesia. *Inter American Institute of Agricultural Sciences* W. F. L. Ed., 262.
- Cros J, Combes MC, Chabrillange N, et al.** (1995) Nuclear DNA content in the subgenus *Coffea* (Rubiaceae): inter- and intra-specific variation in African species. *Canadian Journal of Botany* 73, 14-20.
- Cros J, Combes MC, Trouslot P, et al.** (1998) Phylogenetic analysis of chloroplast DNA variation in *Coffea* L. *Molecular Phylogenetics and Evolution* 9, 109-117.
- Cubry P, Musoli P, Legnate H, et al.** (2008) Diversity in coffee using SSR markers: structure of the *Coffea* genus and perspectives for breeding. *Genome* 51, 50-63.
- Culvenor DS** (2002) TIDA: an algorithm for the delineation of tree crowns in high spatial resolution remotely sensed imagery. *Computers & Geosciences* 28, 33-44.
- Darling JA, Bagley MJ, Roman J, Tepolt CK, Geller JB** (2008) Genetic patterns across multiple introductions of the globally invasive crab genus *Carcinus*. *Molecular Ecology* 17, 4992-5007.

- David G, Pillon P** (1999) La relance caféicole en secteur rural mélanésien. In David Gilbert (dir.), Guillaud Dominique (dir.), Pillon Patrick (dir.). La Nouvelle-Calédonie à la croisée des chemins : 1989-1997 *Source Paris (FRA), Nouméa : IRD, Société des Océanistes*, p. 195-217.
- David G, Siapo P** (1994) Terre, culture et développement en Nouvelle-Calédonie : la caféiculture en milieu mélanésien In *Capecchi B. (ed.). La terre : actes du sixième colloque C.O.R.A.I.L.*, p. 197-210.
- Davis AP, Govaerts R, Bridson DM, Stoffelen P** (2006) An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). *Botanical Journal of the Linnean Society* 152, 465-512.
- de Queiroz K** (2005) Different species problems and their resolution. *Bioessays* 27, 1263-1269.
- de Queiroz K, Donoghue MJ** (1988) Phylogenetic systematics and the species problem. *Cladistics* 4, 317-338.
- Deuss J** (1987) Les projets de développement de la caféiculture au Togo, au Vietnam et en Nouvelle-Calédonie. In: *In. Antheoume et al. (ed.) - Le développement rural: comprendre pour agir* (ed. ORSTOM P-), pp. 47 - 70.
- Dieckmann U, Doebeli M** (1999) On the origin of species by sympatric speciation. *Nature* 400, 354-357.
- Dietz H, Edwards PJ** (2006) Recognition that causal processes change during plant invasion helps explain conflicts in evidence. *Ecology* 87, 1359-1367.
- Dlugosch KM, Parker IM** (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology* 17, 431-449.
- Dobzhansky T** (1937) *Genetics and the Origin of species* Colombia University Press, New York.
- Dobzhansky T** (1940) Speciation as a stage in evolutionary divergence. *American Naturalist* 74, 312-321.
- Dobzhansky T** (1951) *Genetics and the origins of species* Columbia University Press, New York.
- Durka W, Bossdorf O, Prati D, Auge H** (2005) Molecular evidence for multiple introductions of garlic mustard (*Alliaria petiolata*, Brassicaceae) to North America. *Molecular Ecology* 14, 1697-1706.
- Dussert S, Lashermes P, Anthony F, et al.** (1999) Coffee, *Coffea canephora*. In: *Diversité génétique des plantes tropicales cultivées* (eds. Hamon P, Seguin M, Perrier X, Glaszmann J-C), pp. 175-794. CIRAD, Montpellier, France.
- Elliott LJ, Mason DC, Wilkinson MJ, et al.** (2004) The role of satellite image-processing for national-scale estimates of gene flow from genetically modified crops: rapeseed in the UK as a model. *Journal of Applied Ecology* 41, 1174-1184.
- Ellstrand NC, Elam DR** (1993) Population Genetic Consequences of Small Population-Size - Implications for Plant Conservation. *Annual Review of Ecology and Systematics* 24, 217-242.
- Ellstrand NC, Schierenbeck KA** (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences of the United States of America* 97, 7043-7050.
- Erikson M** (2004) Species classification of individually segmented tree crowns in high-resolution aerial images using radiometric and morphologic image measures. *Remote Sensing of Environment* 91, 469-477.
- Estoup A, Angers B** (1998) Microsatellites and minisatellites for molecular ecology: theoretical and experimental considerations. In: *Advances in Molecular Ecology* (ed. Carvalho G). NATO Press, Amsterdam.

- Estoup A, Rousset F, Michalakis Y, et al.** (1998) Comparative analysis of microsatellite and allozyme markers: a case study investigating microgeographic differentiation in brown trout (*Salmo trutta*). *Molecular Ecology* 7, 339-353.
- Estoup A, Tailliez C, Cornuet JM, Solignac M** (1995) Size Homoplasmy and Mutational Processes of Interrupted Microsatellites in 2 Bee Species, *Apis-Mellifera* and *Bombus-Terrestris* (Apidae). *Molecular Biology and Evolution* 12, 1074-1084.
- Evans BJ, Kelley DB, Tinsley RC, Melnick DJ, Cannatella DC** (2004) A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. *Mol Phylogenet Evol* 33, 197-213.
- Falush D, Stephens M, Pritchard JK** (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164, 1567-1587.
- Felsenstein J** (2005) PHYLIP (Phylogeny Inference Package) version 3.6. .
- Ferguson ME, Jarvis A, Stalker HT, et al.** (2005) Biogeography of wild *Arachis* (Leguminosae): distribution and environmental characterisation. *Biodiversity and Conservation* 14, 1777-1798.
- Ferrier S, Guisan A** (2006) Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology* 43, 393-404.
- Fitzpatrick MC, Weltzin JF, Sanders NJ, Dunn RR** (2007) The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? *Global Ecology and Biogeography* 16, 24-33.
- Fjeldsa J, Lovett JC** (1997) Geographical patterns of old and young species in African forest biota: The significance of specific montane areas as evolutionary centres. *Biodiversity and Conservation* 6, 325-346.
- Fontaine C, Lovett PN, Sanou H, Maley J, Bouvet JM** (2004) Genetic diversity of the shea tree (*Vitellaria paradoxa* C.F. Gaertn), detected by RAPD and chloroplast microsatellite markers. *Heredity* 93, 639-648.
- Foody GM** (2007) Editorial: Ecological applications of remote sensing and GIS. *Ecological Informatics* 2, 71-72.
- Fortune PM, Schierenbeck K, Ayres D, et al.** (2008) The enigmatic invasive *Spartina densiflora*: A history of hybridizations in a polyploidy context. *Mol Ecol* 17, 4304-4316.
- Garnier S, Alibert P, Audiot P, Prieur B, Rasplus JY** (2004) Isolation by distance and sharp discontinuities in gene frequencies: implications for the phylogeography of an alpine insect species, *Carabus solieri*. *Mol Ecol* 13, 1883-1897.
- Gause GF** (1934) *The struggle for existence*.
- Gavashelishvili A, Lukarevskiy V** (2008) Modelling the habitat requirements of leopard *Panthera pardus* in west and central Asia. *Journal of Applied Ecology* 45, 579-588.
- Genton BJ, Shykoff JA, Giraud T** (2005) High genetic diversity in French invasive populations of common ragweed, *Ambrosia artemisiifolia*, as a result of multiple sources of introduction. *Molecular Ecology* 14, 4275-4285.
- Gerard PR, Klein EK, Austerlitz F, Fernandez-Manjarres JF, Frascaria-Lacoste N** (2006) Assortative mating and differential male mating success in an ash hybrid zone population. *BMC Evol Biol* 6, 96.
- Geromel C, Ferreira LP, Guerreiro SMC, et al.** (2006) Biochemical and genomic analysis of sucrose metabolism during coffee (*Coffea arabica*) fruit development. *Journal of Experimental Botany* 57, 3243-3258.
- Giordanengo P** (1992) *Biologie, éco-éthologie et dynamique des populations du scolyte des grains de café, Hypothenemus hampei Ferr. (Coleoptera, Scolytidae), en Nouvelle-Calédonie* Thèse Doctorale, Université de Rennes 1.

- Godoy O, Richardson DM, Valladares F, Castro-Diez P** (2009) Flowering phenology of invasive alien plant species compared with native species in three Mediterranean-type ecosystems. *Ann Bot (Lond)* 103, 485-494.
- Gomez C, Batti A, Le Pierres D, et al.** (Submitted) Favorable habitats for *Coffea* interspecific hybridization in central New Caledonia: combined genetic and spatial analyses. *Journal of Applied Ecology*.
- Gomez C, Dussert S, Hamon P, et al.** (2009) Current genetic differentiation of *Coffea canephora* Pierre ex A.Froehn in the Guineo-Congolian African zone: cumulative impact of ancient climatic changes and recent human activities. *BMC Evol Biol* 9, 167.
- Gow JL, Peichel CL, Taylor EB** (2006) Contrasting hybridization rates between sympatric three-spined sticklebacks highlight the fragility of reproductive barriers between evolutionarily young species. *Mol Ecol* 15, 739-752.
- Graham CH, Ron SR, Santos JC, Schneider CJ, Moritz C** (2004) Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution* 58, 1781-1793.
- Grant BR, Grant PR** (1998) Hybridization and speciation in darwin's finches: the role of sexual imprinting on a culturally transmitted trait. In: *Species and speciation*, pp. 404-424. Oxford University Press, New-York.
- Grant V** (1981) Plant speciation. *Columbia University Press, New-york*.
- Guarino L, Jarvis A, Hijmans RJ, Maxted N** (2001) Geographic Information Systems (GIS) and the conservation and use of plant genetic resources.
- Guisan A, Theurillat JP** (2000) Equilibrium modeling of alpine plant distribution: how far can we go? *Phytocoenologia* 30, 353-384.
- Guisan A, Thuiller W** (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8, 993-1009.
- Guisan A, Zimmermann NE** (2000) Predictive habitat distribution models in ecology. *Ecological Modelling* 135, 147-186.
- Guiziou C.** (2009) *Rapport de stage de M1*.
- Haffer J** (1969) Speciation in Amazonian forest birds. *Science* 165, 131-137.
- Hamon P, Siljak-Yakovlev S, Srisuwan S, et al.** (2009) Physical mapping of rDNA and heterochromatin in chromosomes of 16 *Coffea* species: A revised view of species differentiation. *Chromosome Research* 17, 291-304.
- Hamon S, Dussert S, Deu M, et al.** (1998) Effects of quantitative and qualitative Principal Component Score Strategies (PCSS) on the structure of coffee, rice, rubber tree and sorghum core collections. *Genetics Selection Evolution* 30, 237- 258.
- Hanski I** (1998) Metapopulation dynamics. *Nature* 396, 41-49.
- Haralick RM, Shanmugam K, Dinstein I** (1973) Textural features for image classification. *IEEE Trans. Syst. man Cybernet.*, 610-621.
- Hastings A, Harrison S** (1994) Metapopulation Dynamics and Genetics. *Annual Review of Ecology and Systematics* 25, 167-188.
- Hermson J** (1963) The genetic basis of hybrid necrosis in wheat. *Genetica* 33, 245-287.
- Hernández-Martínez G, Manson RH, Contreras Hernández A** (2009) Quantitative classification of coffee agroecosystems spanning a range of production intensities in central Veracruz, Mexico. *Agriculture, Ecosystems and Environment* in press.
- Herrera JC, Combes MC, Anthony F, Charrier A, 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* × *C-canephora*). *Theoretical and Applied Genetics* 104, 661-668.

- Herrera JC, Combes MC, Cortina H, Alvarado G, Lashermes P** (2002b) Gene introgression into *Coffea arabica* by way of triploid hybrids (C-arabica x C- *canephora*). *Heredity* 89, 488-494.
- Herrera JC, Combes MC, Cortina H, Lashermes P** (2004) Factors influencing gene introgression into the allotetraploid *Coffea arabica* L. from its diploid relatives. *Genome* 47, 1053-1060.
- Hewitt GM** (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philos Trans R Soc Lond B Biol Sci* 359, 183-195.
- Hey J** (2001) The mind of the species problem. *Trends Ecol Evol* 16, 326-329.
- Hierro JL, Maron JL, Callaway RM** (2005) A biogeographical approach to plant invasions: the importance of studying exotics in their introduced and native range. *Journal of Ecology* 93, 5-15.
- Hijmans RJ, Graham CH** (2006) The ability of climate envelope models to predict the effect of climate change on species distributions. *Global Change Biology* 12, 2272-2281.
- Holt RD** (1985) Population-Dynamics in 2-Patch Environments - Some Anomalous Consequences of an Optimal Habitat Distribution. *Theoretical Population biology* 28, 181-208.
- Holt RD, Barfield M, Gomulkiewicz R** (2004) Temporal variation can facilitate niche evolution in harsh sink environments. *American Naturalist* 164, 187-200.
- Huntley B, Webb T** (1989) Migration - Species Response to Climatic Variations Caused by Changes in the Earth's Orbit. *Journal of Biogeography* 16, 5-19.
- Husband BC, Barrett SCH** (1996) A metapopulation perspective in plant population biology. *Journal of Ecology* 84, 461-469.
- Hutchinson GE** (1957) Concluding remarks. *Cold Spring Harbour Symp. Quant. Biol.* 22, 415-427.
- Jaffré T, Veillon J-M** (1995) Structural and floristic characteristics of a rain forest on schist in New Caledonia : a comparison with an ultramafic rain forest. *Bulletin du Muséum national d'Histoire naturelle. Section B, Adansonia* 3-4, 201-226.
- Jagoret P, Cilas C, Eskes AB, Charmetant P** (1999) Caractérisation et évaluation des hybrides naturels de caféiers de Nouvelle-Calédonie. *Plantations, recherche, développement* Septembre-Octobre
- Jakob SS, Ihlow A, Blattner FR** (2007) Combined ecological niche modelling and molecular phylogeography revealed the evolutionary history of *Hordeum marinum* (Poaceae) - niche differentiation, loss of genetic diversity, and speciation in Mediterranean Quaternary refugia. *Molecular Ecology* 16, 1713-1727.
- Jarne P, Lagoda PJJ** (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution* 11, 424-429.
- Jarvis A, Williams K, Williams D, et al.** (2005) use of GIS for optimizing a collecting mission for a rare wild pepper (*Capsicum flexuosum* Sendtn.) in Paraguay. *Genetic Resources and Crop Evolution* 52, 671 - 682.
- Ji W, Leberg P** (2002) A GIS-based approach for assessing the regional conservation status of genetic diversity: An example from the southern Appalachians. *Environmental Management* 29, 531-544.
- Johansen K, Coops NC, Gergel SE, Stange Y** (2007) Application of high spatial resolution satellite imagery for riparian and forest ecosystem classification. *Remote Sensing of Environment* 110, 29-44.
- Jones PG, Beebe SE, Tohme J, Galwey NW** (1997) The use of geographical information systems in biodiversity exploration and conservation. *Biodiversity and Conservation* 6, 947-958.

- Joshi C, De Leeuw J, Skidmore AK, van Duren IC, van Oosten H** (2006) Remotely sensed estimation of forest canopy density: A comparison of the performance of four methods. *International Journal of Applied Earth Observation and Geoinformation* 8, 84-95.
- Kaczensky P, Ganbaatar O, von Wehrden H, Walzer C** (2008) Resource selection by sympatric wild equids in the Mongolian Gobi. *Journal of Applied Ecology* 45, 1762-1769.
- Kawecki TJ** (1997) Habitat quality ranking depends on habitat-independent environmental factors: A model and results from *Callosobruchus maculatus*. *Functional Ecology* 11, 247-254.
- Kawecki TJ, Ebert D** (2004) Conceptual issues in local adaptation. *Ecology Letters* 7, 1225-1241.
- Kay KM** (2006) Reproductive isolation between two closely related hummingbird-pollinated neotropical gingers. *Evolution* 60, 538-552.
- Kayitakire F, Hamel C, Defourny P** (2006) Retrieving forest structure variables based on image texture analysis and IKONOS-2 imagery. *Remote Sensing of Environment* 102, 390-401.
- Kearney M, Porter W** (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters* 12, 334-350.
- Kenji K, Larry AR** (1992) A Practical Approach to Feature Selection. In: *Proceedings of the Ninth International Workshop on Machine Learning*. Morgan Kaufmann Publishers Inc.
- Kerr JT, Ostrovsky M** (2003) From space to species: ecological applications for remote sensing. *Trends in Ecology & Evolution* 18, 299-305.
- Kidd DM, Ritchie MG** (2000) Inferring the patterns and causes of geographic variation in *Ephippiger ephippiger* (Orthoptera, Tettigoniidae) using geographical information systems (GIS). *Biological Journal of the Linnean Society* 71, 269-295.
- Kier G, Kreft H, Lee TM, et al.** (2009) A global assessment of endemism and species richness across island and mainland regions. *Proc Natl Acad Sci U S A* 106, 9322-9327.
- Kolher JM, Pillon P** (1983) Impact de l'opération café en milieu mélanésien. Volume II. L'environnement socio-économique: une approche du système de production. Saction de sociologie - ORSTOM, Nouméa.
- Kozak J, Estreguil C, Ostapowicz K** (2008a) European forest cover mapping with high resolution satellite data: The Carpathians case study. *International Journal of Applied Earth Observation and Geoinformation* 10, 44-55.
- Kozak KH, Graham CH, Wiens JJ** (2008b) Integrating GIS-based environmental data into evolutionary biology. *Trends Ecol Evol* 23, 141-148.
- Kuittinen H, Sillanpää MJ, Savolainen O** (1997) Genetic basis of adaptation: flowering time in *Arabidopsis thaliana*. *Theoretical and Applied Genetics* 95, 573-583.
- Ky CL, Barre P, Lorieux M, et al.** (2000) Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). *Theoretical and Applied Genetics* 101, 669-676.
- Lambrinos JG** (2004) How interactions between ecology and evolution influence contemporary invasion dynamics. *Ecology* 85, 2061-2070.
- Lamont BB, He T, Enright NJ, Krauss SL, Miller BP** (2003) Anthropogenic disturbance promotes hybridization between *Banksia* species by altering their biology. *Journal of Evolutionary Biology* 16, 551-557.
- Lashermes P, Combes MC, Robert J, et al.** (1999) Molecular characterisation and origin of the *Coffea arabica* L. genome. *Molecular and General Genetics* 261, 259-266.
- Lashermes P, Combes MC, Trouslot P, 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, Trouslot P, Anthony F, Combes MC, Charrier A** (1996) Genetic diversity for RAPD markers between cultivated and wild accessions of *Coffea arabica*. *Euphytica* 87, 59-64.
- Lefebvre-Pautigny F, Wu F, Philippot M, Rigoreau M, Priyono, Zouine M, Frasse P, Bouzayen M, Broun P, Pétiard V, Tanksley S D and Crouzillat D** (2010) High resolution synteny maps allowing direct comparisons between the coffee and tomato genomes *Tree Genetics & Genomes*
- Le Pierrès D** (1995) *Etude des hybrides inter-spécifiques tétraploïdes de première génération entre Coffea arabica L. et les caféiers diploïdes*, Université de Paris Sud, Centre Orsay.
- Le Pierrès D** (1999) Etude des hybrides naturels entre *Coffea arabica* et *Coffea canephora* de Nouvelle-Calédonie. In: *Field report* (ed. IRD), Nouméa.
- Le Pierrès D, Charmetant P, Yapo A, et al.** (1989) Les caféiers sauvages de Côte-d'Ivoire et de Guinée : bilan des missions de prospection effectuées de 1984 à 1987, 420-428.
- Lek S, Guegan JF** (1999) Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling* 120, 65-73.
- Leroy JF** (1982) L'origine kenyane du genre *Coffea* L. et la radiation des espèces de Madagascar, 413-420.
- Leroy T, Marraccini P, Dufour M, et al.** (2005) Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes. *Theor Appl Genet*, 1-10.
- Leroy T, Montagnon C, Charrier A, Eskes A** (1993) Reciprocal recurrent selection applied to *Coffea canephora* Pierre. 1. Characterization and evaluation of breeding populations and value of intergroupe hybrids. *Euphytica* 67, 179-190.
- Leroy T, Montagnon C, Cilas C, et al.** (1997) Reciprocal recurrent selection applied to *Coffea canephora* Pierre .3. Genetic gains and results of first cycle intergroup crosses. *Euphytica* 95, 347-354.
- Levin DA** (2006) Flowering phenology in relation to adaptive radiation. *Systematic Botany* 31, 239-246.
- Lewontin RC, Birch LC** (1966) Hybridization as a source of variation for adaptation to new environments. *Evolution* 20, 315-336.
- Lexer C, Fay MF, Joseph JA, Nica MS, Heinze B** (2005) Barrier to gene flow between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. *Mol Ecol* 14, 1045-1057.
- Lexer C, Heinze B, Alia R, Rieseberg LH** (2004) Hybrid zones as a tool for identifying adaptive genetic variation in outbreeding forest trees: lessons from wild annual sunflowers (*Helianthus* spp.). *Forest Ecology and Management* 197, 49-64.
- Lexer C, Randell RA, Rieseberg LH** (2003) Experimental hybridization as a tool for studying selection in the wild. *Ecology* 84, 1688-1699.
- Li ZK, Pinson SRM, Paterson AH, Park WD, Stansel JW** (1997) Genetics of hybrid sterility and hybrid breakdown in an intersubspecific rice (*Oryza sativa* L.) population. *Genetics* 145, 1139-1148.
- Lin BB** (2007) Agroforestry management as an adaptive strategy against potential microclimate extremes in coffee agriculture. *Agricultural and Forest Meteorology* 144, 85-94.
- Linder HP** (2001) Plant diversity and endemism in sub-Saharan tropical Africa. *Journal of Biogeography* 28, 169-182.
- Liu K, Muse SV** (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21, 2128-2129.

- Lopez-Gomez AA, Williams-Linera G, Manson RH** (2008) Tree species diversity and vegetation structure in shade coffee farms in Veracruz, Mexico. *Agriculture Ecosystems & Environment* 124, 160-172.
- Lorenzen ED, De Neergaard R, Arctander P, Siegismund HR** (2007) Phylogeography, hybridization and Pleistocene refugia of the kob antelope (*Kobus kob*). *Molecular Ecology* 16, 3241-3252.
- Losos JB, Glor RE** (2003) Phylogenetic comparative methods and the geography of speciation. *Trends in Ecology & Evolution* 18, 220-227.
- Louarn J** (1987) Possibilités d'amélioration de l'espèce cultivée *Coffea canephora* Pierre par hybridation interspécifique avec les caféiers africains diploïdes : données cytogénétiques sur les hybrides F1.
- Louarn J** (1992) *La fertilité des hybrides interspécifiques et les relations génomiques entre caféiers diploïdes d'origine africaine (Genre Coffea L. sous-genre Coffea)*. Thèse d'Etat, Montpellier II, Sciences et Techniques du Languedoc.
- Lowe AJ, Gillies AC, Wilson J, Dawson IK** (2000) Conservation genetics of bush mango from central/west Africa: implications from random amplified polymorphic DNA analysis. *Mol Ecol* 9, 831-841.
- Lowry DB, Modliszewski JL, Wright KM, Wu CA, Willis JH** (2009) The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B-Biological Sciences* 363, 3009-3021.
- MacArthur RH, Wilson EO** (1963) An equilibrium theory of insular zoogeography. *Evolution* 17, 373-387.
- Mack RN, Simberloff D, Lonsdale WM, et al.** (2000) Biotic invasions: Causes, epidemiology, global consequences, and control. *Ecological Applications* 10, 689-710.
- Mahe L, Le Pierres D, Combes MC, Lashermes P** (2007) Introgressive hybridization between the allotetraploid *Coffea arabica* and one of its diploid ancestors, *Coffea canephora*, in an exceptional sympatric zone in New Caledonia. *Genome* 50, 316-324.
- Maley J** (1996) The African rain forest: main characteristics of changes in vegetation and climate from the upper Cretaceous to the Quaternary. *Proceedings of the Royal Society of Edinburgh* 104B, 31-73.
- Maley J** (2001) La destruction catastrophique des forêts d'Afrique centrale survenue il y a environ 2500 ans exerce encore une influence majeure sur la répartition actuelle des formations végétales. In: *Plant systematics and phytogeography for the understanding of african biodiversity. Proceedings of the XVIth AETFAT Congress, held in 2000 at the National Botanic Garden of Belgium* (eds. Robbrecht E, Degreef J, Friis I), pp. 777-796. Syst. Geogr. Pl.
- Maley J, Chepstow-Lusty A** (2001) *Elaeis guineensis* Jacq. (oil palm) fluctuation in central Africa during the late Holocene: climate or human driving forces for this pioneering species ? *Vegetation History and Archeobotany* 10, 117-120.
- Mallet J** (1995) A Species Definition for the Modern Synthesis. *Trends in Ecology & Evolution* 10, 294-299.
- Mallet J** (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* 20, 229-237.
- Mallet J, Beltran M, Neukirchen W, Linares M** (2007) Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evol Biol* 7, 28.
- Maluf MP, Silvestrini M, Ruggiero LMD, Guerreiro O, Colombo CA** (2005) Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP and SSR marker systems. *Scientia Agricola* 62, 366-373.

- Manel S, Schwartz MK, Luikart G, Taberlet P** (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* 18, 189-197.
- Maron JL, Vila M, Bommarco R, Elmendorf S, Beardsley P** (2004) Rapid evolution of an invasive plant. *Ecological Monographs* 74, 261-280.
- Mas AH, Dietsch TV** (2004) Linking shade coffee certification to biodiversity conservation: Butterflies and birds in Chiapas, Mexico. *Ecological Applications* 14, 642-654.
- Maurin O, Davis AP, Chester M, et al.** (2007) Towards a phylogeny for *Coffea* (Rubiaceae): Identifying well-sipported lineages based on nuclear and plastid DNA sequences *Annals of Botany*, 1-19.
- Mayaux P, Richards T, Janodet E** (1999) A vegetation map of Central Africa derived from satellite imagery. *Journal of Biogeography* 26, 353-366.
- Mayr E** (1942) *Systematics and the Origin of Species* Columbia University Press, New-York.
- Mayr E** (1963) *Animal species and evolution* Cambridge, MA: Harvard University Press.
- Mayr E** (1982) *Toward a new philosophy of biology* Havard University Press, Cambridge, MA.
- Mayr E, Ohara RJ** (1986) The Biogeographic Evidence Supporting the Pleistocene Forest Refuge Hypothesis. *Evolution* 40, 55-67.
- Mendez VE, Gliessman SR, Gilbert GS** (2007) Tree biodiversity in farmer cooperatives of a shade coffee landscape in western El Salvador. *Agriculture Ecosystems & Environment* 119, 145-159.
- Milne RI, Abbott RJ, Wolff K, Chamberlain DF** (1999) Hybridization among sympatric species of *Rhododendron* (Ericaceae) in Turkey: morphological and molecular evidence. *Am J Bot* 86, 1776-1785.
- Minsky M, Papert S** (1969) *Perceptrons*, Cambridge (E.U.).
- Mitas L, Mitsova H** (1988) General Variational Approach to the Interpolation Problem. *Computers & Mathematics with Applications* 16, 983-992.
- Mittermeier RA, Robles Gil P, Hoffman M, et al.** (2004) *Hotspots revisited: Earth's biologically richest and most threatened terrestrial ecoregions.*, Mexico City, Mexico.
- Moguel P, Toledo VM** (1999) Biodiversity conservation in traditional coffee systems of Mexico. *Conservation Biology* 13, 11-21.
- Moncada P, McCouch S** (2004) Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. *Genome* 47, 501-509.
- Montagnon C** (2000) *Optimisation des gains génétiques dans le chéma de sélection récurrente réciproque de Coffea canephora P.*, Thèse Ecole Nationale Supérieure Agronomique de Montpellier.
- Montagnon C, Leroy T, Eskes AB** (1998) Amélioration variétale de *Coffea canephora*. I. Critères et méthodes de sélection. II. Les programmes de sélection et leurs résultats. *Plantation, recherche, développement* Mars-avril, 89-94.
- Montagnon C, Leroy T, Yapo A** (1992a) Etude complémentaire de la diversité génotypique et phénotypique des caféiers de l'espèce *C. canephora* en collection en Côte d'Ivoire. In : XVIe Colloque Scientifique International sur le Café. Paris, ASIC. 444-450.
- Montagnon C, Leroy T, Yapo A** (1992b) Genotypic and phenotypic diversity of some coffee groups (*Coffea canephora* Pierre) in the collections - consequences on their use in breeding. *Cafe Cacao The* 36, 187-198.
- Moore WS** (1977) Evaluation of Narrow Hybrid Zones in Vertebrates. *Quarterly Review of Biology* 52, 263-277.
- Mueller-Starck G, Schubert R** (2001) Preface. In: *Genetic response of forest systems to changing environmental conditions*. Kluwer Academic Publishers, Dordrecht.

- Mueller T, Olson KA, Fuller TK, et al.** (2008) In search of forage: predicting dynamic habitats of Mongolian gazelles using satellite-based estimates of vegetation productivity. *Journal of Applied Ecology* 45, 649-658.
- Muller J, Brandl R** (2009) Assessing biodiversity by remote sensing in mountainous terrain: the potential of LiDAR to predict forest beetle assemblages. *Journal of Applied Ecology* 46, 897-905.
- Muloko-Ntoutoume N, Petit RJ, White L, Abernethy K** (2000) Chloroplast DNA variation in a rainforest tree (*Aucoumea klaineana*, burseraceae) in Gabon. *Mol Ecol* 9, 359-363.
- Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GA, Kent J** (2000) Biodiversity hotspots for conservation priorities. *Nature* 403, 853-858.
- N' Diaye A, Noirot M, Hamon S, Poncet V** (2007) Genetic basis of species differentiation between *Coffea liberica* Hiern and *C. canephora* Pierre: Analysis of an interspecific cross. *Genetic Resources and Crop Evolution* 54, 1011-1021.
- N' Diaye A, Poncet V, Louarn J, Hamon S, Noirot M** (2005) Genetic differentiation between *Coffea liberica* var. *liberica* and *C. liberica* var. *dewevrei* and comparison with *C. canephora*. *Plant Systematics and Evolution* 253, 95-104.
- Namur de C, Couturon E, Sita P, Anthony F** (1988) Résultats d'une mission de prospection des caféiers sauvages du Congo, 397-404.
- Nei M, Maruyama T, Chakraborty R** (1975) Bottleneck Effect and Genetic-Variability in Populations. *Evolution* 29, 1-10.
- Nicolas V, Querouil S, Verheyen E, et al.** (2006) Mitochondrial phylogeny of African wood mice, genus *Hylomyscus* (Rodentia, Muridae): implications for their taxonomy and biogeography. *Mol Phylogenet Evol* 38, 779-793.
- Noirot M, Poncet V, Barre P, et al.** (2003) Genome size variations in diploid African *Coffea* species. *Annals of Botany* 92, 709-714.
- Novak SJ, Mack RN** (1993) Genetic-Variation in *Bromus-Tectorum* (Poaceae) - Comparison between Native and Introduced Populations. *Heredity* 71, 167-176.
- Novak SJ, Mack RN, Soltis PS** (1993) Genetic-Variation in *Bromus-Tectorum* (Poaceae) - Introduction Dynamics in North-America. *Canadian Journal of Botany-Revue Canadienne De Botanique* 71, 1441-1448.
- Orka HO, Naesset E, Bollandsas OM** (2009) Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. *Remote Sensing of Environment* 113, 1163-1174.
- Orr HA, Turelli M** (2001) The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution* 55, 1085-1094.
- Orr MR, Smith TB** (1998) Ecology and speciation. *Trends in Ecology & Evolution* 13, 502-506.
- Pacifici F, Chini M, Emery WJ** (2009) A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. *Remote Sensing of Environment* 113, 1276-1292.
- Paetkau D, Calvert W, Stirling I, Strobeck C** (1995) Microsatellite analysis of population structure in Canadian polar bears. *Mol Ecol* 4, 347-354.
- Paillard M, Lashermes P, Petiard V** (1996) Construction of a molecular linkage map in coffee. *Theoretical and Applied Genetics* 93, 41-47.
- Palace M, Keller M, Asner GP, Hagen S, Braswell B** (2008) Amazon forest structure from IKONOS satellite data and the automated characterization of forest canopy properties. *Biotropica* 40, 141-150.
- Parmesan C, Yohe G** (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature* 421, 37-42.

- Paterson HEH** (1985) the recognition concept of species. In: *Species and speciation*, pp. 21-29. Transvaal Museum Monograph No.4, Pretoria.
- Pavlacky DC, Goldizen AW, Prentis PJ, Nicholls JA, Lowe AJ** (2009) A landscape genetics approach for quantifying the relative influence of historic and contemporary habitat heterogeneity on the genetic connectivity of a rainforest bird. *Molecular Ecology* 18, 2945-2960.
- Pearman PB, Guisan A, Broennimann O, Randin CF** (2008a) Niche dynamics in space and time. *Trends in Ecology & Evolution* 23, 149-158.
- Pearman PB, Randin CF, Broennimann O, et al.** (2008b) Prediction of plant species distributions across six millennia. *Ecology Letters* 11, 357-369.
- Pearson RG, Dawson TP** (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography* 12, 361-371.
- Pearson RG, Dawson TP, Berry PM, Harrison PA** (2002) SPECIES: A Spatial Evaluation of Climate Impact on the Envelope of Species. *Ecological Modelling* 154, 289-300.
- Pearson RG, Dawson TP, Liu C** (2004) Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography* 27, 285-298.
- Pease KM, Freedman AH, Pollinger JP, et al.** (2009) Landscape genetics of California mule deer (*Odocoileus hemionus*): the roles of ecological and historical factors in generating differentiation. *Molecular Ecology* 18, 1848-1862.
- Pellegrin F, Seivert B, Kohler F, Van Bercie C, Boccas B** (1983a) La rouille orangée du caféier arabica en Nouvelle-Calédonie. *Café Cacao Thé XXVII*, 27-40.
- Peltzer DA, Ferriss S, FitzJohn RG** (2008) Predicting weed distribution at the landscape scale: using naturalized Brassica as a model system. *Journal of Applied Ecology* 45, 467-475.
- Pendergrast M** (2009) Coffee second only to oil? . *Tea and Coffee trade Journal* April 2009, 38-41.
- Perfecto I, Mas A, Dietsch T, Vandermeer J** (2003) Conservation of biodiversity in coffee agroecosystems: a tri-taxa comparison in southern Mexico. *Biodiversity and Conservation* 12, 1239-1252.
- Perfecto I, Vandermeer J, Mas A, Pinto LS** (2005) Biodiversity, yield, and shade coffee certification. *Ecological Economics* 54, 435-446.
- Pesch R, Pehlke H, Jerosch K, Schroder W, Schluter M** (2008) Using decision trees to predict benthic communities within and near the German Exclusive Economic Zone (EEZ) of the North Sea. *Environ Monit Assess* 136, 313-325.
- Peterson AT, Robins CR** (2003) Using ecological-niche modeling to predict Barred Owl invasions with implications for Spotted Owl conservation. *Conservation Biology* 17, 1161-1165.
- Peterson AT, Sober n J, Sanchez-Cordero VV** (1999) Conservatism of ecological niches in evolutionary time. *Science* 285, 1265-1267.
- Petit RJ, Aguinalalde I, de Beaulieu JL, et al.** (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* 300, 1563-1565.
- Pezzopane JRM, Pedro MJ, de Camargo MBP, Faztoli LC** (2008) Heat requirements of Mundo Novo coffee for the flowering-harvest phenological stage. *Ciencia E Agrotecnologia* 32, 1781-1786.
- Philpott SM, Arendt WJ, Armbrecht I, et al.** (2008) Biodiversity Loss in Latin American Coffee Landscapes: Review of the Evidence on Ants, Birds, and Trees. *Conservation Biology* 22, 1093-1105.

- Philpott SM, Perfecto I, Vandermeer J** (2006) Effects of management intensity and season on arboreal ant diversity and abundance in coffee agroecosystems. *Biodiversity and Conservation* 15, 139-155.
- Pillon P** (1985a) Caféculture : une activité transitoire *Source Actuel Développement*, p 24.
- Pillon P** (1985b) Développement et enjeux sociaux en Nouvelle Calédonie : l'Opération Café *Les Temps Modernes* 41 p. 1623-1653.
- Pinto MA, Rubink WL, Patton JC, Coulson RN, Johnston JS** (2005) Africanization in the United States: replacement of feral European honeybees (*Apis mellifera* L.) by an African hybrid swarm. *Genetics* 170, 1653-1665.
- Piry S, Alapetite A, Cornuet JM, et al.** (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. *J Hered* 95, 536-539.
- Plana V** (2004) Mechanisms and tempo of evolution in the African Guineo-Congolian rainforest. *Philos Trans R Soc Lond B Biol Sci* 359, 1585-1594.
- Plechakova O, Tranchant-Dubreuil C, Benedet F, et al.** (2009) MoccaDB - an integrative database for functional, comparative and diversity studies in the Rubiaceae family. *Bmc Plant Biology* 9, -.
- Poncet V, Dufour M, Hamon P, et al.** (2007) Development of genomic microsatellite markers in *Coffea canephora* and their transferability to other coffee species. *Genome* 50, 1156-1161.
- Poncet V, Hamon P, Minier J, et al.** (2004) SSR cross-amplification and variation within coffee trees (*Coffea* spp.). *Genome* 47, 1071-1081.
- Poncet V, Rondeau M, Tranchant C, et al.** (2006) SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. *Molecular Genetics and Genomics* 276, 436-449.
- Pouliot DA, King DJ, Bell FW, Pitt DG** (2002) Automated tree crown detection and delineation in high-resolution digital camera imagery of coniferous forest regeneration. *Remote Sensing of Environment* 82, 322-334.
- Pounds JA, Fogden MPL, Campbell JH** (1999) Biological response to climate change on a tropical mountain. *Nature* 398, 611-615.
- Powell W, Morgante M, Andre C, et al.** (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* 2, 225-238.
- Prakash NS, Combes M-C, Dussert Sp, Naveen S, Lashermes P** (2005) Analysis of genetic diversity in Indian robusta coffee gene pool (*Coffea canephora*) in comparison with a representative core collection using SSRs and AFLPs. *Genetic Resources and Crop Evolution* 52, 333-343.
- Preston CD, Pearman DA, Dines TD** (2002) *New atlas of the british flora* Oxford University Press.
- Pritchard JK, Stephens M, Donnelly P** (2000) Inference of population structure using multilocus genotype data. *Genetics* 155, 945-959.
- Pulliam HR** (1988) Sources, sinks and population regulation. *American Naturalist* 132, 652-661.
- Pulliam HR** (2000) On the relationship between niche and distribution. *Ecology Letters* 3, 349-361.
- Quinlan JR** (1993) *C4.5.: Programs for machine learning*, Morgan Kaufmann Publishers.
- Rallo P, Tenzer I, Gessler C, et al.** (2003) Transferability of olive microsatellite loci across the genus *Olea*. *Theor Appl Genet* 107, 940-946.
- Rannala B, Mountain JL** (1997) Detecting immigration by using multilocus genotypes. *Proc Natl Acad Sci U S A* 94, 9197-9201.

- Read JM, Clark DB, Venticinque EM, Moreira MP** (2003) Application of merged 1-m and 4-m resolution satellite data to research and management in tropical forests. *Journal of Applied Ecology* 40, 592-600.
- Reznick DN, Ghalambor CK** (2001) The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica* 112-113, 183-198.
- Rhymer JM, Simberloff D** (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics* 27, 83-109.
- Ricciardi A** (2007) Are modern biological invasions an unprecedented form of global change? *Conservation Biology* 21, 329-336.
- Richardson JE, Pennington RT, Pennington TD, Hollingsworth PM** (2001) Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science* 293, 2242-2245.
- Rieseberg LH** (1997) Hybrid origins of plant species. *Annual Review of Ecology and Systematics* 28, 359-389.
- Rieseberg LH, Archer MA, Wayne RK** (1999a) Transgressive segregation, adaptation and speciation. *Heredity* 83, 363-372.
- Rieseberg LH, Raymond O, Rosenthal DM, et al.** (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301, 1211-1216.
- Rieseberg LH, Sinervo B, Linder CR, Ungerer MC, Arias DM** (1996) Role of gene interactions in hybrid speciation: Evidence from ancient and experimental hybrids. *Science* 272, 741-745.
- Rieseberg LH, Whitton J, Gardner K** (1999b) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 152, 713-727.
- Rieseberg LH, Willis JH** (2007) Plant speciation. *Science* 317, 910-914.
- Rieseberg LH, Kim SC, Randell RA, et al.** (2007) Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica* 129, 149-165.
- Riley SPD, Shaffer HB, Voss SR, Fitzpatrick BM** (2003) Hybridization between a rare, native tiger salamander (*Ambystoma californiense*) and its introduced congener. *Ecological Applications* 13, 1263-1275.
- Robbrecht E** (1996a) Generic distribution patterns in subsaharan African Rubiaceae (Angiospermae). *Journal of Biogeography* 23, 311-328.
- Robbrecht E** (1996b) Generic Distribution Patterns in Subsaharan African Rubiaceae (Angiospermae). *Journal of Biogeography* 23, 311-328.
- Robertson MP, Villet MH, Palmer AR** (2004) A fuzzy classification technique for predicting species' distributions: applications using invasive alien plants and indigenous insects. *Diversity and Distributions* 10, 461-474.
- Rouse J, Haas R, Schell J, Deering D, Harlan J** (1974) Monitoring the vernal advancement of retrogradation of natural vegetation (ed. NASA / GSFC TI, Final Report), p. 371, Greenbelt, MD, US.
- Rousset F** (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145, 1219-1228.
- Rova JHE, Delprete PG, Andersson L, Albert VA** (2002) A trnL-F cpDNA sequence study of the Condamineae-Rondeletieae-Sipaneeae complex with implications on the phylogeny of the Rubiaceae. *American Journal of Botany* 89, 145-159.
- Saatchi S, Buermann W, Ter Steege H, Mori S, Smith TB** (2008) Modeling distribution of Amazonian tree species and diversity using remote sensing measurements. *Remote Sensing of Environment* 112, 2000-2017.

- Salomon M** (2001) Evolutionary biogeography and speciation: essay on a synthesis. *Journal of Biogeography* 28, 13-27.
- Schlotterer C** (2004) The evolution of molecular markers - just a matter of fashion? *Nature Reviews Genetics* 5, 63-69.
- Seehausen O** (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution* 19, 198-207.
- Seehausen O, Takimoto G, Roy D, Jokela J** (2008) Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology* 17, 30-44.
- Sehgal RNM, Jones HI, Smith TB** (2005) Molecular evidence for host specificity of parasitic nematode microfilariae in some African rainforest birds. *Molecular Ecology* 14, 3977-3988.
- Servedio MR, Noor MAF** (2003) The role of reinforcement in speciation: Theory and data. *Annual Review of Ecology Evolution and Systematics* 34, 339-364.
- Shaw KL** (1998) Species and the diversity of natural groups. In: *Endless forms: species and speciation*, pp. 44-56. Oxford University Press, Oxford.
- Silvertown J** (2004) Plant coexistence and the niche. *Trends in Ecology & Evolution* 19, 605-611.
- Silvestrini M, Junqueira MG, Favarin AC, et al.** (2007) Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellites markers. *Genetic Resources and Crop Evolution* 54, 1367-1379.
- Simpson GG** (1961) *Principles of animal taxonomy* Columbia University Press, New York.
- Solis-Montero L, Flores-Palacios A, Cruz-Angon A** (2005) Shade-coffee plantations as refuges for tropical wild orchids in central Veracruz, Mexico. *Conservation Biology* 19, 908-916.
- Sonke B, Stoffelen P** (2004) Une nouvelle espèce de *Coffea* L. (Rubiaceae, Coffeae) du Cameroun et quelques notes sur ses affinités avec les espèces voisines. *Adansonia* 26, 153-160.
- Soto-Pinto L, Romero-Alvarado Y, Caballero-Nieto J, Segura Warnholtz G** (2001) Woody plant diversity and structure of shade-grown-coffee plantations in northern Chiapas, Mexico. *Rev Biol Trop* 49, 977-987.
- Soto-Pinto L, Villalvazo-Lopez V, Jimenez-Ferrer G, et al.** (2007) The role of local knowledge in determining shade composition of multistrata coffee systems in Chiapas, Mexico. *Biodiversity and Conservation* 16, 419-436.
- Stebbins GL** (1974) *Flowering Plants: Evolution Above the Species Level*, Cambridge, MA.
- Stoffelen P, Robbrecht E, Smets E** (1997) Pollen morphology of *Coffea* and *Psilanthus* (Rubiaceae-Coffeae), mainly from Africa. *Grana* 36, 313-327.
- Swenson NG** (2008) The past and future influence of geographic information systems on hybrid zone, phylogeographic and speciation research. *Journal of Evolutionary Biology* 21, 421-434.
- Taylor EB, Boughman JW, Groenenboom M, et al.** (2006) Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology* 15, 343-355.
- Templeton AR** (1989) The meaning of species and speciation: a genetic perspective. In: *Speciation and its consequences*, pp. 3-27. Sinauer Associates, Sunderland MA.
- Tercinier G** (1958) Caféiculture et sols en Nouvelle-Calédonie. *Revue Agricole Nouvelle-Calédonie* N°3, 1 - 20.
- Thomas CD, Cameron A, Green RE, et al.** (2004) Extinction risk from climate change. *Nature* 427, 145-148.
- Thompson JN** (1998) Rapid evolution as an ecological process. *Trends in Ecology & Evolution* 13, 329-332.

- Thuiller W, Albert C, Araujo MB, et al.** (2008) Predicting global change impacts on plant species' distributions: Future challenges. *Perspectives in Plant Ecology Evolution and Systematics* 9, 137-152.
- Thuiller W, Araujo MB, Lavorel S** (2004) Do we need land-cover data to model species distributions in Europe? *Journal of Biogeography* 31, 353-361.
- Tiranti B, Negri V** (2007) Selective microenvironmental effects play a role in shaping genetic diversity and structure in a *Phaseolus vulgaris* L. landrace: implications for on-farm conservation. *Mol Ecol* 16, 4942-4955.
- Turelli M, Barton NH, Coyne JA** (2001) Theory and speciation. *Trends Ecol Evol* 16, 330-343.
- Turner M** (2001) *Landscape Ecology in theory and practice* Springer-Verlag.
- Turner W, Spector S, Gardiner N, et al.** (2003) Remote sensing for biodiversity science and conservation. *Trends in Ecology & Evolution* 18, 306-314.
- Valbuena-Carabana M, Gonzalez-Martinez SC, Sork VL, et al.** (2005) Gene flow and hybridisation in a mixed oak forest (*Quercus pyrenaica* Willd. and *Quercus petraea* (Matts.) Liebl.) in central Spain. *Heredity* 95, 457-465.
- van Valen L** (1976) Ecological species, multispecies, and oaks. *Taxon* 25, 233-239.
- Vayssières MP, Plant RE, Allen-Diaz BH** (2000) Classification trees: an alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science* 11, 679-694.
- White F** (1979a) The guineo-congolian region and its relationships to the other phytocoria. *Bull. Jard. Bot. Nat. Belg.*, 55.
- White F** (1979b) The guineo-congolian region and its relationships to the other phytocoria. *Bulletin du Jardin botanique national de Belgique* 49.
- Whitmore T, C.** (1982) On pattern and process in forests. In: *The plant community as a working mechanism* (ed. Newman EI), pp. 45-59. Blackwell Scientific Publications, Oxford, UK.
- Wiley EO** (1978) The evolutionary species concept reconsidered. *Systematic Zoology* 27, 17-26.
- Wilkinson MJ, Elliott LJ, Allainguillaume J, et al.** (2003) Hybridization between *Brassica napus* and *B. rapa* on a national scale in the United Kingdom. *Science* 302, 457-459.
- Wittaker RJ, Fernandez-palacios JM** (2007) *Island Biogeography* Oxford University Press, Oxford, UK.
- Witten I, H., Frank E** (2005) *Data Mining: Practical machine learning tools and techniques*, 2nd edn. Morgan Kaufman, 2005, San Francisco.
- Woodward FI, Cramer W** (1996) Plant functional types and climatic changes: Introduction. *Journal of Vegetation Science* 7, 306-308.
- Wright JW** (1976) Introduction to forest genetics. *Academi Press, san Diego, CA*.
- Yesson C, Culham A** (2006) Phyloclimatic modeling: combining phylogenetics and bioclimatic modeling. *Syst Biol* 55, 785-802.
- Zhang BS, Valentine I, Kemp P, Lambert G** (2006) Predictive modelling of hill-pasture productivity: integration of a decision tree and a geographical information system. *Agricultural Systems* 87, 1-17.
- Zhang BS, Valentine I, Kemp PD** (2005) A decision tree approach modelling functional group abundance in a pasture ecosystem. *Agriculture Ecosystems & Environment* 110, 279-288.
- Zhang XS, Wang JL, Hill WG** (2004) Redistribution of gene frequency and changes of genetic variation following a bottleneck in population size. *Genetics* 167, 1475-1492.
- Zimmermann NE, Edwards TC, Moisen GG, Frescino TS, Blackard JA** (2007) Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *Journal of Applied Ecology* 44, 1057-1067.

Résumé :

Les caféiers sont des arbres ou arbustes natifs des forêts intertropicales d'Afrique, de Madagascar, des Mascareignes et des Comores et leurs aires d'origine ou de diversification sont en forte régression. Parmi les 103 espèces du genre *Coffea* seules trois sont cultivées : *C. arabica*, *C. canephora* et *C. liberica*. L'espèce *C. canephora* est celle du genre qui a la plus grande distribution dans les forêts tropicales africaines, et sa diversité génétique est structurée en cinq groupes bien distincts. L'étude fine de la structuration de sa diversité génétique a révélé une importance particulière de la rupture floristique du Dahomey gap en tant que barrière biogéographique, ainsi que des variations climatiques du passé sur la structuration de la diversité (modèle d'étude des refuges forestiers du Quaternaire). Une collection de référence (core-collection) de la diversité sauvage et cultivée (représentée par une majorité d'hybrides inter-groupes) a été mise en place pour caractériser la diversité introduite en Nouvelle-Calédonie (depuis 1856). Cette diversité s'est révélée être élevée pour *C. canephora* et *C. liberica* et comme attendu, faible pour *C. arabica*. *C. canephora* et *C. arabica* (*C. liberica* de manière plus anecdotique) ont été introduites un peu partout en Nouvelle-Calédonie pour leur culture. Pour des raisons économiques, beaucoup de plantations traditionnelles ont été abandonnées au moment de la deuxième guerre mondiale, laissant les arbres de caféiers évoluer de manière naturelle. Des hybridations inter-spécifiques se produisent entre ces trois espèces (à hauteur de 3% dans une population tri-spécifique) grâce à une levée de barrières à la reproduction, permise par une conjoncture environnementale particulière. Une expertise environnementale fine, basée sur la comparaison des conditions environnementales entre la région de Sarraméa et les régions d'origine des trois espèces en Afrique, a permis de mettre en exergue les combinaisons environnementales favorables à la cohabitation des espèces et à leur hybridation inter-spécifique. L'équilibre entre drainage et humidité, un régime pluviométrique particulier et la présence d'un couvert forestier, sont des paramètres environnementaux importants qui maintiennent un microclimat favorable. L'étude du couvert forestier par télédétection a permis de renforcer le modèle de prédiction des niches favorables aux hybridations inter-spécifiques grâce à l'information apportée quant à l'hétérogénéité et la présence de gros arbres dans la canopée ainsi que à l'amélioration de la résolution spatiale qui permet la détection fine de micro-habitats. Les niches favorables sont des micro-environnements répartis dans la région d'étude. Il s'agit de mettre l'accent sur la conservation d'une telle néo-diversité adaptée et de valoriser ce patrimoine naturel.

Abstract :

Coffee trees are originated from inter-tropical forests of Africa and Madagascar and their diversification origin areas are highly threatened. Among the 103 species of the *Coffea* genus, only three are cultivated: *C. arabica*, *C. canephora* and *C. liberica*. *C. canephora* has the widest natural distribution in tropical African forests, and its genetic diversity is structured in five distinct groups. The study of the genetic diversity structuring has revealed a particular importance of the Dahomey gap floristic breaking as biogeographic barrier, as well as the past climatic variations on the diversity structuring (model to study the "refuge theory"). A core set of both wild and cultivated (represented in majority by inter-group hybrids) accessions was selected to characterize the introduced diversity in New-Caledonia (since 1856). On introduced New-Caledonian sample, high levels of genetic diversity have been revealed for *C. canephora* and *C. liberica* while as expected, low level of diversity has been revealed for *C. arabica*. These three species (*C. liberica* to a lesser extent) have been introduced in different regions of New Caledonia for their culture. Because of economical reasons, a lot of traditional coffee plantations have been abandoned during the Second World War, leaving the cultivars to evolve in natural conditions. Inter-specific hybridizations are occurring between coffee species (at a level of 3% according to the study of a tri-specific population) thanks to a removal of reproductive barriers facilitated by particular environmental conditions. An environmental expertise based on the comparisons of environmental conditions between Sarramea region and African origin regions of the three species has underlined the favorable environmental combinations to sympatry and inter-specific hybridizations. Equilibrium between draining and humidity, precipitations sequences and the forest cover are important parameters maintaining a favorable microclimate. The canopy forest study by remote sensing tools has reinforced the predicting distribution model of favorable niches to inter-specific hybridization thanks to the information brought on canopy heterogeneity and large tree crowns and to the spatial resolution improving the detection of micro-habitats. Favorable niches are micro-environments distributed throughout the study area. The question is to underline the preservation of such de novo adapted coffee genetic resources.