

Université des Antilles et de la Guyane

Faculté des Sciences exactes et naturelles

École doctorale pluridisciplinaire : Santé, Environnement et Sociétés dans les Amériques

Thèse pour le doctorat en génétique écologique

MONTAIGNE William

***Diversité génétique et adaptation au milieu
chez les arbres forestiers tropicaux
Etude chez le genre Virola (Myristicaceae)***

Sous la direction

d'Ivan SCOTTI et de Caroline SCOTTI-SAINTAGNE

Soutenue le 15 décembre 2011 à Kourou

N : 2011AGUY0480

Jury :

Giovanni VENDRAMIN, Directeur de Recherche, Plant Genetic Institut – CNR, Université de Florence, Italie, Rapporteur

Rosane COLLEVATTI, Professeur, Université fédérale de Goiás, Brésil, Rapporteur

Henry CARON, Ingénieur de Recherche, INRA UMR BioGeCo, Bordeaux 1, Examineur

Daniel SABATIER, Chargé de Recherche, IRD UMR AMAP, Montpellier 2, Examineur

Jean-Christophe ROGGY, Chargé de Recherche, INRA UMR EcoFoG, UAG, Examineur

Ivan SCOTTI, Chargé de Recherche, INRA UMR EcoFoG, UAG, Directeur de Thèse

*A Thom & Moïse,
A la vôtre vieux bouzins !*

Remerciements

Evidemment, je tiens avant tout à te remercier Ivan, et ce pour plusieurs raisons. Tout d'abord, merci de m'avoir accueilli dans ton labo et de m'avoir fait confiance pour ce travail. Merci de m'avoir initié à la génétique des populations et de m'avoir permis de mener à bien cette étude. Merci de m'avoir fait partager tes grandes connaissances scientifiques et ta vision de la recherche. Et, le plus important, MERCI d'avoir été bien plus qu'un simple directeur de thèse : je pense à ces moments extraprofessionnels de complicité, aussi bien sur nos vélos, sur le stade « Thomas Leclerc » et bien sûr dans les soirées kourouciennes... Je trouve que ça commence à faire un peu trop de merci pour un seul homme, mais bon, tu le mérites. Pour les mêmes raisons, je t'adresse toute ma reconnaissance Caro (sauf pour le côté sport, mais y'a un début un tout !). Encore merci à vous deux.

Je souhaite également remercier tout le labo de génétique écologique pour leurs divers petits coups de pouce et la bonne ambiance. Merci Valérie et Eliane pour l'aide apportée en laboratoire. Merci Saint-Omer et Pierre pour votre aide lors des échantillonnages sur le terrain. Merci à vous Delphine, Maxime, Steph et Louise d'avoir partagé une partie de votre vie de thésard avec moi. Merci tous les stagiaires pour votre passage, ça fait du bien de voir des nouvelles têtes de temps en temps. Et plus généralement, merci à tous les autres collègues, qui de près ou de loin, ont permis à cette thèse d'aboutir.

Merci G. Vendramin et R. Collevatti d'avoir accepté d'être les rapporteurs de cette thèse. Et également Jean-Chris, Henry et Daniel d'avoir bien voulu compléter ce jury. Daniel, j'en profite aussi pour te remercier de ton œil d'expert dans la reconnaissance des Virola. Un autre merci à François Lefèvre et Joëlle Ronfort de m'avoir guidé lors des comités de thèse.

Parce qu'il faut savoir profiter de la vie un maximum (même si c'est un peu moins évident en fin de thèse...), je tiens à adresser un ENORME merci à vous tous mes Amis. Merci les colocs : Dodo et Greg, d'abord potes en Gwada, puis colocs en Guyane... On se rejoint dans quelle partie du monde la prochaine fois ? Louise, on se recroisera très certainement dans la capitale ou dans nos campagnes perdues. Merci Claire et Xav l'hélicoptère, auteurs de mes plus grands moments de poilades. Merci Luc et Seb pour votre zénitude. Merci, Juliette (on n'a pas encore eu trop le temps de se connaître, mais je sais déjà que tu es une super coloc). Merci Merlin, on est arrivé ensemble, on repartira ensemble ! Si on avait été tous colocs en même temps, ça aurait fait un joyeux bordel ! Un gargantuesque merci à tous les copains rencontrés ici, je ne fais pas la liste car vous êtes trop nombreux et j'ai la flemme ^^ . Merci

Bouc et Po d'avoir fait le déplacement pour me faire un coucou. Merci les bouzins, c'est toujours un immense plaisir de rentrer et de vous revoir.

Merci à vous les parents de ne jamais vous être opposés à mes envies, et d'être toujours aussi présents malgré les milliers de kilomètres qui nous séparent. Je pense fort à vous.

Mes excuses à tous ceux que j'ai pu oublier.

A Fanny, pour ce que tu sais et que je n'ai pas besoin d'écrire ici. Il serait trop difficile de décrire tout ce que tu m'as apporté avec seulement quelques mots.

Résumé

La diversité génétique est une composante essentielle de la biodiversité. Le maintien de ressources génétiques suffisamment larges est nécessaire pour assurer la viabilité et le potentiel évolutif des populations naturelles. Dans le contexte actuel des changements globaux, une question essentielle est de savoir si ces espèces pourront s'adapter à de nouvelles conditions environnementales, notamment chez les arbres forestiers tropicaux qui sont à la base de la structuration d'un des écosystèmes les plus riches de la planète. Cette thèse a le principal objectif de caractériser la diversité et la variabilité génétique chez les arbres forestiers tropicaux pour décrire les processus évolutifs qui en sont à l'origine. L'étude menée au niveau interspécifique et intraspécifique se focalise sur le genre *Virola* (Myristicaceae) qui constitue un bon modèle pour l'étude de la génétique écologique.

Si une forte diversité génétique globale est gage de prospérité pour le futur des espèces, il paraît primordial d'effectuer des études d'impact suite à certaines perturbations du milieu. Ainsi, dans la première partie de cette thèse, une étude de la régénération d'un échantillonnage exhaustif de *V. michelii* a été menée dans une parcelle du dispositif expérimental de Paracou ayant subi une exploitation forestière de faible intensité et comparée à une parcelle témoin. La diversité génétique mesurée à partir de marqueurs AFLP (*Amplified Fragment-Length Polymorphism*, N = 229) en zones perturbées s'est révélée être plus grande qu'en zones non-perturbées.

L'étude de la diversité génétique intra-spécifique peut s'avérer utile pour la compréhension de la capacité d'adaptation d'une espèce aux variations de son environnement et donc de sa pérennité dans le temps. Alors, dans la seconde partie, celle-ci a été étudiée au sein d'une population de *V. michelii* à l'échelle locale sur le site expérimental de Paracou. Une approche par criblage génomique avec des marqueurs AFLP (N = 229) a été entreprise sur 77 individus adultes et 401 individus juvéniles, afin d'identifier des différences dans la composition génétique de sous-populations associées à des conditions contrastées pour différentes variables environnementales. Certains loci (N = 2) ont été révélés comme soumis à la sélection divergente pour des habitats contrastés, indiquant un signe d'adaptation.

Dans la dernière partie, l'étude des niveaux de divergence génétique chez trois espèces de *Virola* du bouclier guyanais (*V. michelii*, *V. surinamensis* et *V. kwatae*) pour des marqueurs moléculaires nucléaires et chloroplastiques a été menée. Il en résulte que deux d'entre elles (*V. surinamensis* et *V. kwatae*) montrent de fortes similarités génétiques malgré leur

distribution sur des environnements contrastés. Les analyses par le coalescent ont révélé d'une part une divergence récente entre ces deux espèces laissant supposer une spéciation écologique, et d'autre part des flux de gènes interspécifiques agissant sur ces deux espèces-sœurs.

Ce travail a permis d'aborder différents processus évolutifs à l'origine de la diversité génétique actuelle chez ces espèces forestières tropicales. Il fournit alors une contribution pour appréhender le devenir des populations et peut avoir une utilité dans les programmes de conservation afin de préserver le maximum de diversité dans la nature.

Mots-clés : diversité génétique, divergence adaptative, arbres forestiers tropicaux *Viola*, sélection naturelle, flux de gènes interspécifiques, spéciation écologique.

Abstract

Genetic diversity is an essential component of biodiversity. The maintenance of sufficient genetic resources is needed to ensure the adaptive potential and the viability of natural populations. In the current context of global changes, the study of adaptation in living organisms is a key task, particularly for tropical forest trees that are dominant components (in terms of biomass and as ecological drivers) of some of the most diverse ecosystems on Earth. The main objective of this thesis is to characterize genetic diversity and genetic variability to understand the evolutionary processes that act on them. This ecological-genetic study was carried out at the interspecific and intraspecific level in the *Virola* genus.

If overall high levels of genetic diversity are a guarantee of prosperity for the future of the species, it seems essential to perform studies on the impact of environmental disturbance on genetic diversity. In the first section, the genetic consequences of regeneration dynamics were studied in an exhaustive sample of *V. michelii* in a low-intensity logging plot and in a control plot at the Paracou experimental site. A greater genetic diversity, measured from AFLP markers (Amplified Fragment Length Polymorphism, $N = 229$), was found in perturbed areas.

Because studying genetic diversity within species may be useful for understanding species adaptation to environmental changes, in the second section. I studied local adaptation in a population of *V. michelii* on the Paracou experimental site. A genome scan approach with AFLPs ($N = 229$) was conducted on 77 adult individuals and 401 juveniles to identify genetic differences between populations associated to contrasting conditions for an array of environmental variables. Some loci ($N = 2$) were found to be subject to divergent selection, indicating adaptation to contrasting habitats.

In the third section, the study of levels of genetic divergence in three *Virola* species of the Guiana Shield (*V. michelii*, *V. surinamensis* and *V. kwatae*) was investigated for nuclear and chloroplast molecular markers. *V. surinamensis* and *V. kwatae* showed strong genetic similarities despite their contrasting habitats preferences. Coalescent analyses have revealed, on one hand, a recent divergence between these two species suggesting an ecological speciation, and on the other hand that interspecific gene flow occurs between these sister-species.

This work focuses on understanding evolutionary processes shaping genetic diversity and provides a useful contribution for biodiversity conservation programs.

Keywords: genetic diversity, adaptive divergence, tropical forest *Virola* genus, natural selection, interspecific gene flow, ecological speciation.

Avant-propos

Ce manuscrit présente les résultats de ma thèse au sein de l'INRA de l'Unité Mixte de Recherche « Ecologie des forêts de Guyane » (UMR EcoFoG) portant sur la diversité génétique et de l'adaptation chez les espèces d'arbres tropicaux. Plus particulièrement, ces vastes domaines disciplinaires ont été entrepris chez le genre *Virola* (Myristicacées) qui m'a servi de modèle d'étude.

Ce travail de recherche a été réalisé depuis la conception du sujet avec Ivan Scotti, jusqu'à l'élaboration de ce mémoire, en passant par les phases d'échantillonnage sur le terrain, de manipulation en laboratoire, d'analyse de données et par la fameuse phase de rédaction d'articles scientifiques pour des publications à comité de lecture.

Cette thèse est structurée en trois grandes parties :

- (i) La première constitue l'introduction générale et décrit le contexte et les bases scientifiques et méthodologiques se référant au sujet.
- (ii) La deuxième partie correspond aux différentes études entreprises pour cette thèse. Elle est composée de trois chapitres (qui peuvent être lu indépendamment) qui sont présentés en anglais, sous la forme d'articles scientifiques. Toutefois, un résumé en français précédera chaque article.
- (iii) La dernière partie propose une conclusion générale qui synthétise les résultats obtenus et offre quelques perspectives de recherche.

Cette thèse s'inscrit dans plusieurs projets menés par l'UMR EcoFoG:

- Le projet ENERGI-RAVI (Feder 2008 - 2011) a pour but d'étudier les mécanismes qui caractérisent la grande diversité du vivant en Amazonie ainsi que la distribution de cette richesse. Plus particulièrement, ce programme vise à approfondir la connaissance des mécanismes génétiques qui déterminent et résultent de l'interaction entre les organismes et leur environnement. La génétique écologique définit ce champ d'étude qui croise les domaines scientifiques de l'évolution, de l'écologie et de la génétique. Cette discipline vise à comprendre comment la diversité génétique peut être utilisée et gérée, et donc permettra d'aider à la définition de plans de gestion et d'exploitation durable des ressources du vivant.
- Le projet BRIDGE (ANR 2007 - 2010) cherche à améliorer nos connaissances sur les différentes dimensions de la diversité biologique des arbres de Guyane française (spécifique, fonctionnelle, évolutive).

- Le projet « influence de l'intensité d'exploitation et du degré d'ouverture de la canopée en forêt tropicale humide sur le maintien et la dynamique de la biodiversité » du programme BGF (Biodiversité et Gestion Forestière, Ministère de la Recherche, GIP-ECOFOR 2005 - 2009) vise l'étude de la dynamique des peuplements et des traits d'espèces ainsi que la définition de règles de gestion durable. Ce projet a pour but d'étudier les conséquences des ouvertures de la canopée sur la dynamique des communautés d'arbres, la richesse spécifique et la diversité génétique.

Bonne lecture

Tables des matières

INTRODUCTION	15
I. Contexte	16
A. Des changements environnementaux sans précédent.....	16
B. L'adaptation comme issue de secours	17
II. L'origine de la diversité génétique	19
A. Le polymorphisme génétique.....	19
B. Les forces évolutives	20
1. Les mutations.....	20
2. La dérive génétique.....	24
3. La migration	24
4. La sélection naturelle	26
C. L'impact de la démographie sur la diversité génétique.....	28
1. La taille efficace des populations.....	28
2. Les changements démographiques	28
III. La diversité génétique comme moteur de l'adaptation et de la spéciation.....	33
A. La détection de la sélection naturelle	33
1. La théorie neutre de l'évolution	33
2. Les tests de neutralité.....	34
a) généralités.....	34
b) L'approche de criblage génomique (« genome scan »).....	35
3. Les interférences avec la démographie	36
B. L'adaptation locale	37
C. De l'adaptation vers la spéciation	38
1. Les modes de spéciation	39
2. La spéciation écologique.....	40
3. L'hybridation et les complexes d'espèces.....	41
IV. Problématique de thèse.....	43
A. Modèles d'étude	43
B. Sites d'étude.....	44
1. La Guyane française	44
2. Le dispositif expérimental de Paracou.....	45
C. Objectifs	48

RESULTATS	51
I. Article 1: Regeneration dynamics of a tropical rainforest tree: does intermediate disturbance maintain genetic diversity?	52
A. Résumé	52
B. Article	54
II. Article 2: Natural selection and within-population divergence in response to microenvironmental heterogeneity in a Neotropical nutmeg tree (Myristicaceae)	92
A. Résumé	92
B. Article	95
C. Travaux supplémentaires hors-article.....	120
III. Article 3: Sisters, but not alike: ecological divergence with gene flow in a trio of Neotropical nutmeg tree species (Myristicaceae)	124
A. Résumé	124
B. Article	129
CONCLUSION	161
BIBLIOGRAPHIE.....	167

Tables des illustrations

Table des encadrés

Encadré 1: Caractérisation de la diversité génétique.....	21
Encadré 2: Les différents marqueurs moléculaires	22
Encadré 3 : L'équilibre de Hardy-Weinberg.....	23
Encadré 4: Le modèle en île de Wright.....	25
Encadré 5 : Le modèle de Wright-Fisher	28
Encadré 6 : La théorie de la coalescence.....	29
Encadré 7 : Les méthodes bayésiennes	30
Encadré 8 : Les conséquences de l'exploitation forestière	31
Encadré 9 : Le test du K_A / K_S	34
Encadré 10 : Les différents génomes des plantes et quelques caractéristiques.....	49

Table des figures

Figure 1: Illustration des différentes formes de sélection au niveau moléculaire au sein d'une population.....	27
Figure 2 : les différents types de spéciation	40
Figure 3 : carte des pluviométries en Guyane française.....	45
Figure 4 : schéma du dispositif de Paracou.....	47
Figure 5: réseau haplotypique formé à partir de 35 séquences chez <i>Virola sebifera</i> et <i>V. lorentensis</i> (d'après Steeves 2011).....	165

INTRODUCTION

I. Contexte

A. Des changements environnementaux sans précédent

Le monde du vivant est impressionnant de par son incroyable diversité de formes et d'adaptations. Les forêts tropicales humides, et notamment la forêt amazonienne, sont considérées comme le premier réservoir mondial de diversité biologique terrestre (Hoorn *et al.* 2010), notamment en Amérique centrale et du sud qui possèdent la plus grande surface forestière humide. La forêt amazonienne abrite une extraordinaire diversité biologique tant au niveau de la faune que de la flore ; à titre d'exemple la forêt guyanaise peut enregistrer de 120 à 200 espèces d'arbres sur un hectare, soit au moins autant que dans tout le territoire métropolitain français (Blanc *et al.* 2003) ! Malheureusement, ces hot-spots de biodiversité tropicale terrestre ne cessent de décroître à cause de la déforestation engendrée par l'Homme afin, par exemple, d'étendre ses cultures de rente et son élevage intensif de bétail, de construire ses routes et autres infrastructures ou d'exploiter les matières premières parfois non-renouvelables (Puig 2001). Ces activités ont également des conséquences sous-jacentes indirectes comme la production de gaz à effet de serre ayant pour effet le réchauffement de la planète. En effet, une augmentation de 80% du rejet en dioxyde de carbone a été enregistrée entre 1970 et 2004 (IPCC 2007). Celle-ci a pour conséquence de contribuer fortement à l'augmentation des températures d'une moyenne de 0,26°C par décennie au niveau de forêts tropicales (Malhi & Wright 2004) ; cette augmentation de température pourrait atteindre des seuils de 3°C à 8°C d'ici 2100. Il est attendu que ce changement climatique restera irréversible pour au moins mille ans (Solomon *et al.* 2009). En plus des changements de températures, le régime des précipitations est également perturbé depuis ces dernières années. En effet, des augmentations significatives des événements de sécheresse ont déjà été décrites pour la région amazonienne et ces événements sont considérés par certains auteurs comme la menace la plus importante pour les forêts tropicales (Phillips *et al.* 2009; Asner & Alencar 2010; Meir & Ian Woodward 2010). De plus, une augmentation de l'intensité des saisons sèches est également attendue sur le bassin amazonien (Li *et al.* 2008; Malhi *et al.* 2008).

Dans ce contexte des changements environnementaux, la biodiversité, et plus particulièrement la diversité génétique, est d'autant plus menacée qu'elle est essentielle pour assurer la viabilité et le potentiel évolutif des populations naturelles. Effectivement, une diversité génétique suffisante dans les populations naturelles leur permet une plus grande adaptabilité

face aux modifications du milieu, une réduction du risque de dépression de consanguinité affectant la fécondité et la survie, et diminue ainsi le risque d'extinction (Frankham 2003, 2005). La génétique écologique (qui est au croisement des domaines scientifiques de l'évolution, de l'écologie et de la génétique) a alors une implication majeure pour la biologie de la conservation (Stockwell *et al.* 2003; Lowe *et al.* 2004). Les chercheurs sont pessimistes sur les conséquences de telles perturbations relativement récentes, dont certaines sont encore inconnues. Nous ne savons pas encore comment réagira l'ensemble des organismes vivants face à ces rapides et importants changements environnementaux considérés comme responsables de l'actuelle 6^{ème} grande phase d'extinction des espèces depuis l'origine de notre planète.

Face à ces contraintes, trois types de réponses s'offrent aux organismes vivants (Aitken *et al.* 2008):

- la migration des populations vers des conditions environnementales plus favorables,
- l'adaptation des populations aux nouvelles conditions,
- ou fatalement, l'extinction des populations.

La migration chez les arbres peut sembler difficile étant donnée la rapidité des changements (Petit & Hampe 2006); ceux sont des organismes sessiles possédant de long temps de génération. De plus, les espèces ligneuses représentent l'essentiel de la biomasse de l'écosystème forestier tropical humide, et sont par leurs interactions avec différents organismes de la forêt à la base du réseau trophique. Il paraît donc indispensable d'étudier l'adaptation des espèces d'arbres et de leurs populations à leur milieu pour prédire leur devenir.

B. L'adaptation comme issue de secours

La notion d'adaptation est un sujet important de la biologie évolutive (Darwin 1859; Gould & Lewontin 1979; Orr 2005). L'adaptation peut être décomposée en deux mécanismes évolutifs distincts : la plasticité phénotypique et l'adaptation génétique. La plasticité phénotypique correspond à la capacité d'un organisme à modifier sa physiologie, sa morphologie ou son développement en réponse à des changements environnementaux (Callahan *et al.* 1997). Il s'agit essentiellement d'un mécanisme au niveau individuel qui s'exerce sur une durée relativement courte (vie de l'organisme). En revanche, la notion d'adaptation étudiée au sein de cette thèse : l'adaptation génétique, est le résultat de l'action

de la sélection naturelle qui augmente la fréquence des phénotypes les plus favorables à la survie et à la reproduction. On dit alors qu'ils ont une meilleure fitness, aussi appelée valeur sélective ou valeur adaptative. Par l'élimination des individus les moins adaptés à leur environnement, la fitness globale au niveau populationnel aura tendance à augmenter. Pour que l'adaptation puisse s'opérer, il faut que les caractères liés à la fitness soient au moins partiellement héritable, c'est-à-dire que la valeur de ces caractères soit contrôlée par des gènes transmissibles de génération en génération. D'autres caractères héritable peu liés à la valeur sélective peuvent évoluer de façon aléatoire par la dérive génétique (Lande 1976). Il paraît donc essentiel d'étudier la part de l'évolution qui est due à la dérive génétique et celle due à la sélection naturelle, autrement dit à l'adaptation, voir même d'essayer d'identifier l'action de cette sélection et les régions qui y sont soumises.

II. L'origine de la diversité génétique

La diversité génétique est une composante essentielle de la biodiversité. Elle décrit la variabilité des gènes entre ou à l'intérieur des espèces et de leurs populations. L'intérêt actuel porté à la biodiversité montre à quel point il est nécessaire de veiller au maintien d'une ressource génétique suffisamment large pour garantir l'adaptation des organismes face aux changements environnementaux directs (les exploitations forestières et les changements d'usage des terres par exemple) et indirects sur le long terme (changements climatiques globaux). En effet, cette diversité génétique est très importante car elle représente le support de base sur lequel peut agir la sélection. Il est admis que plus la diversité est grande dans un groupe d'individus (sous-population, population, espèce) plus il sera facile pour certains individus de s'adapter à de nouvelles conditions environnementales. En plus de permettre une plus grande adaptabilité des individus, elle permettra également de réduire la dépression de consanguinité et diminuera ainsi le risque d'extinction (Frankham 2003, 2005). L'étude de la diversité génétique à différents niveaux hiérarchiques d'organisation peut apporter une aide précieuse à la biologie des populations et à la biologie évolutive ; deux disciplines importantes pour la biologie de la conservation. La diversité génétique est alors devenue un outil primordial pour définir des buts, des méthodes et des priorités dans des programmes de conservation (Stockwell *et al.* 2003).

A. Le polymorphisme génétique

Le polymorphisme génétique est à la base de la diversité génétique, il correspond à des variations de séquences d'ADN au sein d'un groupe d'individus. Ces variations naturelles sont dues à des mutations successives au cours de l'évolution qui permettent de caractériser la diversité génétique entre individus et populations. En général, la majorité des polymorphismes sont neutres (Kimura 1968, 1983), mais une partie de ces variations peut influencer les différences phénotypiques observées entre individus qui pourront leur permettre d'avoir une meilleure fitness dans leur environnement. Le polymorphisme peut nous renseigner sur les différents processus qui façonnent la variabilité génétique. Ils peuvent être de nature démographique (taille des populations, migrations) et sélectif (lié à l'environnement). L'étude

de l'importance relative de ces différentes composantes façonnant le polymorphisme permet de mieux comprendre l'histoire évolutive des populations et des espèces.

B. Les forces évolutives

Le niveau de diversité génétique des populations et des variations de fréquences alléliques (Encadré 1) dépend de l'action respective de quatre forces évolutives pouvant interagir les unes avec les autres : la mutation, la sélection, la migration et la dérive. Elles sont à l'origine de la structure de la diversité génétique et de son évolution.

1. Les mutations

Les mutations sont des modifications de l'information génétique. Elles peuvent survenir aléatoirement au cours de la réplication de l'ADN mais peuvent aussi être dues à des agents mutagènes physiques (par exemple, les rayonnements ultra-violet), chimiques ou biotiques (virus et transposons). La recombinaison est un processus qui entraîne un brassage de matériel génétique intra- ou inter-chromosomique et dont la conséquence est la création de nouvelles combinaisons alléliques, c'est-à-dire d'haplotypes. Les mutations (et recombinaisons) qui augmentent ainsi la diversité génétique globale en générant de nouveaux allèles et de nouveaux haplotypes peuvent être de différentes classes. Le polymorphisme chromosomique peut être dû soit à une variation du nombre des chromosomes soit à un changement de leur structure (délétion, duplication, inversion, translocation). Le polymorphisme d'insertion caractérise des éléments d'ADN capables de s'insérer dans différentes régions du génome. Inversement, le polymorphisme de délétion supprime certaines régions du génome, alors que le polymorphisme d'inversion comme son nom l'indique provoque un changement dans l'ordre de certaines séquences dans le génome. Le polymorphisme ponctuel signifie des changements à l'échelle du nucléotide (SNP pour « Single Nucleotide Polymorphism »). Enfin, le polymorphisme de taille correspond à une variation du nombre de motifs répétés sur un brin d'ADN, comme nous pouvons l'observer chez les microsatellites (SSR pour « Simple Sequence Repeat »). Différents marqueurs moléculaires existent pour travailler sur ces différents types de polymorphismes (Encadré 2).

Les mutations affectent l'ensemble du génome. Celles-ci sont silencieuses lorsqu'elles n'ont pas d'impact évolutif. Par contre, des changements fonctionnels peuvent avoir lieu d'une part, lorsqu'elles entraînent des changements d'acides aminés (mutation non-synonyme) dans les régions codantes et donc des modifications structurales et fonctionnelles des protéines.

D'autre part, les mutations affectant les régions non-codantes peuvent aussi avoir un impact évolutif. En effet, elles peuvent entraîner des modifications dans les régions de l'ADN ayant un rôle important dans la régulation de la transcription et dans l'organisation du génome et qui sont donc soumises à des pressions de sélection indiquant qu'elles ont un rôle évolutif et fonctionnel majeur (Wang *et al.* 1999).

Le phénomène de mutation est la source fondamentale de variation génétique et correspond à un changement héréditaire dans le matériel génétique de l'organisme. Les mutations sont importantes car elles apportent de la richesse par la création de nouveaux allèles.

Encadré 1: Caractérisation de la diversité génétique

- Calcul de la diversité génétique :

Il existe plusieurs manières de calculer la diversité génétique. La diversité génétique de Nei He (1973) correspond à l'hétérozygotie attendue et caractérise le niveau de variabilité des gènes au sein d'une même population, c'est-à-dire qu'elle exprime la probabilité pour que deux variants (allèles) tirés au hasard dans une population soient différents (c'est une mesure équivalente à l'indice de Simpson (1949) en écologie). Ce paramètre fournit une bonne estimation de la variabilité génétique de la population :

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

Avec p_i la fréquence allélique pour chaque allèle et n le nombre d'allèle au locus considéré.

- Structuration de la diversité génétique :

Les F-statistiques (Wright 1969; Weir & Cockerham 1984) permettent de calculer la différenciation génétique entre populations (F_{ST}), aussi appelée indice de fixation (Nei & Chesser 1983). Cet indice permet de mesurer la différenciation des populations à partir du polymorphisme génétique et est calculé dans le cadre d'un modèle en île sous l'équilibre migration/dérive (Wright 1951). Les valeurs de cette mesure sont comprises entre 0 et 1 ; si les populations présentent les mêmes fréquences alléliques, alors $F_{ST} = 0$. Le F_{ST} peut être interprété comme la part de variance des fréquences alléliques due aux différences entre populations :

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

Avec H_S , l'hétérozygotie attendue pour une sous-population supposée à l'équilibre de Hardy-Weinberg (Encadré 3) et H_T la diversité moyenne des H_S sur les sous-populations et H_T l'hétérozygotie attendue pour une population globale supposée à l'équilibre de Hardy-Weinberg.

- Structuration spatiale de la diversité génétique:

Pour comprendre l'origine de la structuration de la diversité génétique entre populations, il est parfois utile de déterminer des distances génétiques entre paires de populations (par exemple avec les F_{ST}). Les matrices de distances génétiques peuvent alors être comparées à des matrices de distances (géographiques ou écologiques) entre sites afin d'évaluer leur corrélation. Le test de Mantel (Mantel 1967; Smouse *et al.* 1986) permet d'évaluer

cette corrélation par régression entre matrices à expliquer (distance génétique) et matrices explicatives (distance géographique ou écologique).

L'autocorrélation génétique spatiale permet une estimation du patron de répartition de la variable génétique. Elle va permettre de renseigner si les individus les plus proches géographiquement sont aussi les plus proches génétiquement. L'indice générique I de Moran (1950) est le plus souvent utilisé comme estimateur de l'autocorrélation des fréquences alléliques (Smouse & Peakall 1999):

$$I_k = \frac{n \sum_{i=1}^n \sum_{j \neq i}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{W \sum_{i=1}^n (x_i - \bar{x})^2} \quad W = \sum_{i=1}^n \sum_{j \neq i}^n w_{ij}$$

Avec n , nombre total d'individus ; $w_{ij} = 1$ si les individus i et j appartiennent à la classe de distance k , $w_{ij} = 0$ dans les autres cas ; $x_i = 1, 0.5$ ou 0 selon que l'individu i a 2, 1 ou 0 fois l'allèle considéré, \bar{x} représente la moyenne sur les n individus des fréquences alléliques individuelles.

D'autres méthodes pour analyser l'autocorrélation génétique spatiale s'avèrent plus spécifiques, notamment en utilisant des coefficients de parenté (Hardy & Vekemans 1999). Certaines analyses sont même conçues spécifiquement pour les marqueurs dominant en supposant des coefficients de consanguinité (Hardy 2003).

Encadré 2: Les différents marqueurs moléculaires

Le séquençage de tout le génome sur un grand nombre d'individus, bien que techniquement faisable, est encore aujourd'hui coûteux et n'est réalisable que pour quelques espèces modèles (e.g. Manel *et al.* 2010). De ce fait, les études de génétique des populations reposent le plus souvent sur l'analyse du polymorphisme au niveau de marqueurs moléculaires pour déterminer et comparer des génotypes. Ces marqueurs ciblent une partie du génome à étudier et sont utilisés pour mettre en évidence des différences moléculaires entre les individus. Il existe plusieurs types de marqueurs dont certains peuvent détecter des mutations ponctuelles de la séquence d'ADN ou alors des différences du nombre de copies de certains motifs répétés. Les marqueurs dits codominants sont capables de distinguer les individus possédant deux copies du même allèle (homozygotes) alors que les marqueurs dominants ne sont pas capables de différencier les individus hétérozygotes des individus homozygotes pour l'allèle dominants. Trois types de marqueurs ont été utilisés pour cette thèse et sont présentés brièvement ici :

□ Les AFLP (« Amplified Fragment Length Polymorphism ») sont des marqueurs dominants de type présence ou absence d'une séquence d'ADN de longueur donnée ; ils ne ciblent pas une région précise du génome. Cette technique aléatoire permet de travailler sur tout le génome et peut révéler simultanément une centaine de loci de taille moléculaire différente. De part leur universalité et le nombre important de marqueurs générés répartis sur l'ensemble du génome, les AFLP sont des marqueurs de premier choix pour les études de génomique des populations. Cependant, cette technique présente certaines limites. D'une part, du fait de leur dominance, il est difficile d'estimer la fréquence des hétérozygotes et des homozygotes pour l'allèle dominant, d'autre part, ils peuvent être sujets à l'homoplasie (identité des allèles en taille, mais correspondant régions génomiques différentes). De plus, ils nécessitent des tests de répétabilité afin d'estimer le taux d'erreur du phénotypage.

- Les microsatellites (ou SSR « Simple Sequence Repeat ») sont des marqueurs codominants, hypervariables se caractérisant par des motifs de quelques bases répétées plusieurs fois dans le génome. Le nombre de répétitions définit les différents allèles du marqueur. Ils sont supposés neutres et sont facilement reproductibles. De ce fait, ils sont utilisés dans de nombreux domaines de la génétique des populations. Comme les AFLP, les SSRs peuvent présenter des homoplasies de taille
- Les SNP (« Single Nucleotide Polymorphism ») se caractérisent par des variations au niveau des bases qui constituent l'ADN. La majorité du polymorphisme du génome est constitué de SNP. Ces marqueurs codominants permettent les analyses de grande précision.

Encadré 3 : L'équilibre de Hardy-Weinberg

La variabilité génétique d'une population peut être difficile à prévoir au fil des générations. De nombreux facteurs peuvent modifier la fréquence des allèles au sein des populations (i.e. les forces évolutives).

Les fréquences alléliques et génotypiques peuvent être prédites facilement pour un modèle théorique simple d'une population « idéalisée », dont les caractéristiques sont les suivantes :

- Population de taille infinie (pour éviter la dérive génétique) pour des individus diploïdes
- Croisements par reproduction sexuée aléatoires (population panmictique) pour des générations non-chevauchantes
- Absence de migration
- Absence de sélection
- Absence de mutation

Si ces conditions sont respectées, la loi de Hardy-Weinberg (Hardy 1908; Weinberg 1908) permet de prédire exactement les fréquences génotypiques à partir des fréquences alléliques de la population idéale. Les implications du principe de Hardy-Weinberg sont qu'une seule génération de panmixie suffit pour atteindre les fréquences génotypiques de Hardy-Weinberg et que les fréquences alléliques et les fréquences génotypiques (c'est à dire la structure génétique de la population) restent stables de génération en génération. La relation entre les fréquences alléliques et les fréquences génotypiques d'une population à l'équilibre de Hardy-Weinberg est la suivante: Soit un locus ayant deux formes alléliques A et a, avec p : la fréquence dans la population de l'allèle A et q : la fréquence dans la population de l'allèle a (avec $p + q = 1$ car il n'y a que 2 allèles). Alors, dans cette population idéale, les unions aléatoires des individus définiront 3 génotypes possibles ; AA, Aa ou aa, avec p^2 pour la valeur fréquence génotypique de AA, q^2 pour aa et $2pq$ pour Aa (donc avec $p^2 + 2pq + q^2 = 1$, voir Figure).

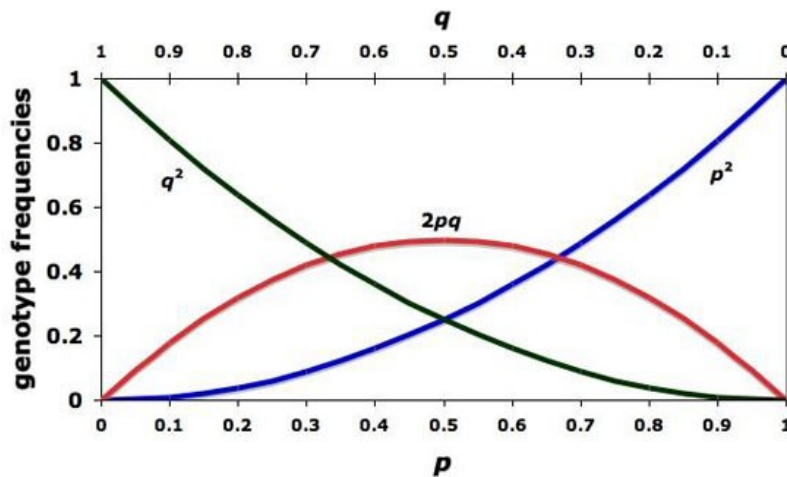


Figure : fréquence génotypique en fonction de la fréquence allélique sous l'équilibre d'Hardy-Weinberg (d'après Andrews 2010)

L'équilibre d'Hardy-Weinberg constitue un outil fondamental en génétique des populations, car une violation de celui-ci pourrait indiquer qu'au moins un des ces cinq processus agissant sur les fréquences alléliques (dérive, migration, sélection, mutation, croisement non aléatoire) opère dans la population.

2. La dérive génétique

La dérive génétique, qui agit sur tout le génome, représente un changement aléatoire des fréquences alléliques au cours des générations dans une population de taille finie (Wright 1931). En effet, dans une population de taille finie, les fréquences alléliques varient sous l'effet du hasard d'une génération à l'autre. Plus la population sera petite, plus les effets de la dérive seront grands car les écarts de fréquences alléliques y seront d'autant plus visibles d'une génération à l'autre. La dérive peut aboutir à l'élimination ou à la fixation d'un allèle pour tous les individus de la population. Les chances de fixation ou d'élimination d'un allèle dépendent évidemment de sa fréquence dans la population. Plus il est fréquent, plus il aura de chances d'être fixé tandis que plus l'allèle est rare, plus il sera sujet à l'élimination (s'il ne confère bien entendu aucun avantage sélectif). La probabilité de fixation, pour un allèle qui vient d'apparaître dans la population par mutation, en absence de sélection, est de $1/2N$ (avec $2N$: nombre total de gamètes pour une population diploïde). Contrairement aux mutations, la dérive génétique aura tendance à réduire la diversité génétique.

3. La migration

En génétique des populations, la migration se réfère au mouvement de gènes entre populations. La migration limite, voir empêche (selon son intensité), la divergence génétique entre populations engendrée par l'effet de la dérive génétique. Le système de dispersion du

pollen et des graines est bien évidemment directement lié à ce processus (Hamrick *et al.* 1993).

Les flux géniques sont le résultat de la contribution génétique d'une population migrante sur la future génération de la population « puits ». Ils ont pour effet l'augmentation de la richesse allélique pour les populations qui reçoivent ces migrants. A long terme, lorsque les flux de gènes sont réciproques d'une population à une autre, les fréquences alléliques entre populations s'homogénéisent (Slatkin 1985) induisant une réduction de la différenciation génétique. L'homogénéisation est d'autant plus forte que les flux de gènes sont importants. Il est utile de noter que les flux de gènes ont généralement lieu au sein d'une même espèce, mais il existe aussi des flux de gènes interspécifiques, qui seront abordés dans une autre partie de cette thèse.

Les marqueurs génétiques offrent une aide précieuse pour l'estimation des flux de gènes (Slatkin 1985; Neigel 1997). Ces estimations reposent sur des modèles évolutifs comme le modèle continent-île qui prend surtout en compte des flux de gènes unidirectionnels (de la population la plus grande à la population la plus petite), comme le modèle en îles (Wright 1951) où toutes les populations sont de même taille et échangent le même nombre de migrants (Encadré 4) ou comme le modèle de « stepping-stone » (Wright 1943) qui prend en compte les distances entre populations.

Encadré 4: Le modèle en île de Wright

Le modèle démographique en île de Wright (1951) est un modèle évolutif simple (sans sélection) avec des populations de tailles identiques échangeant un nombre égal de gènes (Figure ci-dessous).

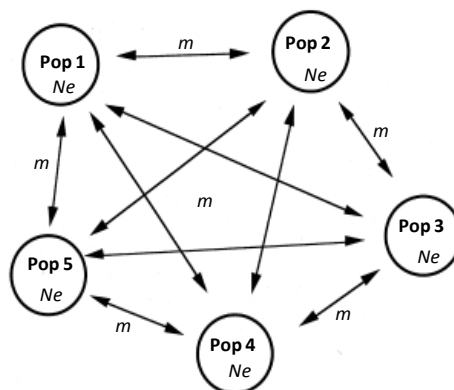


Figure : Représentation du modèle en îles de Wright.

La dérive génétique tend à augmenter la différenciation entre les populations alors que la migration provoque l'effet contraire. Ainsi le niveau de différenciation (F_{ST} , voir Encadré 1) permet de renseigner sur le niveau de

flux de gènes entre populations. La valeur du F_{ST} à l'équilibre migration/dérive sous ce modèle est liée à la migration selon l'équation:

$$F_{ST} = \frac{1}{4N_e m + 1}$$

où N_e est la taille efficace de chaque population et m le taux de migration ($N_e m$ correspond alors au nombre de migrants). Si les populations ne présentent aucun allèle en commun et qu'il n'y a pas de migration entre elles, alors la valeur de F_{ST} sera de 1, ce qui signifie que les populations sont complètement différenciées sous l'effet de la dérive génétique.

Le modèle en île, couramment utilisé pour estimer les taux de migration et néanmoins critiqué (Whitlock & McCauley 1999), est robuste pour les estimations de la migration, même si les populations ne se conforment pas exactement aux hypothèses du modèle (Bohonak 1999).

4. La sélection naturelle

Avancée par Charles Darwin en 1859 dans « L'Origine des espèces », la sélection naturelle est un mécanisme qui contribue à l'évolution des espèces. Elle fait le tri entre les individus montrant un différentiel de survie et de reproduction dans un environnement donné. Elle augmente ainsi la fréquence des phénotypes les plus favorables à la survie et à la reproduction (donc ayant une plus grande fitness) tandis que les phénotypes les moins adaptés tendront à disparaître. Pour que la sélection agisse, il est nécessaire que (i) les individus d'une population ne soient pas tous identiques (variation de traits), (ii) qu'une part de cette variation affecte les traits liés à la capacité de survivre et/ou de se reproduire, et (iii) que les différences de survie/reproduction soient dues à des variations génétiques héritables (par conséquent, ces différences doivent être contrôlées par des gènes transmissibles de génération en génération). Les porteurs des mutations avantageuses ont plus de descendants qui à leur tour auront plus de descendants et ainsi de suite, aboutissant alors à une population dont les individus sont adaptés à leur milieu. Contrairement à la migration et à la dérive génétique qui affectent l'ensemble du génome, la sélection n'agit que sur certains loci, et sur leur voisinage direct (locus liés physiquement).

An niveau intra-populationnel, trois types de sélection naturelle sont définis qui engendrent différentes signatures moléculaires (Figure 1): la sélection purifiante, la sélection positive et la sélection balancée :

- (i) La sélection purifiante (également appelée sélection purificatrice ou négative) élimine les allèles délétères.

- (ii) Inversement, la sélection positive entraîne l'augmentation en fréquence des allèles qui confèrent un avantage sélectif. Une forte sélection positive peut conduire à un balayage sélectif correspondant à une réduction de la diversité autour de l'allèle sélectionné par effet d'entraînement (*hitchhiking*) des variants neutres avoisinants qui sont en déséquilibre de liaison avec l'allèle sélectionné (Maynard Smith & Haigh 1974).
- (iii) La sélection balancée maintient dans une population deux ou plusieurs allèles à un locus donné. Ce maintien de polymorphisme peut être en partie expliqué par l'hétérogénéité spatiale et/ou temporelle de l'environnement (Hedrick 1986; Nielsen 2005) ou encore par un avantage aux hétérozygotes (Takahata *et al.* 1992).

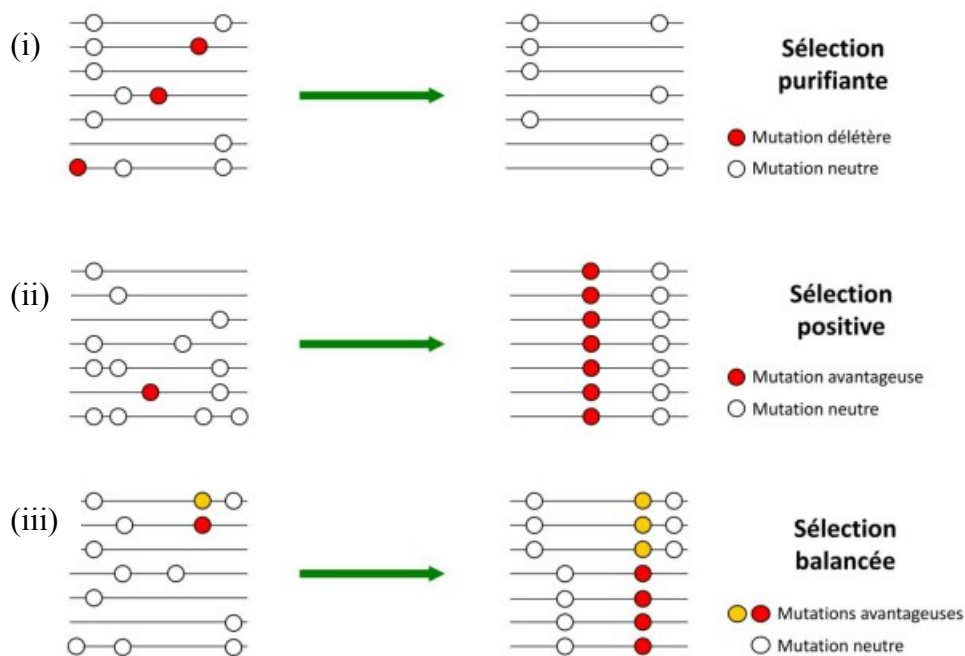


Figure 1: Illustration des différentes formes de sélection au niveau moléculaire au sein d'une population, d'après Audigeos (2010) : (i) la sélection purifiante élimine rapidement les mutations défavorables, (ii) la sélection positive fixe rapidement les mutations favorable et (iii) la sélection balancée maintient plusieurs mutations bénéfiques en fréquence intermédiaire.

Au niveau inter-populationnel, on fait référence à la sélection divergente. Elle agit dans les populations de façon opposée sur des caractères phénotypiques identiques. Il en résulte l'accumulation de fort niveau de différenciation génétique au niveau des loci soumis à sélection divergente (Nosil *et al.* 2009). Comme nous le verrons ultérieurement, elle peut être la résultante de l'adaptation locale.

Les interactions des quatre forces évolutives présentées ici constituent le moteur de l'évolution des êtres vivants. En résumé, les mutations qui sont générées par hasard, sont à

l'origine même de la diversité génétique par la création de nouveaux variants. Au contraire, la dérive qui est également une force stochastique tendra à diminuer cette diversité. Quant à la migration, elle vise à la cohésion des populations, en empêchant la différenciation génétique entre elles. Enfin, la sélection naturelle, en faisant le tri des variants les plus avantageux, est la force qui permet d'expliquer l'adaptation, un fait majeur de l'évolution pouvant aboutir à long terme à la spéciation.

C. L'impact de la démographie sur la diversité génétique

1. La taille efficace des populations

Du fait de la possibilité d'une variation du nombre d'individus au cours des générations, de l'absence de reproduction de certains individus et du sex-ratio déséquilibré, la taille démographique des populations (N) n'est pas un paramètre pertinent pour évaluer l'effet de la dérive génétique sur l'histoire génétique des populations. On utilise la taille efficace (N_e) dont le concept a été élaboré par Wright (1931). La taille efficace correspond au nombre d'individus d'une population idéale de Wright-Fisher (Encadré 5) qui présenterait les mêmes fluctuations de polymorphisme (c'est-à-dire le même degré de dérive génétique) que la population réelle. Le concept de taille efficace englobe la mesure de capacité d'une population à transmettre sa variabilité génétique d'une génération à l'autre.

Encadré 5 : Le modèle de Wright-Fisher

Le modèle de Wright-Fisher (Wright 1929; Fisher 1930) est un modèle classique de la génétique des populations. Il suppose une population panmictique (croisement aléatoire des gamètes), de taille constante ($2N$ allèles pour les individus diploïdes) à l'équilibre démographique, avec des générations non-chevauchantes, se reproduisant de façon sexuée avec possibilité d'autofécondation, sans migration ni sélection avec des mutations neutres apparaissant à taux μ constant.

Il peut être utilisé pour prédire l'évolution du polymorphisme moléculaire sous le modèle neutre de l'évolution à l'équilibre mutation / dérive et peut donc être utilisé pour détecter d'éventuels signes de sélection ou de changements démographiques. Il peut également être utilisé pour la théorie de coalescence pour reconstruire l'histoire généalogique d'un échantillon de gènes.

2. Les changements démographiques

Au cours des générations, la taille efficace des populations naturelles est rarement stationnaire. En effet, une population peut connaître des phases d'augmentation ou de

réduction de son effectif efficace qui ne sont pas sans conséquences sur la diversité génétique de la population.

L'expansion démographique se traduit par une augmentation de la taille efficace (N_e) d'une population. Il en résulte, qu'avec un plus grand nombre d'individus, ce processus démographique provoque une augmentation de la variabilité génétique. En effet, une taille efficace plus grande apporte une probabilité plus grande d'apparition de nouveaux allèles et donc une plus grande richesse allélique. Mais par cet excès d'allèles rares, la diversité génétique observée dans la population en expansion serait alors plus faible que la diversité calculée pour une population en équilibre démographique à partir du nombre d'allèles observée (Cornuet & Luikart 1996). Inversement, un goulot d'étranglement (« bottleneck ») qui signifie une réduction rapide de la taille efficace d'une population, a pour conséquence une diversité génétique de la population supérieure à la diversité génétique sous l'équilibre démographique par l'élimination plus rapide des variants génétiques rares sous l'effet du hasard (du fait des effets importants de la dérive génétique sur les plus petites populations) et provoquera aussi une augmentation de la consanguinité au sein de la population. La théorie de la coalescence (Encadré 6), par la reconstruction de l'histoire généalogique des gènes, offre, entre autre, la possibilité de détecter et même de dater de changements démographiques passés.

Encadré 6 : La théorie de la coalescence

La théorie de la coalescence (Kingman 1982, 2000; Rosenberg & Nordborg 2002; Nordborg 2004) vise à reconstruire, rétrospectivement, l'histoire généalogique d'un échantillon de gènes. Le modèle de Wright-Fisher est couramment utilisé (Encadré 5) pour retracer l'évolution de la diversité génétique du passé vers le présent, à chaque génération, les gènes présents sont donc issus du pool génétique de la génération précédente (tirage avec remise car des parents peuvent avoir plusieurs descendants et d'autres aucun). De génération en génération, des gènes sont donc perdus par dérive génétique (extinction de lignées). Dans l'approche rétrospective du coalescent, on visualise ce processus de dérive génétique en remontant dans le passé jusqu'à la rencontre de l'ancêtre commun des gènes échantillonnés (voir figure). On dit que deux gènes coalescent quand ils se rejoignent sur le gène ancestral commun (MRCA : « Most Recent Common Ancestor »). Les gènes issus d'un gène ancêtre commun sont identiques par descendance. La probabilité que deux lignées coalescent à la génération précédente, pour des organismes diploïdes de taille N_e constante, est de $1/2N_e$ et le nombre attendu de mutations par génération est de $2N_e\mu$.

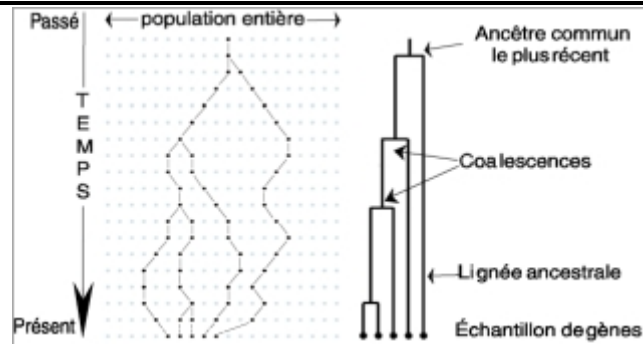


Figure : principe de la théorie de la coalescence

A gauche, visualisation par rétrospection des lignées ancestrales de l'échantillon de gènes considérés au sein de la population actuelle. A droite, l'arbre de coalescence de ce même échantillon

Les analyses de coalescence se décomposent de la manière suivante : une fois les données génétiques collectées, les généalogies possibles sont retracées selon des modèles évolutifs choisis. Puis, pour chaque généalogie, un calcul de vraisemblance des données est réalisé (la vraisemblance totale des données est la somme de la vraisemblance pour chaque arbre, pondérée par la probabilité de chaque généalogie sous ce modèle) qui permettra l'estimation des paramètres du modèle qui maximisent la vraisemblance des données. Ces paramètres peuvent être d'origine démographique (ex : taille efficace, flux de gènes, expansion ou réduction démographique) ou historique (ex : temps de divergence et temps des événements démographiques). Les approches bayésiennes fournissent un outil idéal pour les estimations de ces paramètres (Encadré 7).

Le coalescent utilise ainsi le polymorphisme moléculaires neutre pour connaître l'histoire évolutive des populations. Ainsi, il aide à comprendre les forces évolutives qui sont les plus compatibles avec la généalogie observée.

Encadré 7 : Les méthodes bayésiennes

Les méthodes bayésiennes (Bayes & Price 1763) fournissent des méthodes de calculs efficaces pour différents domaines de la génétique des populations (Beaumont & Rannala 2004). En particulier, elles peuvent être d'un grand recours en coalescent pour les calculs d'estimation de paramètres évolutifs tels que la taille efficace des populations, le temps de divergence entre populations ou les événements de migrations (Drummond *et al.* 2005; Minin *et al.* 2008). Les méthodes bayésiennes estiment la distribution postérieure des paramètres θ_i , à partir des données D et des priors sur les paramètres θ_i . La notion de probabilité postérieure est la probabilité du paramètre θ sachant les données D : $P(\theta | D)$. La probabilité postérieure d'un paramètre peut être calculée par le théorème de Bayes ; elle est fonction de la vraisemblance des données sachant le paramètre ($P(D|\theta)$) et de la probabilité *a priori* $P(\theta)$ de ce paramètre comme le montre l'équation:

$$P(\theta | D) = \frac{P(D | \theta) \times P(\theta)}{P(D)}$$

où $P(D | \theta)$ est la fonction de vraisemblance, $P(\theta)$ la probabilité *a priori* du paramètre θ et $P(D)$ la probabilité des données.

Les méthodes bayésiennes permettent donc d'estimer les paramètres du modèle qui maximisent la probabilité d'observer les données. Il existe plusieurs approches pour calculer la probabilité postérieure des paramètres à estimer. Deux sont brièvement présentées ici :

□ Les méthodes bayésiennes par MCMC (Markov Chain Monte Carlo), permettent d'explorer toutes les valeurs de probabilité postérieures des paramètres associées aux données. La convergence des chaînes de Markov pourra permettre une bonne estimation des paramètres. Cette approche est notamment utilisée, par le modèle d'Isolation with Migration' (IMa Hey 2006).

□ L'ABC ("Approximate Bayesian Computation" : Bertorelle *et al.* 2010; Csilléry *et al.* 2010 pour des reviews) est une approche par simulations, particulièrement utile quand il s'agit de modèles évolutifs complexes avec de nombreux paramètres, pour lesquels le bayésien classique par MCMC peut présenter des problèmes de convergence des chaînes et donc d'estimation des paramètres. Un jeu de données est simulé sous le modèle de coalescence (avec par exemple SIMCOAL2 (Excoffier *et al.* 2000)) pour différentes valeurs d'un paramètre θ que l'on cherche à estimer. Un résumé des données génétiques simulées est ensuite obtenu en utilisant des statistiques résumées (la diversité génétique H_e ou la différenciation génétique F_{ST} , par exemple). On calcule ensuite l'écart dij entre les statistiques résumées issues de simulation et les statistiques résumées observées en utilisant par exemple le modèle de régression comme implémentée dans ABCest2 (Beaumont *et al.* 2002; Excoffier *et al.* 2005). La distribution des écarts entre les statistiques résumées simulées et observées permet d'estimer la probabilité à posteriori du paramètre θ sachant que l'écart entre les données simulées et les données observées est inférieur à un seuil de tolérance que l'on se fixe : $P(\theta | D_{sim} - D_{obs} < \epsilon)$.

Cette méthode, qui n'est donc pas véritablement une méthode purement bayésienne, permet l'estimation de différents paramètres de manière indépendante comme par exemple l'estimation des paramètres démographiques suite à des scénarios de changements de tailles démographiques.

Les changements démographiques, au sein d'une population, agissent sur la taille efficace N_e et modifient ainsi la diversité génétique intra-populationnelle.

Cependant, il est important de noter qu'une modification de l'effectif d'une population d'arbres provoque également des changements physiques directs du milieu et en conséquence modifie la dynamique d'autres populations. En particulier, l'exploitation forestière réduit les densités de populations et cette réduction, par coupe forestière, va modifier l'environnement physique par la formation de trouées perturbant l'environnement local. Ainsi, l'exploitation forestière, au-delà de l'être perçue comme un bottleneck pour les populations concernées, peut aussi modifier les patrons génétiques des populations d'autres espèces occupant le même environnement qui se retrouve ainsi modifié, et peut donc agir sur leurs histoires évolutives (Encadré 8).

Encadré 8 : Les conséquences de l'exploitation forestière

L'exploitation forestière provoque des modifications directes de l'environnement avec les coupes de certains arbres, mais aussi des modifications indirectes pour les populations comme la perturbation de la régénération (e.g. Fredericksen & Mostacedo 2000). L'étude de la diversité génétique est alors ici d'une importance capitale quant au devenir des populations exploitées.

En forêt tropicale, qui est fortement exploitée pour la qualité de certaines essences (Bawa & Seidler 1998), il existe un nombre important d'espèces subissant les effets de l'exploitation (de manière directe ou indirecte) pour lesquelles aucune étude d'impact n'a été réalisée, notamment sur leur diversité génétique et sur les conditions biotiques et abiotiques nécessaires à leur bonne régénération. Ces connaissances paraissent pourtant primordiales afin de maintenir les conditions nécessaires au maintien de la diversité pour les futures générations.

La déforestation réduit les densités et donc, tend à augmenter les distances séparant les individus et les populations ; ce qui peut causer une perte d'allèles et affecter sérieusement la diversité génétique, la structure spatiale et la structure génétique spatiale (Young *et al.* 1996; Aldrich & Hamrick 1998; White *et al.* 1999; Dick 2001), la fitness et le succès reproductif de la plante par les activités des pollinisateurs (Aizen & Feinsinger 1994), le dépôt de pollen et de graines (Quesada *et al.* 2003). Cependant, il a déjà été montré que ces activités humaines sur la forêt peuvent avoir des effets positifs (White *et al.* 2002; Dick *et al.* 2003) ou neutres (Lee 2000; Cascante *et al.* 2002; Fuchs *et al.* 2003) sur plusieurs traits liés à la biologie de la reproduction et de la régénération.

III. La diversité génétique comme moteur de l'adaptation et de la spéciation

Il est communément admis qu'une population possédant globalement une grande variabilité génétique aura une grande capacité à faire face aux variations de l'environnement dans l'espace et dans le temps même si, d'après la théorie neutre de l'évolution, ayant été avancée par Kimura en 1968, seule une minorité de loci s'avère adaptative. Pour chaque environnement donné, le phénotype possédant une plus grande fitness tendra à être sélectionné comme vu précédemment. Ainsi, l'adaptation au milieu, via la sélection naturelle, permet le maintien d'une diversité phénotypique, et donc génétique (au moins pour certains loci), et cette diversité pourrait s'avérer essentielle pour la survie des populations en cas de perturbation du milieu.

A. La détection de la sélection naturelle

1. La théorie neutre de l'évolution

La théorie neutre de l'évolution (Kimura 1968) suggère que la majorité des mutations n'ont pas d'effet sur la fitness des organismes : on dit que ces mutations sont sélectivement neutres car elles ne confèrent ni avantage (mutations favorables) ni désavantage (mutations délétères) aux organismes. En effet, au lieu d'être le produit de différents types de sélection, la plupart du polymorphisme n'a pas d'effet sur la fitness des individus et est donc influencée par le changement aléatoire des fréquences alléliques par dérive génétique. Cette théorie accepte qu'une mutation puisse avoir un effet sur la fitness, mais elle suggère d'une part que ces mutations seraient tellement délétères, qu'elles seraient très rapidement éliminées de la population par la sélection négative ; d'autre part elle affirme que très peu de mutations sont avantageuses et que celles-ci tendront alors à se fixer rapidement. Ces mutations non-neutres n'auraient alors qu'un effet mineur sur le niveau de polymorphisme et de divergence par rapport à l'importance de la dérive génétique.

2. Les tests de neutralité

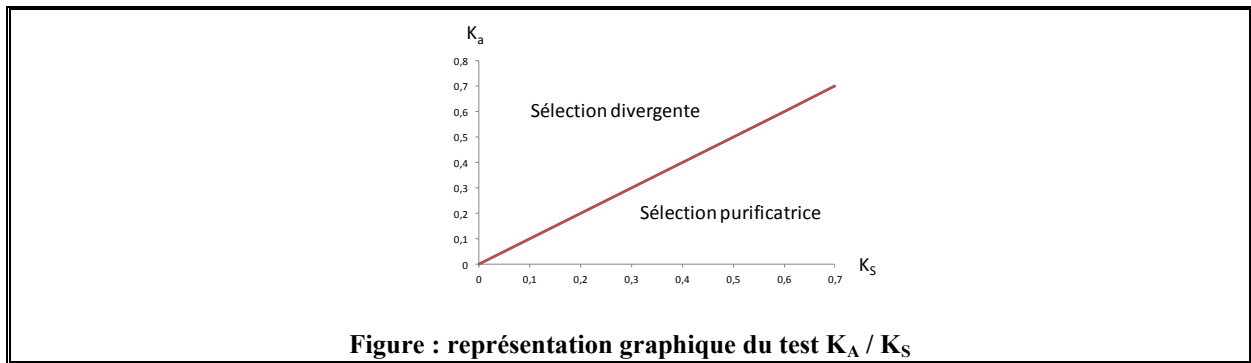
a) généralités

Il est difficile de mettre en évidence l'effet de la sélection naturelle, et donc d'une possible adaptation, de manière directe. Il faudrait pour cela suivre expérimentalement la fitness d'individus dans des conditions environnementales contrôlées sur plusieurs générations. Une manière plus rapide mais indirecte consiste à détecter les signatures moléculaires engendrées par l'action de la sélection naturelle par le biais de comparaisons avec une population « idéale » dérivant du modèle neutre standard de Wright-Fisher sous l'équilibre mutation/dérive (Encadré 5). Par exemple, il est possible de comparer les estimateurs de diversité calculés à partir de notre échantillon aux estimateurs de diversité sous l'hypothèse nulle ($\theta = 4N_e\mu$). L'écart entre les deux valeurs sont analysés analytiquement ou par simulation, et peuvent ainsi renseigner de la présence ou non de l'action sélection naturelle. Plus précisément, l'écart à ce modèle peut être interprété soit par une signature de la sélection naturelle soit par un écart à l'équilibre démographique (du fait de la taille constante N_e du modèle neutre).

Il existe différents tests pour identifier les signatures de la sélection naturelle; ils se différencient en fonction de leur principe de base, de leur cadre d'application ou bien du nombre de loci qu'il est nécessaire d'étudier. Certains tests sont basés sur les fréquences alléliques tel que le test de Ewens (1972) – Watterson (Watterson 1974), d'autres sur la diversité nucléotidique (Tajima 1983; Fu & Li 1993) ou haplotypique (Hudson *et al.* 1994; Fu 1996). Il existe aussi des tests basés sur l'hétérogénéité entre plusieurs loci comme le test HKA (Hudson *et al.* 1987), le test du K_A / K_S (Encadré 9, McDonald & Kreitman 1991) ou la méthode de criblage génomique qui sera plus détaillée dans la section suivante.

Encadré 9 : Le test du K_A / K_S

Le test du K_A / K_S est un test d'indication de pression sélective agissant sur une région codante du génome. Ce test compare le rapport du taux de substitutions non-synonymes (K_A) et du taux substitutions synonymes (K_S) par site polymorphe entre populations ou espèces proches pour la séquence étudiée (Kimura 1980; Li *et al.* 1985; McDonald & Kreitman 1991; Yang & Nielsen 2000). A l'équilibre (en l'absence de sélection), K_A est égal à K_S . Par contre, Si K_A est supérieur à K_S , la sélection est divergente. Au contraire un K_A inférieur à K_S , renseigne d'une sélection purificatrice (figure).



b) *L'approche de criblage génomique (« genome scan »)*

La génomique des populations (Luikart *et al.* 2003; Scotti-Saintagne *et al.* 2004; Bonin 2008) est une discipline récente et populaire qui consiste à analyser des données génétiques issues de criblage génomique (« genome scan ») de différentes populations pour identifier des régions génomiques sous sélection. De part leur grand nombre de loci ciblés sur tout le génome, les marqueurs anonymes AFLP constituent un outil idéal pour une telle approche notamment pour les espèces non modèles, pour lesquelles aucune informations génomique n'est disponible (Mueller & Wolfenbarger 1999; Bensch & Akesson 2005). Les régions sous sélection (également appelées « outliers ») sont détectés par le fait qu'elles présentent des niveaux de différenciation génétique différents des loci neutres. En effet, ils tendront à avoir un niveau de différenciation génétique plus élevé qu'attendu lorsqu'ils sont soumis à une sélection divergente et plus faible lorsqu'ils sont soumis à une sélection balancée (Lewontin & Krakauer 1973; Luikart *et al.* 2003; Beaumont 2005; Storz 2005).

Plusieurs approches sont possibles pour détecter les loci soumis à sélection (e.g. Beaumont & Nichols 1996; Vitalis *et al.* 2001; Beaumont & Rannala 2004; Foll & Gaggiotti 2008). Dans cette thèse deux méthodes ont été utilisées. La première méthode (Beaumont & Nichols 1996) utilise le modèle de coalescence (voir Encadré 6) pour simuler des loci ayant évolué sous le modèle neutre et identifier les outliers en comparant les valeurs de différenciation (F_{ST}) de chaque locus avec les valeurs simulées de F_{ST} . La deuxième approche (Foll & Gaggiotti 2008) a l'avantage de ne pas dépendre de la distribution de loci neutres simulés puisqu'elle estime directement, en utilisant une approche bayésienne, la probabilité postérieure pour un locus donné d'être soumis à sélection. Pour chacun des deux modèles, la probabilité d'un locus d'être un outlier est estimée en testant son écart par apport au modèle neutre. Chaque approche ayant sa spécificité, il peut être utile combiner ces méthodes (Oetjen & Reusch 2007; Pérez-Figueroa *et al.* 2010; Narum & Hess 2011; Tice & Carlon 2011), en particulier pour limiter la détection de faux positifs (détection de « faux » outliers : « Erreur de type I»), ou de faux négatifs (absence de détection de « vrais » outliers : « Erreur de type II»).

La génomique des populations peut permettre la détection des facteurs environnementaux responsables de la structure génétique observée. En effet, si les populations diffèrent par un facteur environnemental précis et qu'une sélection divergente est détectée (par des outliers à fort niveau de divergence entre populations) alors ce facteur pourra être identifié comme responsable de la sélection. Cette approche qui vise à détecter la corrélation entre les patrons génétiques et les facteurs environnementaux se nomme la génomique du paysage (« landscape genomics », Manel *et al.* 2003; Joost *et al.* 2007; Manel *et al.* 2010). Par la technique de genome scan, chez différents organismes il a déjà été montré que des pressions de sélection sur certaines régions du génome peuvent être exercées par différents facteurs environnementaux : l'altitude (Bonin *et al.* 2006; Fischer *et al.* 2011), la pollution chimique (Williams & Oleksiak 2008; Paris *et al.* 2010), la salinité (Chen & Yang 2009; Coyer *et al.* 2011), la radioactivité (Kuchma & Finkeldey 2011), le stress hydrique (Altinkut *et al.* 2003), les zones de balancement des marées (Oetjen & Reusch 2007). Le pourcentage d'outliers détecté est très variable selon les études menées (entre 0.5 et 20%: Jump *et al.* 2006; Miller *et al.* 2007; Williams & Oleksiak 2008; Chen & Yang 2009; Kuchma & Finkeldey 2011), car il dépend avant tout de la méthode utilisée. En effet, l'approche bayésienne (Foll & Gaggiotti 2008) se révèle plus stringente vis-à-vis de cette détection en limitant la détection de faux positifs (Narum & Hess 2011; Tice & Carlon 2011).

3. Les interférences avec la démographie

L'utilisation des tests de neutralité sélective peut produire des résultats ambigus, puisque les écarts à l'attendu peuvent également refléter l'histoire démographique de la population considérée plutôt que l'effet de la sélection naturelle (Nielsen 2005). En effet, parmi les postulats du modèle neutre, la taille de la population (N_e) est constante. Si les populations étudiées ne respectent pas ces conditions alors aucune conclusion ne peut être tirée de l'écart à la valeur de la diversité attendue à l'équilibre mutation-dérive. En effet, les facteurs démographiques comme la sélection naturelle pourraient expliquer cet écart. Par exemple une contraction de population (« bottleneck »), peut avoir les mêmes effets que la sélection balancée représentant une augmentation de la proportion des allèles à fréquence intermédiaire (Bryja *et al.* 2007), ainsi qu'une expansion démographique qui peut imiter la sélection positive ou négative par une augmentation de la proportion des allèles rares (Hahn *et al.* 2002; Mes 2004). Néanmoins, à la différence de la sélection naturelle qui cible son action sur quelques loci, les transitions démographiques laissent une empreinte homogène sur l'ensemble du génome, Il devient alors possible de distinguer ces deux processus en

comparant le profil de la variabilité génétique de la région candidate à la sélection au profil global des autres régions du génome (Bamshad & Wooding 2003; Luikart *et al.* 2003). Le profil génétique de ces régions « neutres » servira à construire le modèle nul d'absence de sélection en présence de transition démographique. L'écart à ce modèle reflétera alors seulement l'effet de la sélection naturelle. L'utilisation de la génomique des populations peut alors s'avérer un outil idéal pour s'affranchir de l'inférence démographique lors de la recherche de signature de sélection.

B. L'adaptation locale

L'adaptation locale mesure la concordance entre l'hétérogénéité spatiale de l'environnement et la variabilité génétique adaptative des populations reliées par des flux de gènes (Kawecki & Ebert 2004). C'est un mécanisme important qui va permettre à une population de maintenir plusieurs phénotypes dans un environnement contrasté (Hedrick 1986). Ainsi une sélection divergente qui apparaît à l'échelle de la population globale est la résultante des adaptations locales au milieu qui peuvent se faire à des échelles spatiales réduites (microenvironnements). Il est parfois difficile de distinguer clairement deux habitats relativement proches d'un point de vue géographique. Souvent, on parle de gradients environnementaux pour lesquels des changements de caractéristiques se font continuellement. L'adaptation locale se traduit alors généralement par un gradient de fréquences alléliques et phénotypiques le long d'un gradient environnemental (Mullen & Hoekstra 2008). Comme nous l'avons vu précédemment, il peut être difficile de distinguer les événements sélectifs des événements démographiques car ils peuvent laisser la même empreinte moléculaire. L'étude de l'adaptation locale, en travaillant sur une seule population, a l'avantage de s'affranchir des problèmes liés à la possibilité de confondre les différenciations génétiques causées par les événements démographiques différents avec le résultat d'une sélection divergente. En effet, les sous-populations ont une histoire commune.

L'adaptation locale résulte d'un équilibre entre la sélection, qui tend à différencier des populations vivant dans des milieux différents, et le flux de gènes, qui contribue au contraire à homogénéiser ces populations (Sambatti & Rice 2006; Savolainen *et al.* 2007). Des pressions de sélection divergente importantes permettront une plus grande différenciation entre sous-populations donc un niveau d'adaptation locale plus important. Alors que des flux de gènes importants engendreront une homogénéisation rapide au niveau de toute la population et la détection de l'adaptation locale tendra alors à disparaître. A l'inverse, certaines études

suggèrent qu'un taux de migration moyen à élevé, en augmentant la variance génétique dans la population, permet une sélection plus efficace et donc un niveau d'adaptation locale plus élevé (Gandon & Michalakis 2002). Les flux de gènes peuvent ainsi affecter le potentiel adaptatif des populations. En effet, l'apport de nouveaux allèles avantageux pourront permettre aux populations de mieux répondre aux pressions de l'environnement (Lenormand 2002). Dans un environnement à la fois variable dans le temps et dans l'espace, d'autres facteurs peuvent influencer sur l'adaptation locale. En effet, plus le temps de génération est court (ce qui n'est pas le cas des arbres), plus les populations s'adapteront rapidement (par la sélection des individus sur la base de leur fitness au fil des générations plus courtes, Savolainen *et al.* 2004). Si les études phénotypiques de l'adaptation locale sont assez nombreuses chez les arbres forestiers en particulier par les approches de génétique quantitative dans des dispositifs expérimentaux de jardins communs (« common garden ») ou bien de transplantations réciproques (Mátyás 1996; Cordell *et al.* 1998; Kawecki & Ebert 2004; Baltzer *et al.* 2005; Williams *et al.* 2008), les bases génétiques restent incertaines, notamment en milieu naturel (Savolainen *et al.* 2007). Les nouveaux outils moléculaires, tels que les approches génomiques, permettront sûrement de pallier ce manque d'informations.

C. De l'adaptation vers la spéciation

La spéciation est le processus par lequel de nouvelles espèces sont formées. Il existe de nombreuses définitions pour décrire la notion d'espèce (Mayr 1942; Hey 2001; Rieseberg *et al.* 2006) et débattre sur ce sujet sort du cadre de ce travail. Selon Darwin dans l'Origine des espèces (1859), la spéciation est un événement progressif et continu dont le moteur est la sélection naturelle et qui peut être la conséquence de l'adaptation à de nouvelles niches écologiques. Cette définition peut être complétée par la notion d'isolement reproducteur entre les espèces, par le biais des barrières géniques (Dobzhansky 1937) qui maintiennent l'intégrité des espèces. De plus, il est important de noter que la dérive génétique peut favoriser le processus l'apparition de nouvelles espèces notamment par l'effet fondateur (Templeton 1980; Barton & Mallet 1996; Turelli *et al.* 2001) qui se traduit par la migration d'un petit nombre d'individus fondant une nouvelle population. Ce petit nombre d'individus n'emporte avec lui qu'une partie du pool génétique de la population mère, et l'importance des effets de la dérive sur cette nouvelle petite population peut permettre une divergence génétique progressive entre les deux populations pouvant aboutir à la spéciation.

1. Les modes de spéciation

Traditionnellement, les modes de spéciation font référence à la répartition géographique des populations lors de l'isolement reproducteur (Figure 2):

- la spéciation allopatrique caractérise une divergence de populations en réponse à un isolement géographique qui empêche les flux de gènes. Ce scénario semble être le plus courant pour l'origine des espèces (Coyne & Orr 2004). En l'absence de flux de gènes due aux barrières géographiques, la pression de sélection ne doit pas nécessairement être élevée pour engendrer cette spéciation. Un effet fondateur peut être à l'origine de cette spéciation.
- la spéciation sympatrique décrit une divergence de populations mais en absence de barrière géographique. Elle est souvent la conséquence d'une sélection divergente (ou disruptive) qui agit dans la population de façon opposée sur des mêmes caractères (Dieckmann & Doebeli 1999; Via 2001). Ici, par l'absence de barrière géographique, des flux de gènes sont possibles entre les deux populations. Mais alors, ces flux devront être contrebalancés par une sélection très importante pour aboutir à la spéciation. Bien que considérée comme rare, elle a été décrite chez certains organismes comme chez les poissons (Schliewen *et al.* 2001; Barluenga *et al.* 2006) ou chez les oiseaux (Coyne & Price 2000) comme chez le fameux pinson de Darwin (Grant & Grant 1979).
- la spéciation parapatrique est un intermédiaire entre les deux précédents modes de spéciation. La séparation géographique est incomplète, deux populations partagent une même frontière perméable aux flux de gènes.

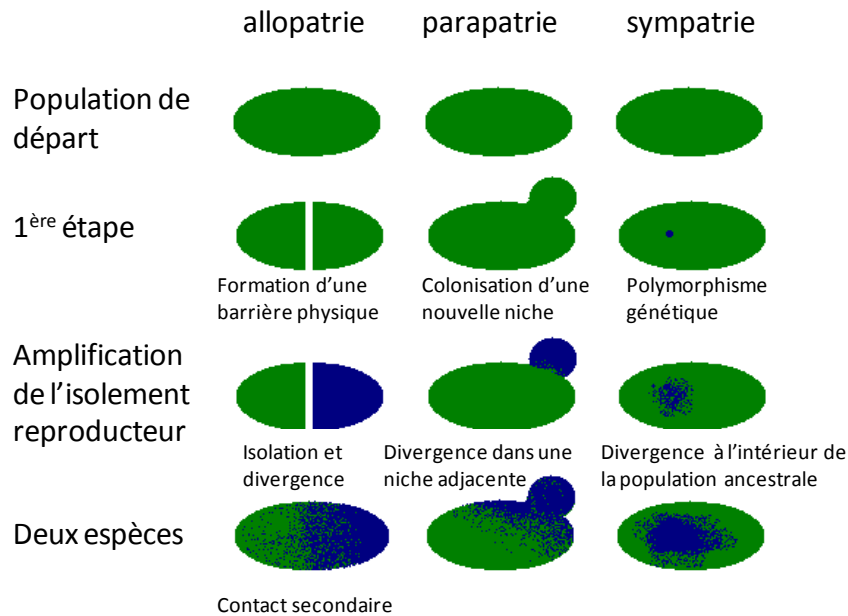


Figure 2 : les différents types de spéciation : la spéciation allopatric, parapatric et sympatric

La classification des modes de spéciation selon la répartition géographique des populations en cours de spéciation met l'accent sur la présence ou pas des flux de gènes dans le processus de spéciation. Une seconde façon de classer les modes de spéciation est de considérer les mécanismes à l'origine de l'évolution de l'isolement reproducteur. Par exemple, la spéciation sexuelle repose sur l'incompatibilité des organes reproducteurs ou la présence d'un décalage dans la saison de reproduction (isolement prézygotique). Il existe aussi la spéciation par réarrangement chromosomique qui engendre une fécondité réduite des hétérozygotes (donc des hybrides) due à des problèmes de ségrégation des chromosomes à la méiose (isolement postzygotique). Dans cette thèse, un intérêt particulier sera porté à la spéciation par sélection naturelle autrement appelée, spéciation écologique.

2. La spéciation écologique

La spéciation écologique est le résultat de la sélection divergente qui agit de façon opposée sur des traits entre des environnements contrastés. Elle mène directement ou indirectement à un isolement reproducteur, alors appelé, isolement par adaptation (Schluter 2001; Via 2001; Wu 2001; Rundle & Nosil 2005; Nosil *et al.* 2008; Via 2009) et peut être observée quelque soit le degré d'isolement géographique entre les populations. L'adaptation locale peut donc aboutir à un processus de spéciation (Kawecki & Ebert 2004). Les variations de l'habitat dans l'espace et/ou dans le temps sont à l'origine de la sélection naturelle divergente qui participera à l'isolement reproducteur (pré- ou postzygotique) et à la diminution ou l'arrêt des flux de gènes entre les populations. Cependant, il a déjà été

remarqué que les flux de gènes ont une grande importance en facilitant l'adaptation pour un nouvel environnement par l'apport de nouveaux allèles avantageux (Morjan & Rieseberg 2004; Seehausen 2004; Savolainen *et al.* 2006; Gaggiotti 2011). Même si elle a déjà été démontrée chez quelques espèces de poissons, de mollusques, d'oiseaux ou d'insectes (Rundle *et al.* 2000; Quesada *et al.* 2007; Nosil *et al.* 2008; Schluter 2009), elle reste relativement rarement observée dans la nature, et rarement documentée chez les plantes (Rieseberg & Willis 2007), particulièrement chez les arbres forestiers (Lexer & Fay 2005; Roda 2011) sauf chez quelques espèces modèles comme les chênes (Cavender-Bares & Pahlisch 2009) ou les peupliers (Lexer *et al.* 2005).

3. L'hybridation et les complexes d'espèces

La reproduction entre deux individus d'espèces différentes (hybridation), n'est pas un événement rare chez les végétaux (Anderson & Stebbins 1954; Stace 1975; Mallet 2005; Lepais *et al.* 2009), notamment chez les arbres tropicaux (Leigh *et al.* 2004). Elle facilite l'apparition et l'établissement de nouvelles espèces qui peut se faire par changement de ploïdie (alloploïdie) comme chez le blé (e.g. Gaut *et al.* 2000) ou non (homoploïdie) comme étudié chez le tournesol (e.g. Baack *et al.* 2005). L'hybridation peut également faciliter la conquête de nouveaux milieux grâce aux phénotypes extrêmes des hybrides (Rieseberg *et al.* 1999) et pourrait alors favoriser la spéciation écologique (Barton 2001; Levin 2004; Seehausen 2004; Rieseberg & Willis 2007; Ma *et al.* 2010). Au contraire, dans certains cas, il est très difficile de démontrer des signes d'hybridation sur la seule base de différences morphologiques avec les espèces parentale, on peut alors avoir recours à l'utilisation de marqueurs moléculaires (Palmé 2002; Petit 2004). Par exemple, les méthodes d'assignation ou de regroupement (« clustering ») permettent, comme l'indique leur nom, d'assigner ou de regrouper des individus dont l'origine peut être inconnue (identité spécifique ou localisation spatiale) à des groupes prédéfinis (méthode de classification) ou nouvellement construits (méthode de clustering) sur la base de leur génotype (Pritchard *et al.* 2000; Anderson & Thompson 2002; Corander *et al.* 2003; Gompert & Alex Buerkle 2010). Ces approches bayésiennes visent à estimer le nombre le plus probable de groupes sous l'équilibre d'Hardy-Weinberg (Encadré 3) sachant les données et à les délimiter pour ensuite calculer la probabilité d'appartenance de chaque individu à chaque groupe. Ainsi, ce genre d'approche permet de caractériser les individus qui sont proches génétiquement. Par exemple, il sera alors possible de détecter des individus intermédiaires aux groupes génétiques c'est-à-dire des individus issus d'une hybridation si l'étude est réalisée au niveau interspécifique.

L'introggression se définit comme le transfert de certains gènes d'une espèce à l'autre par hybridation (Anderson & Hubricht 1938). Le transfert de certains gènes, plus que d'autres, dépend de la compatibilité des allèles entre espèces. En effet, certaines associations rendront des hybrides stériles ou létaux (Brideau *et al.* 2006). Les niveaux de différenciation génétique entre certaines espèces proches peuvent alors être différents selon les régions et le type de génome. Certaines régions génomiques, potentiellement liées à la divergence adaptative entre espèces, peuvent montrer un fort niveau de différenciation alors que d'autres régions montrent un faible niveau de différenciation qui représente probablement des zones perméables aux flux de gènes interspécifiques (Rieseberg & Burke 2001; Wu 2001). Cependant malgré les flux de gènes interspécifiques, l'intégrité des espèces semble souvent maintenue (Petit & Hampe 2006). La spéciation par hybridation (l'hybride donnant une espèce nouvelle, Avise 2004) pourrait être à l'origine des fortes diversités spécifiques présentes dans certains milieux comme les forêts tropicales.

Les flux de gènes interspécifiques entre espèces proches, aussi appelées espèces-sœurs, caractérisent un complexe d'espèces. Ce concept fut proposé par Pernès (1984); il définit le complexe d'espèces comme un ensemble de taxons qui échangent des gènes directement ou indirectement par le relais d'autres espèces. Les complexes d'espèces sont fréquents chez les plantes dans la nature (Rieseberg & Carney 1998; Leigh *et al.* 2004; Mallet 2005; Mallet 2008; Lepais *et al.* 2009) mais d'avantages d'études sont nécessaires pour estimer leur fréquence. L'isolement reproducteur est une notion-clé de la définition biologique des espèces. Néanmoins, les cas d'hybridations dans la littérature sont nombreux et le débat sur les processus de spéciation et la notion d'espèce reste encore d'actualité.

IV. Problématique de thèse

A. Modèles d'étude

Les espèces d'arbres forestiers tropicaux constituent la composante dominante, en termes de biomasse et de capacité d'influencer la structure de l'écosystème, de l'un des biomes plus diversifiés de la planète, la forêt tropicale humide. Leur adaptation au milieu et le potentiel de faire face aux changements (surtout d'origine climatique) du milieu déterminent donc à la fois la structure des communautés végétales tropicales et le potentiel d'adaptation de ces communautés aux changements globaux.

Les Myristicacées forment une ancienne famille botanique très répandue en Amérique du Sud et centrale, très diversifiée en particulier en Amazonie (Holbrook *et al.* 2006; Queenborough *et al.* 2007) et dont les espèces se différencient pour leurs préférences de milieu et par rapport à la contrainte hydrique (Baraloto *et al.* 2007). Cette famille comprend une vingtaine de genres représentant environ 500 espèces d'arbres, de buissons ou de lianes se localisant dans des niches écologiques variées (Janovec & Neill 2002; Queenborough *et al.* 2007). Malgré l'âge avancé de cette famille (début Eocène : $55,8 \pm 0,2$ à $48,6 \pm 0,2$ millions d'années, Doyle *et al.* 2008) et cette forte diversité spécifique, écologique, mais aussi morphologique, de faibles niveaux de divergence moléculaire ont été observés, ce qui complique la résolution phylogénétique, toujours incomplète, au sein de cette famille (Sauquet *et al.* 2003). Ses membres constituent alors un outil de premier choix pour étudier les mécanismes évolutifs sous-jacents.

Le genre néotropical *Virola* (nom vernaculaire : yayamadou) est constitué d'une quarantaine d'espèces (Wilson 2004) dont cinq sont présentes en Guyane française : *Virola michelii* Heckel, *V. surinamensis* Warburg, *V. kwatae* Sabatier, *V. sebifera* Aublet and *V. multicosata* Ducke. Ces espèces présentent des différences morphologiques distinctes et des différences de distribution tant au niveau géographique qu'au niveau de l'environnement local (Sabatier 1997). Par exemple, le yayamadou montagne (*V. michelii*) se situe préférentiellement sur les plateaux, ainsi que *V. kwatae* (Sabatier 1997) alors que le yayamadou marécage (*V. surinamensis*) affectionne les bas-fonds et bas de pente inondables (Baraloto *et al.* 2007; Macía 2011), tout comme l'indique leur noms vernaculaires. Toutes les deux présentent une

large aire de distribution à l'échelle continentale et sont très présentes dans les forêts guyanaises tandis que *V.kwatae* ne se situe que dans une petite partie du bouclier guyanais (Sabatier 1997). Certaines espèces comme *V. surinamensis* et *V. michelii* sont fortement exploitées pour la confection de matériaux de construction.

A l'intérieur de ce genre, plusieurs espèces de *Virola* ont été étudiées de par leur capacité de dispersion de graines (zoochorie par les toucans et les grands singes), en particuliers chez *V. michelii* (Forget *et al.* 2000), *V. kwatae* (Forget & Cuijpers 2008), *V. sebifera* (Howe 1981), *V. calophyla* (Russo & Augspurger 2004) et *V. flexuosa* (Holbrook & Loiselle 2009), ou pour les propriétés de leur molécules chimiques (e.g. Kawanishi & Hashimoto 1987) notamment pour les effets hallucinogènes recherchés par les chamanes amérindiens (Miller 1999). Cependant, seulement un très faible nombre d'études génétiques ont été portées chez les espèces de *Virola*. En particulier, celles-ci étaient des études focalisées sur la diversité et la structure génétique de populations de l'espèce *V. michelii* avec l'étude conjointe d'autres espèces d'arbres tropicaux (Caron 2000; Degen *et al.* 2001). Celles-ci ont révélé une bonne diversité génétique et une structure génétique spatiale significative à fine échelle.

B. Sites d'étude

1. La Guyane française

La Guyane française est le plus grand département français, recouvert à plus de 90% de forêt, et se situe en Amérique du Sud sur le plateau des Guyanes. Ce territoire comprend des différences environnementales et climatiques selon les régions et selon les saisons. La Guyane subit un climat équatorial humide : les températures sont stables toute l'année avec une moyenne de 26°C, avec de très faibles variations selon les saisons, et une humidité atmosphérique très importante (de 80 à 90%). Les déplacements saisonniers de la zone intertropicale de convergence (ZIC) déterminent l'existence de quatre saisons : une petite saison des pluies (novembre-janvier), une petite saison sèche (février-mars), une grande saison des pluies (avril-juillet) et une grande saison sèche (août-octobre). Les limites entre les saisons peuvent varier de semaines selon les années, certaines saisons peuvent même s'effacer. Les différences de précipitations entre saisons sèches et saisons de pluies peuvent être très importantes.

Les précipitations annuelles sont de l'ordre de 3000 mm par an mais avec de grandes disparités selon les régions de la Guyane (Figure 3). En effet, un fort gradient décroissant de

pluviométrie vers l'ouest et vers le sud est observé. Ainsi, la région de Kaw-Cacao-Régina reçoit en moyenne plus de 4000 mm de pluie par an alors que Saint-Laurent du Maroni a à peine 2000 mm par an.

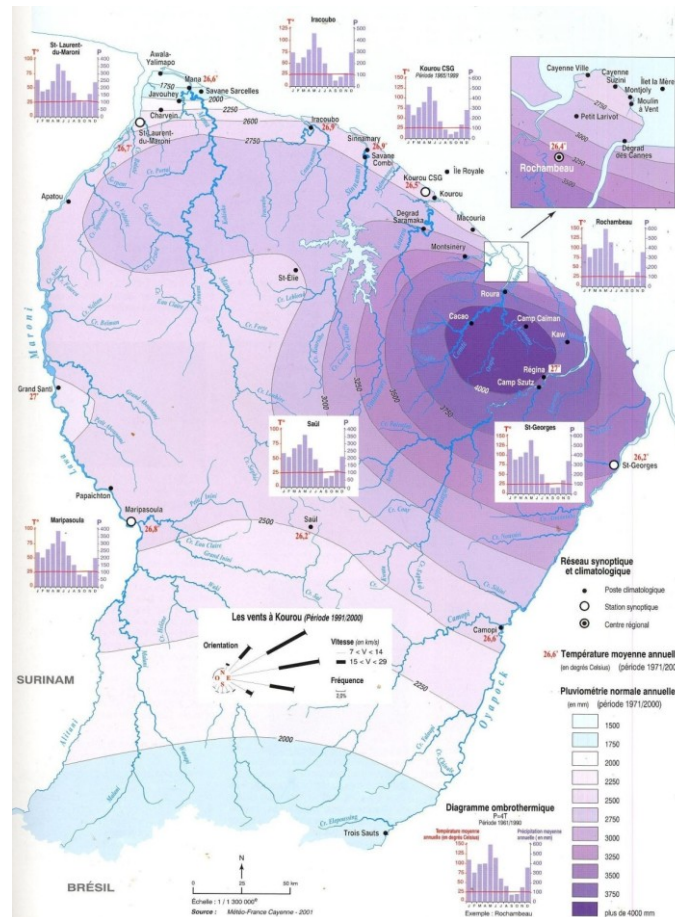


Figure 3 : carte des pluviométries en Guyane française (ATLAS de la Guyane, planche 14)

Il existe alors, en Guyane, de grandes différences quant aux précipitations, à la fois dans le temps et dans l'espace. Cette dernière composante est potentiellement intéressante pour étudier l'adaptation au niveau du stress hydrique à l'échelle régionale (Rambal & Leterme 1987). Cependant, l'adaptation locale à un environnement hétérogène à fine échelle paraît également une force importante pouvant intervenir dans les processus de divergence de population via la sélection adaptative (Lechowicz & Bell 1991; Ellis & Weis 2006; Parisod & Christin 2008).

2. Le dispositif expérimental de Paracou

Le dispositif de recherche sylvicole de Paracou (5°18N, 52°55W ; <http://www.ecofog.gf/spip.php?article174>) est installé depuis 1984 par le Centre Technique

Forestier Tropical (devenu depuis CIRAD) à environ 50 km au Nord-ouest de Kourou en forêt primaire en Guyane Française. Il a été mis en place dans le but d'améliorer les connaissances de l'écosystème forestier tropical et plus particulièrement d'étudier les effets de différents traitements sylvicoles sur la dynamique des espèces d'arbres. Au total, le dispositif de Paracou comprend 16 parcelles pour une surface totale de 120 hectares (Figure 4). Entre 1986 et 1988, des parcelles de 6,25 hectares (9 hectares en incluant les zones tampon entre chacune de ces parcelles) ont été exploitées à des niveaux de traitements différents :

- Traitement 0 : absence d'exploitation (parcelles témoins)
- Traitement 1 : exploitation "traditionnelle" pour le bois d'œuvre, soit prélèvement d'environ 10 arbres/hectare faisant plus de 50 ou 60 cm d.b.h. (« diameter breast height » ; diamètre à 1m30 de hauteur), choisis parmi une cinquantaine d'essences forestières de valeur.
- Traitement 2 : exploitation pour le bois d'œuvre comme pour le traitement 1, puis éclaircie par dévitalisation d'environ 30 arbres/hectare de plus de 40 cm d.b.h. appartenant à des essences sans intérêt commercial.
- Traitement 3 : exploitation pour le bois d'œuvre comme pour le traitement 1, prélèvement d'environ 15 arbres/hectare supplémentaires faisant entre 40 et 50 cm d.b.h., pour alimenter une unité de production d'énergie, puis éclaircie par dévitalisation d'environ 20 arbres/hectare de plus de 50 cm d.b.h. appartenant à des essences sans intérêt commercial.

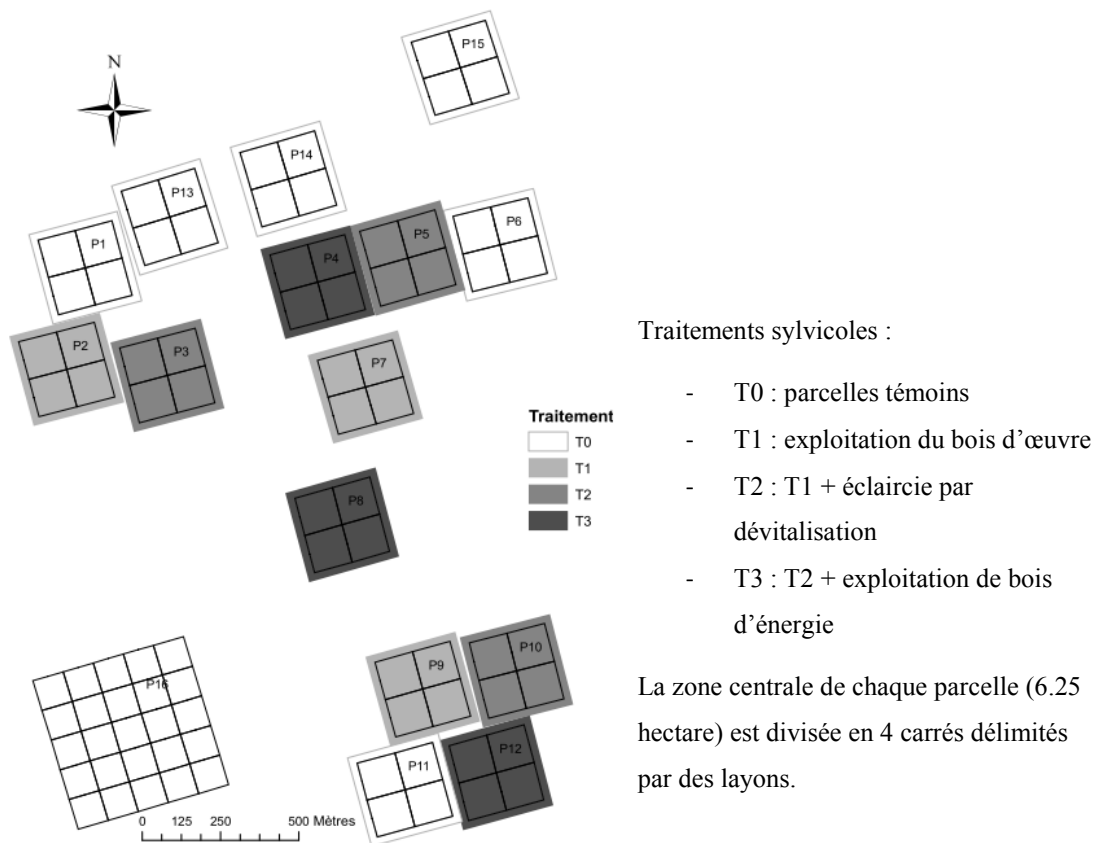


Figure 4 : schéma du dispositif de Paracou

Dans chaque parcelle, les arbres de plus de 10 cm de d.b.h. ont été identifiés botaniquement et cartographiés. Les trois composantes de la dynamique forestière (accroissement, mortalité et recrutement) sont suivies depuis 1984 pour les 12 premières parcelles et depuis 1991 pour les autres.

Outre les données botaniques, plusieurs types de données caractérisant l'environnement des parcelles sont disponibles :

- la topographie (courbes de niveaux équidistantes de 2,5 m)
- une cartographie des criques permettant de localiser les cours d'eau
- une cartographie fine des conditions édaphiques
- une cartographie des dégâts (trouées, pistes de débardage) consécutifs aux traitements sylvicoles

La topographie permet de différencier différentes zones : les plateaux qui restent émergés continuellement (altitude relativement plus élevée), les bas-fonds occasionnellement inondés en saison des pluies (zone d'altitude faible) et les pentes (zones de transition).

Plusieurs facteurs édaphiques relatifs à l'humidité des sols ont aussi été décrits tel que le drainage (Morneau 2007) qui utilise la typologie établie par Vincent Freycon (pédologue au CIRAD de Montpellier) :

- DVD (« Deep Vertical Drainage ») : le drainage vertical libre profond. Le sol a une structure microagrégée sur au moins 1,2 m de profondeur et est une caractéristique des plateaux.
- Alt : apparition d'une allotérite (zone d'altération dans laquelle l'organisation originelle du matériau parental n'est plus reconnaissable (Ferry *et al.* 2003). Le drainage est toujours vertical et profond, mais ralenti au niveau de l'allotérite. Comme pour DVD, Cette caractéristique environnementale ne se retrouve qu'au sommet de quelques collines (plateaux).
- SLD (« Superficial Lateral Drainage ») : le drainage latéral superficiel. La circulation de l'eau est latérale, au dessus de l'allotérite.
- UhS (« Uphill System ») : le système hydromorphe amont. Il s'agit de sols proches des SLD mais pour lesquels l'écoulement latéral est très lent du fait de leur situation sur pente faible.
- DhS (« Downhill System ») : le système hydromorphe aval. Ces sols sont soumis aux battements d'une nappe permanente et se situent donc en bas de pente, non loin des cours d'eau.
- SH : les sols hydromorphes qui englobent les bas-fonds.

L'engorgement des sols correspondant à l'hydromorphie de surface (Cantet 2004) a également été mesuré et indexé.

Grace à cette caractérisation fine de l'environnement, le site de Paracou fournit un outil privilégié pour diverses études écologiques en milieu forestiers et notamment l'étude de l'adaptation locale.

C. Objectifs

L'objectif de cette thèse est de caractériser la diversité et la variabilité génétique chez les arbres forestiers tropicaux pour décrire les processus évolutifs qui en sont à l'origine. L'étude menée au niveau interspécifique et intraspécifique se focalise sur le genre *Virola* (Myristicaceae) qui constitue un bon modèle pour l'étude de la génétique écologique.

Si une diversité génétique suffisante (au moins à certains loci) est gage de prospérité pour le futur des espèces, il paraît primordial d'effectuer des études d'impacts suite à certaines perturbations du milieu, notamment suite aux exploitations forestières si nombreuses dans la forêt amazonienne. Le site de Paracou offre un site expérimental adéquat pour ce genre d'analyse car certaines parcelles ont été exploitées et d'autres non, permettant de faire une comparaison des deux environnements. Ainsi dans cette première partie de thèse, l'étude de la régénération a été suivie à travers l'analyse de la structure génétique et spatiale par des marqueurs AFLP chez des jeunes arbres de *V. michelii* issus de la perturbation ou non.

L'étude de la diversité génétique intra-spécifique s'avère utile pour la compréhension de la capacité d'adaptation d'une espèce aux variations de son environnement et donc de sa pérennité dans le temps. Ainsi, dans la deuxième partie, une étude de l'adaptation locale, par une approche de génomique des populations, a été menée chez *V. michelii* sur le dispositif de Paracou. Même si cette espèce montre des préférences pour des milieux relativement peu humide (Baraloto *et al.* 2007), il a été montré que certains individus présentent des différences de répartition vis-à-vis de la contrainte hydrique. Un important échantillonnage d'individus de différents stades de vie de l'espèce *V. michelii* a été entrepris sur le site de Paracou. Ainsi, les données génétiques (AFLP) mises en corrélations avec la géolocalisation des individus (et donc avec leur données environnementales, notamment hydriques), a permis d'étudier l'adaptation des individus à leur microenvironnement. De plus, si une signature adaptative peut être renseignée, il paraît intéressant de savoir à partir de quel stade de développement de l'arbre elle est détectable, c'est-à-dire à partir de quel moment les filtres environnementaux auront agité.

La distribution et la grande diversité morphologique et d'habitats écologiques des espèces de *Virola* en Guyane française laissent présager une spéciation écologique. Pour tester cette hypothèse, trois espèces de *Virola* ont été choisies : *V. michelii*, *V. surinamensis* et *V. kwatae*. Pour cette étude, présentée dans la troisième partie, dix marqueurs génétiques microsatellites nucléaires et deux marqueurs de séquences chloroplastiques (Encadré 10) ont été utilisés, chez une trentaine d'individus par espèce, pour essayer de reconstruire l'histoire évolutive de ces espèces à partir des niveaux de diversité et de différenciation génétique interspécifique.

Encadré 10 : Les différents génomes des plantes et quelques caractéristiques

Trois génomes sont présents chez les plantes : le génome nucléaire, chloroplastique et mitochondrial (Birky Jr *et al.* 1989; Ennos *et al.* 1999). Les forces évolutives n'agissent pas identiquement sur ces différents

génomés qui possèdent des spécificités qui leur sont propres (seuls le génome chloroplastique et nucléaire seront abordés dans cette thèse).

Les génomes présentent des modes de transmissions différents : chez les angiospermes le génome chloroplastique (haploïdes) ne se transmet généralement que par la mère (donc ne dépend que du flux de graines), alors que le génome nucléaire (diploïde) présente une transmission à la fois maternelle et paternelle (dispersion par le pollen et par les graines). L'haploïdie du génome chloroplastique lui confère une taille efficace plus petite et donc une plus grande sensibilité à l'effet de la dérive génétique. Les marqueurs chloroplastiques montrent alors de plus grandes divergences génétiques entre populations que les marqueurs nucléaires (Petit *et al.* 2005).

Le génome chloroplastique ne recombine pas (on parle de transmission clonale) et présente un taux de mutation assez faible (entre 1.10^{-9} et 3.10^{-9} (Wolfe *et al.* 1987)). Les marqueurs chloroplastiques peuvent alors s'avérer utile pour des études phylogénétiques (Palmer *et al.* 1988; Gielly & Taberlet 1994) ou phylogéographiques (Avice 2000; Shaw *et al.* 2005). Cependant, des captures de génome chloroplastique par une autre espèce en cas d'introgression (par hybridation interspécifique) sont possibles (Rieseberg & Soltis 1991; Tsitrone *et al.* 2003) De ce fait, son histoire ne reflète pas nécessairement celle de la population ou même de l'espèce (Avice 2000; Ballard & Whitlock 2004).

Il est alors recommandé d'utiliser conjointement des marqueurs pour ces deux types de génomes pour reconstruire l'histoire évolutive des populations.

Enfin dans une conclusion générale, les principaux résultats seront rappelés et mis en relation pour qu'émerge une vision globale des forces pouvant générer les niveaux de diversité génétique chez les arbres tropicaux, en particulier chez le genre *Virola*.

RESULTATS

I. Article 1: Regeneration dynamics of a tropical rainforest tree: does intermediate disturbance maintain genetic diversity?

A. Résumé

Etude de la régénération chez *Virola michelii* (Myristicaceae) suite à une perturbation anthropique (exploitation forestière) : une perturbation intermédiaire peut être garante d'une diversité génétique suffisante.

Les interactions entre les événements démographiques et les patrons de diversité génétiques des populations peuvent avoir des conséquences sur la viabilité de celles-ci. En effet, si la taille démographique influence la reproduction, une diversité génétique importante empêche les risques de dépression de consanguinité et augmente le potentiel adaptatif des populations.

L'exploitation forestière est une activité importante et réduit inévitablement la taille des populations. L'extraction des essences commerciales a aussi des conséquences directes sur le milieu, notamment par la création de trouées (i.e. ouvertures de la canopée.) Pourtant, relativement peu d'études ont été menées pour connaître ces impacts sur les structures génétiques des populations futures. Il est difficile d'entreprendre des études génétiques sur les conséquences de ces modifications de milieux chez les arbres adultes car les temps de génération sont généralement longs. Par contre des études sur la régénération semblent être intéressantes. En effet, l'étude de la structure génétique et spatiale des plantules et des jeunes arbres forestiers est d'une forte importance car elle déterminera la structure de la future cohorte des adultes reproducteurs et donc l'avenir des prochaines populations, et peut alors avoir une grande utilité dans les programmes de conservation.

Un échantillonnage exhaustif de juvéniles de l'espèce *Virola michelii* a été réalisé sur une parcelle ayant subi une exploitation forestière caractérisée de faible intensité et sur une parcelle témoin du bloc sud du dispositif de Paracou. Les caractéristiques génétiques et démographiques ont été étudiées chez des juvéniles *V. michelii* suite à cette exploitation forestière. 229 marqueurs AFLP ont été utilisés pour étudier l'impact de la perturbation sur la diversité génétique sur 383 juvéniles. Différentes comparaisons ont été réalisées pour étudier les effets des perturbations : d'une part nous avons testé l'effet de perturbations artificielles en

comparant les individus provenant de la parcelle exploitée avec les individus provenant de la parcelle témoin et d'autres part sachant que des perturbations naturelles existent aussi (chablis, i.e. chutes d'arbres), nous avons aussi choisi de regrouper les individus par zones de fortes densités ou de faible densité. En effet, l'apport lumineux engendré par les trouées naturelles et artificielles tend à augmenter la régénération et donc augmenter la densité des jeunes arbres.

Les analyses d'autocorrélation génétique spatiale (qui mesurent la corrélation entre les degrés de similitude génétique et les distances géographiques) montrent que les individus issus des zones perturbées présentent une plus forte structure génétique spatiale que chez les autres individus mais avec une diversité génétique plus élevée. Etant donné que la taille des populations peut interagir sur les valeurs de diversité, une méthode de raréfaction par bootstrap (RaBoT) a été élaborée pour pouvoir s'affranchir de ces éventuels effets stochastiques lors des comparaisons de diversité entre populations.

Le logiciel d'assignation par le bayésien BAPS a permis de visualiser la structuration génétique spatiale des individus. Cette analyse indique que des groupes génétiquement proches sont disposés en patch et explique la forte structuration observée. Pour essayer de comprendre ce phénomène, et ayant à disposition un échantillonnage de tous les individus adultes *V. michelii* potentiellement matures du dispositif de Paracou (N=77), une étude de parentalité a été effectuée par le logiciel FaMoZ (Gerber *et al.* 2003). Celle-ci suggère que seul un faible nombre de reproducteurs contribue à la régénération, que les gamètes de ces arbres sont largement dispersés sur le site d'étude et que les patches identifiés par l'assignation bayésienne ne comprennent pas forcément des juvéniles présentant les mêmes parents.

Plusieurs mères se situent loin de leurs progénitures, mettant en avant une capacité de dispersion des graines à grande distance (d'une centaine de mètres à plus d'un kilomètre). Associé aux fortes structurations spatiales observées, cela pourrait montrer que des graines provenant du même arbre sont dispersées ensemble. Une des hypothèses émises est que les disséminateurs de graines (toucans et singes principalement) préfèrent consommer les fruits qu'ils avaient récoltés dans les arbres aux alentours, au niveau des trouées. Outre la dispersion à longue distance, ils assureraient un bon mixage des génotypes. Une autre hypothèse, qui n'est pas forcément incompatible avec la précédente, est que la plus grande diversité génétique des zones perturbées (surtout en leur périphérie, du fait que les individus se positionnent préférentiellement en périphérie des trouées) pourrait être due à la diversité des micro-habitats présents en ces zones (notamment par le gradient de lumière) favorisant la

survie d'un plus grand nombre de génotypes. Bien que l'identification du sexe des individus soit incomplète, nous pouvons aussi émettre l'hypothèse que les trouées facilitent la dispersion du pollen en réduisant les barrières physiques aux flux polliniques, donc que potentiellement plus de mâles contribuent à la reproduction augmentant ainsi la diversité génétique.

Cette étude de la régénération montre que certaines espèces d'arbres comme *V. michelii* peuvent répondre positivement à certains types de perturbation. En effet, nous avons remarqué une augmentation de la diversité génétique au niveau des trouées d'exploitation forestière.

B. Article

Regeneration dynamics of a tropical rainforest tree: does intermediate disturbance maintain genetic diversity?

Ivan Scotti¹, William Montaigne², Klára Cseke³ & Stéphane Traissac⁴

¹ INRA - Institut National de la Recherche Agronomique, Unité Mixte de Recherche « Ecologie des Forêts de Guyane » (EcoFoG), Kourou, French Guiana

² UAG - Université des Antilles et de la Guyane, Unité Mixte de Recherche « Ecologie des Forêts de Guyane » (EcoFoG), Kourou, French Guiana

³ Forest Research Institute, Department of Forest Tree Breeding, Experiment Station and Arboretum, Sárvár, Hungary

⁴ AgroParisTech, Unité Mixte de Recherche « Ecologie des Forêts de Guyane » (EcoFoG), Kourou, French Guiana

Cet article a été soumis à la revue scientifique *Ecology*.

Abstract

The interplay between genetic and demographic processes can have major consequences on the population viability. Demographic size is key to sustained population reproduction, while genetic diversity can insure viability by reducing risks of inbreeding depression and by maintaining adaptive potential. Forest trees generally display high levels of within-population genetic diversity, moderate levels of spatial genetic structure and relatively large effective sizes. The impact of population reductions is well studied; less attention has been paid to the consequences of census bursts on genetic diversity and population structure. We have studied the properties of a cohort of saplings of the Neotropical tree, *Virola michelii* (Myristicaceae) after forest disturbance caused by experimental low-intensity logging. To do this, we have developed a statistical test to compare multi-locus diversity in uneven-sized populations at the genome level and applied it to 229 AFLP loci. The new test (RaBoT) is based on a rarefaction approach with bootstrap re-sampling. Post-disturbance cohorts display a more local spatial genetic structure than background regeneration cohorts, but they are globally more genetically diverse. Distinct groups of genetically related saplings could be identified by Bayesian assignment analyses; these groups are spatially clumped and can explain the presence of strong spatial genetic structures. Parentage analysis suggested that a limited number of parents contributed to most of the regeneration, but that efficient mixing of these progenies contributed to the increased diversity of cohorts occupying canopy gaps. This suggests that, on the short term, *V. michelii* genetic diversity globally benefits from small-scale forest disturbance.

Keywords: Neotropics, population genetics, tropical trees, plot dynamics

Introduction

Regeneration dynamics influence the ability of plant species and populations to maintain sustainable levels of genetic diversity, not only through the effects of sheer demographic size, but also through the consequences of variations in effective sample size (N_e), which in turn determines the degree of relatedness of individuals in a gene pool, and ultimately a population's level of genetic diversity (Charlesworth 2009). Although a minority of the saplings that grow in a forest will eventually become fertile adults, sapling cohorts are the stock those adults are drawn from; the genetic structure of future adult populations has its foundations in the genetic diversity of present-day juvenile cohorts. In highly diverse ecosystems, where a given species' population density can be extremely low, such as tropical forests, the study of parental population effective size and of new generations' genetic diversity is therefore of paramount importance. In some cases it has been proven that parental fertility is highly unequal (Murawski & Hamrick 1991), leading to small effective population sizes, while in other instances (Nason et al. 1998) N_e has been found to be surprisingly large. The effect of the unequal contribution of parental trees to regeneration has drawn much attention in the case of sudden bursts of regeneration after disturbance or at forest edges. Sudden forest growth after disturbance in secondary forests can be associated to the establishment of patches of genetically uniform individuals derived from seed dispersal from a limited number of parent trees (Sezen *et al.* 2005; Sezen *et al.* 2007; Davies *et al.* 2010). The same pattern can be observed in regeneration at forest edges (Piotti *et al.* 2009), in forest patches established by rare long-distance dispersal events (Pardini & Hamrick 2008; Fayard *et al.* 2009), in species with limited seed dispersal distance (Born et al. 2008b) or because of intrinsic individual variation in fecundity (Klein et al. 2008). Conversely, dispersal agents such as birds and mammals can actively compensate local diversity loss by actively moving groups of seeds across large distances, so that in some cases local potential parents are under-represented in regeneration after disturbance (Aldrich & Hamrick 1998; Aldrich *et al.* 1998). In other cases, a reduction in local density, caused by forest fragmentation, can result in increased long-distance pollination and increased diversity by enhanced mixing of pollen clouds (White *et al.* 1999; White *et al.* 2002; Bizoux *et al.* 2009). The effect of forest logging on genetic diversity levels in post-logging generations has been studied experimentally (Cloutier et al. 2007) and, in species with known mating system parameters, by a modelling approach (Degen et al. 2006). These studies did not detect any significant change in genetic diversity on the short and long term for moderate logging intensity.

The outcome of genetic processes associated to regeneration in natural forest tree populations depends strongly on the spatial distribution of genetic diversity both in the parental and the offspring population. Spatial patterns of genetic variation (Epperson 2003) have major consequences on the availability of genetic diversity for selection and on the long-term individual inbreeding levels, especially for species with leptokurtic dispersal curves. The development of general models of the distribution of genetic diversity, such as the island and the stepping-stone models (Wright 1943), has ultimately been motivated by the need to predict the impact of spatial structuring on within-population genetic diversity levels. Spatial genetic structure of plant populations has been the focus of intense study, often under the form of detection of spatial autocorrelation of genetic identity (Epperson 1990) or genetic relatedness (Hardy & Vekemans 1999; Vekemans & Hardy 2004) among individuals. Spatial genetic structure under the form of spatial autocorrelation is a general, albeit synthetic, way of representing population structure. It has been systematically studied in temperate and tropical forest trees as a method to detect demographic differences between populations, between age classes and between regeneration cohorts having undergone different demographic processes. Spatial genetic structure in adult populations is very often found in forest trees (Brunel & Rodolphe 1985; Ueno *et al.* 2000, 2002; Degen *et al.* 2004; Dutech *et al.* 2005; Luna *et al.* 2005; Hardy *et al.* 2006; Born *et al.* 2008a; Born *et al.* 2008b; Scotti *et al.* 2008), sometimes with strength proportional to population density (Valbuena-Carabana *et al.* 2007), and sometimes without a corresponding structure in seedlings or saplings (Epperson & Alvarez-Buylla 1997; Latouche-Hallé *et al.* 2003; Jones & Hubbell 2006; Ng *et al.* 2006; Davies *et al.* 2010); in other cases, spatial genetic structure appears in young cohorts but disappears in older age classes (Hamrick *et al.* 1993; Chung *et al.* 2003b; Jones *et al.* 2007; Biscaia de Lacerda *et al.* 2008). Rarely, any structure at all is found (Chung *et al.* 2003a; Hardesty *et al.* 2005). Clearly, spatial genetic structure is not in absolute terms a property of one or the other age class, and the properties of the spatial distribution of genetic diversity depend on several factors, such as mating system, pollination and dispersal distance, degree of overlap of generations and rate of population turnover.

Given the variability of patterns described above, it is impossible to provide a generalisation of what the “ideal” level of genetic diversity and spatial genetic structure should be. On the other hand, evaluating the effect of habitat changes on genetic diversity and spatial structure is a realistic goal, and an interesting one. However, comparing mature stands having undergone disturbances in the far past with undisturbed stands is complicated, because it is hard to find

precise records of very old disturbances and therefore it is virtually impossible to define the “cases” and the “controls”. Sezen *et al* (2005; 2007) have compared adult and sub-adult cohorts in a secondary forest and in an adjacent old-growth forest, finding reduced diversity and loss of spatial structure in the secondary growth; Ueno *et al* (2006) have compared sapling genetic diversity and spatial genetic structure in cohorts growing under forest cover and in gaps, finding weaker structure in gaps and comparable levels of genetic diversity. These two examples show that changes in the demography of juvenile age classes can modify population genetic structure and motivate further studies. In particular, it is worthwhile to assess the impact of canopy gaps of natural or anthropogenic origin on regeneration dynamics in primary tropical rainforests, in which the genetic consequences of habitat change are poorly known.

We have studied this question in the common primary forest Neotropical tree species, *Virola michelii* (virola), in two forest plots in Paracou (French Guiana), one having undergone one bout of experimental logging between 1986 and 1988 and the other left undisturbed. The Paracou forest inventory site is the subject of recurrent censuses, and the availability of detailed pre- and post-disturbance forest inventory data and of forest disturbance maps makes it a unique tool for the study of regeneration processes. Disturbance caused by logging is expected to induce a sudden, short-lived burst of regeneration of light-responding species, while regeneration in control plots is driven by long-term natural disturbances and background understory growth. The study of cohorts growing in disturbed and control areas, and the study of relatedness and population spatial genetic structure allowed us to draw conclusions on the consequences of moderate disturbance on the level and distribution of genetic diversity in the stand.

Materials and methods

Virola michelii has been classified either as a light-demanding (Delcamp et al. 2008) or as a shade-tolerant (Forget 1991) dioecious canopy tree species that preferentially grows on *terra firme* soils. *V. michelii* has a short seed dormancy of 4-6 weeks, during which embryos develop prior to germination (Forget 1991). The species is common in the Guiana Shield, and its seedlings can rapidly colonise the understory and forest canopy gaps. Indeed, Forget (1991) has shown that survival of juveniles is greater in gaps than understory and defined the species as dependent on small canopy gaps. However, seedlings can colonise the understory and survive until a canopy gap favouring their growth eventually appears (Forget 1991). *V. michelii* is common in secondary forests, and man-made forest disturbances are therefore an important source of favourable sites for its regeneration. The conspecific, ecologically close *V. kwatae* has been shown to quickly colonise forest disturbances (Forget & Sabatier 1997).

The dynamics of *V. michelii* regeneration was studied at Paracou (French Guiana; 5°18'N, 52°55'W; <http://www.ecofog.gf/spip.php?article174>) experimental site, which is part of the GUYAFOR forest plot network (<http://www.ecofog.gf/spip.php?article364>). *V. michelii* adult populations show intermediate levels of spatial genetic structure in the studied area (Degen *et al.* 2001; Hardy *et al.* 2006). The study was carried out in two contiguous 6.25 ha plots (plots 9 and 11), separated by a 50m-wide buffer zone. Plot 9 has undergone experimental logging between 1986 and 1988: a traditional selective felling for timber, representing about 10 trees/ha of diameter at breast height (d.b.h.) >50-60 cm, belonging to commercial species. Plot 11 was a control plot without any logging. An exhaustive sampling of all *V. michelii* saplings (d.b.h. < 10 cm, minimum height = 130 cm) in these two plots, as well as of all *V. michelii* inventoried trees with d.b.h. > 20 cm (*i.e.* potentially mature adults) across the Paracou site plots, was realized in 2008. In particular, two plots (Plot 10 and Plot 12) lie within 300 m from the two focal plots, and adult trees collected in these two plots are considered hereafter as belonging to the focal plots' "neighborhood". Diameter at breast height (d.b.h.) and geographical position were measured for each individual sapling. Sexing of reproductive trees was performed during one flowering and fruiting season. Each tree was inspected weekly for flowers and fruits. When large amounts of fruits (with or without seed) were found under a tree, it was considered to be a female. Flowers were collected under each tree, brought back to the laboratory and inspected with a binocular microscope to distinguish pistillate and staminate flowers. Flower type was used to infer the sex of the corresponding tree. Trees for which fewer than ten flowers could be described were assigned a tentative sex

(“likely male” or “likely female”). For genetic analysis, leaf or cambium samples were collected from these individuals, and kept in 2 mL tubes with silica gel until DNA extraction. Logging disturbance maps were matched to sapling geographical positions using the ArcGis9 software package to characterise sapling distribution relative to logging-generated canopy gaps. As both natural and man-made canopy gaps enhance virola regeneration, and its saplings are not preferentially located inside the limits of mapped canopy gaps (see results), we decided to partition virola saplings in two different ways: (a) as two cohorts, one from the treatment plot, and the other from the control plot, with the former including natural and man-made disturbances as well as undisturbed zones, and the latter including only natural disturbances and undisturbed areas; (b) as two cohorts, occurring at high or low sapling density, with the former including man-made disturbances and putative natural disturbances, and the latter including undisturbed areas. The method of Goreaud and Pélissier (1999) was implemented to compute local densities and to partition the two plots into high density (HD) and low-density (LD) zones. The density threshold separating these two zones depends on frequency of local density values. The threshold was established based on density distribution in the control plot. For both studied plots, we calculated sapling local density at each node of a 10m grid path, inside a 15m radius. To perform these analyses we used the ADS module of ADE4 R package. It has to be noted that the two partitions are non independent, as more high-density saplings are found in the treatment than in the control plot.

For all samples, total genomic DNA was extracted from cambium or leaf tissues according to the hexadecetyl trimethyl ammonium bromide (CTAB) procedure (Doyle & Doyle 1987). DNA quality was assessed by spectrophotometry or by agarose gel electrophoresis.

Amplified fragment-length polymorphism (AFLP) reactions were performed following Vos *et al* (1995). DNA samples were digested with the restriction enzymes *EcoRI* and *MseI*, and the digested fragments were ligated with *EcoRI* and *MseI* adapters. Pre-amplification was conducted using a primers combination with one additional selective nucleotide: *EcoRI+A* + *MseI+A*. Selective amplifications were performed with the following primer combinations: *MseI+AAA* / *EcoRI+AAG*, *MseI+AAA* / *EcoRI+ACC*, and *MseI+AAA* / *EcoRI+ATG*. The *EcoRI* primers for the selective amplifications were labeled fluorescently. Both amplification steps were performed using an ABI 9700 thermocycler (Applied Biosystems, USA). AFLP fragments were separated on an ABI 3130xl capillary sequencer (Applied Biosystems, USA) and scored using GENEMAPPER® in terms of the presence or absence of identified AFLP fragments. Peaks were visually inspected and only those showing a clear presence/absence

pattern, and whose peaks were higher than 50 units, were retained for subsequent analyses. To exclude low-quality samples, individual samples were ranked according to the total number of peaks produced with all primer combinations, and the 10% lower tail was excluded from subsequent analyses. Finally, the same procedure was applied to individual AFLP markers, and the markers in the 10% tail with the smallest numbers of peaks across all samples were dropped.

Spatial autocorrelation of kinship was estimated based on two kinship estimators specifically developed for dominant markers, “Kinship” and “Relationship” statistics (equations (1) and (2) in Hardy (2003); the two statistics estimate similar parameters, but the latter corrects inter-individual relatedness by both individual inbreeding and population-averaged relatedness, whereas the former takes only average relatedness into account). Calculations were performed by SPAGeDi version 1.3 (Hardy & Vekemans 2002) by setting even samples sizes for each distance class and a total of ten classes for each analysis. Distances within pairs varied between 0 and 600 m. Three independent analyses were executed for HD and LD cohorts and for the treatment and the control plots with inbreeding coefficient F set to -0.1, 0 and 0.1 to take unknown mating system parameters into account, based on independent inbreeding estimates obtained on codominant markers (Montaigne et al., in prep.).

Population genetic structure was assessed by automated Bayesian assignment using BAPS (Corander *et al.* 2006; Corander *et al.* 2008a; Corander *et al.* 2008b). The algorithm was allowed to pick the most likely K (number of clusters), with $1 < K < 30$ and five runs for each K . Bayesian clustering was performed with and without spatial information as a prior, while keeping all the remaining settings equal. Departures from random spatial distribution of Bayesian clusters were tested by autocorrelation analysis. Moran’s I was computed on individual trees’ cluster identity using SPAGeDi version 1.3 (Hardy & Vekemans 2002) and departures from random distribution of clusters was tested with the same procedures as above.

To test for differences in diversity between high- and low-density sapling cohorts, between genetic clusters and the global sapling population, between the control and the treatment plot and between sapling and adult cohorts, a bootstrap approach was applied to the comparison of Simpson diversity index (H) observed for each group, as obtained based on frequencies of presence/absence of the PCR fragment in each group for each of the 229 AFLP markers. Populations with different census sizes may display different levels of diversity, at a given marker, either because their levels of diversity are truly different or because stochastic effects increase the variance of observed diversity in smaller populations, thus increasing the chances

that their observed diversity, at any given marker, happens to be smaller or larger than in the larger population only due to random fluctuations. Thus, a difference in observed diversity values between populations with different census sizes is not *per se* an indication that demographic events have affected one or the other population. The bootstrap-based, rarefaction method developed here (Rarefaction-by-Bootstrap Test, “RaBoT”) is analogous to rarefaction-based measures of species and allele richness (Hurlbert 1971; Petit *et al.* 1998) and aims to suppress the stochastic effects in the comparisons of diversity values in pairs of populations with different census sizes. To obtain this result, we proceed as follows. Census sizes are recorded as N_{small} , N_{large} for the smaller and larger population being compared, respectively. The observed diversity in the smallest population (H_{small}) is computed for each AFLP marker. In the larger population, 1000 samples are randomly drawn with sample sizes equal to N_{small} , H is computed for each replicate and for each marker, and the median $m_{H,large}$ is computed from the set of 1000 H values obtained for each marker. Finally, for each marker $\Delta H = H_{small} - m_{H,large}$ is computed. The genome-wide distribution of ΔH (i.e. the distribution of ΔH values drawn from all markers) is inspected. If there is no real difference in diversity between the small and the large populations, beyond the effect of population sampling stochasticity, then H_{small} values should be drawn, for each locus, from the same distribution as the random subsamples from the large population and the same locus. Under this hypothesis, the observed H_{small} value for each marker is a random outcome from the same distribution as the corresponding distribution of H values for the random subsamples. As a consequence, individual-marker ΔH values should have the same probability of being positive or negative, and at the genome level there should be equal numbers of positive and negative ΔH values. Departures from this expectation are tested by a Chi-square test and are taken as suggestive of real, genome-wide differences in diversity. This method has been applied to the following comparisons of H values (smaller population listed first for each pair): saplings from the control and the treatment plot; LD and HD cohorts; each Bayesian cluster and the global population (RaBoT was not applied to clusters with fewer than five individuals.); cohort of adult trees within the study area (*not* the population of all potential sampled parents) and the juvenile cohort.

Parentage analyses were performed to identify the relative contribution of individual trees to the regeneration cohort, and to assess the effect of dispersal and of fertility variation on the genetic diversity of the regeneration cohorts in undisturbed and disturbed portions of the forest. Parentage analyses were performed using FaMoz (Gerber *et al.* 2003), with typing

error rate set to 0.001 for both LOD calculation and simulations, and the same three levels of inbreeding coefficient F (-0.1, 0, 0.1) used for the spatial autocorrelation of kinship estimate. Samples with missing data were included in the calculations. Simulations were first performed to assess the power and accuracy of the test, and then parentage analyses were run on the real sapling dataset to identify the most likely parent pair for each sapling. All sexed and not sexed trees with diameter at breast height > 20 at the date of controlled logging were used as potential parents. A parent pair was assigned to a seedling when the Δ LOD relative to the second most-likely parent pair was higher than 3.0. Parent pairs with incompatible sexes (two males or two females) were not retained. Parent pair analyses were preferred over single-parent analyses because parent identification was unclear in the latter case (i.e. Δ LODs were generally small), although it provided a larger number of potential parents.

Results

A total of 627 juveniles (374 in the logged plot, 253 in the control plot) and 90 adult trees (in the whole Paracou site) were collected and genotyped. After dropping AFLP bands with extreme frequencies and individuals with poor banding profiles, a dataset including genotypes for 77 mature trees and 383 juveniles, characterised at 229 AFLP markers, was obtained. Generally speaking, DNA extracted from cambium samples performed better than DNA extracted from leaves, partly explaining the difference in discard rate between adults (exclusively sampled from cambium) and juveniles (sampled from either tissue based on ease of collection).

The distribution of saplings in the treatment and control plots only poorly overlapped the contours of canopy gaps as detected after controlled logging, although the densest areas correspond to the zones in, around and between large gaps (Figure 1a). To identify groups of saplings whose growth had or had not been induced by (natural or man-made) disturbance, the study plots were subdivided into several “high density” (HD) zones and a “low (background) density” (LD) zone, instead of separating samples on the base of canopy gap contours, which would have artificially split continuous groups of seedlings.. The threshold was set at 60 stems/ha, corresponding to the strongest break in density distribution slope for both plots (Supplementary Figure 1;). The layout of corresponding isodensity level curves allowed us to define the contour separating HD and LD sub-areas (figure 1b). Zones with high seedling density were detected in the control plot as well as in the treatment plot, suggesting that unmapped natural canopy gaps may have favoured *virola* regeneration. Overall, 118 saplings were assigned to the low-density (LD) area and 265 to high-density (HD) areas. Since no differences were observed between spatial autocorrelation analyses run for three values of the inbreeding coefficient F , only the results for $F = 0.1$ are shown here. Sharp differences could be observed between the autocorrelograms of the treatment and the control plots (Figure 2a): the treatment plot showed almost strictly monotonous curves, with highly significant slopes (b (slope) values: -0.0005 and -0.001 for the Kinship and the Relationship parameter, respectively; P -value < 0.0001 for both), while the control plot showed virtually no structure (b (slope) values: -0.00005, $P = 0.18$, and -0.00006, $P = 0.18$, for the Kinship and the Relationship parameter, respectively). On the other hand, autocorrelation patterns for both the LD and the HD cohorts (Figure 2b) show a negative slope for kinship as a function of spatial distance (b (slope) values: for LD, -0.0002 and -0.0004 for the Kinship and the Relationship parameter, respectively: for HD, -0.0001 and -0.0002 for the Kinship and the Relationship

parameter, respectively; P -value < 0.0001 for all autocorrelograms) and are qualitatively similar at short distances, but differ at longer distances, with autocorrelation values tending to become non-significant at longer spatial distances in HD.

To identify the genetic structure underlying these patterns, an automated Bayesian analysis was run to infer clusters of genetically related saplings. Since the results obtained with and without prior spatial information are qualitatively similar, only the analysis with prior spatial information is described here. Fourteen clusters were identified as the most likely subdivision (Supplementary Figure 2). The analysis of spatial autocorrelation of cluster distribution reveals (Supplementary Figure 3a) that individuals belonging to the same cluster are strongly aggregated. The pattern appears to be similar for the treatment and the control plot (Plot 9 and Plot 11; Supplementary figures 3b and 3c, respectively), but, in agreement to the above results where autocorrelation is stronger in the treatment than in the control plot.

Genome-wide levels of diversity were compared among cohorts by the Rarefaction-by-Bootstrap Test (RaBoT) method. Diversity was significantly lower in the control than in the treatment plot, with 72% of negative ΔH values ($\chi^2_{1 \text{ d.f.}} = 53.7$, P -value < 0.0001 ; figure 3), and in LD than in HD cohorts (Supplementary Figure 4a), with 59% of negative ΔH values ($\chi^2_{1 \text{ d.f.}} = 8.07$, P -value = 0.004), indicating that increased sapling density is associated to increased diversity. The diversity of adult trees growing in the two focal plots was marginally lower than the diversity of the whole juvenile cohort (56% of negative ΔH values; $\chi^2_{1 \text{ d.f.}} = 3.23$, P -value = 0.072; Supplementary figure 4b), but it was larger than in the HD juvenile cohort (42% of negative ΔH values, $\chi^2_{1 \text{ d.f.}} = 4.18$, P -value = 0.040; Supplementary figure 4c) and identical to LD cohort's (50% of negative ΔH values, $\chi^2_{1 \text{ d.f.}} = 0.005$, P -value = 0.943; Supplementary figure 4d). Finally, clusters identified by automated Bayesian assignment showed lower-than-average diversity: for seven clusters out of ten, ΔH distribution is significantly skewed towards negative values (P -value < 0.05 ; Supplementary Table 1), indicating lower diversity than in the global juvenile population (Supplementary Figure 5). The fraction of markers with negative ΔH values increases with cluster size (Supplementary figure 3, lower right pane), indicating that the result is not an artefact caused by small sample sizes.

Parentage analyses were performed to identify patterns of reproduction and dispersal that may be related to the observed genetic structures and levels of diversity. Simulation tests showed that, conditional on sampled parent genotypes, parenthood was correctly identified (i.e. the

most likely parents and parent pairs were the correct ones) for 88% to 100% of the simulated progenies (out of 5000 simulated offspring genotypes and over three values of inbreeding coefficient F (-0.1, 0, 0.1)). A parent pair was identified among the 77 genotyped mature trees for 217 saplings out of 383 (57%). Fifty-nine saplings out of 383 (15%) had at least one parent within the study plots, and one had both. Of these 57 samplings, 47 had their second parent in the neighbourhood of the study plots (<300 m apart); 110 additional saplings had at least one parent in the same neighbourhood. Twenty-seven putatively mature trees, out of the 77 trees sampled, contributed to identified parent pairs. Of these twenty-seven trees, six were females, ten were males, and eleven could not be unequivocally sexed. Two females out of six were in the plots or in the neighbourhood, as well as four males out of ten. Six trees of both sexes contributed individually more than 5% of the offspring and 84% cumulatively (Table 1). Of these six trees, sex could be determined for three (Table 1). The best contributor (which could not be sexed; 27% of the total parenthood) was in the neighbourhood of the study area, while the only contributor tree inside the study area was a female having contributed 14% of the seeds (Table 1). Seedlings derived from all six parents were found scattered throughout the study area (Figure 4), with no particular trend for clumped distributions. Nevertheless, Bayesian clusters, which are non-randomly distributed (see above), had non-overlapping sets of potential parents. Differences in the composition of the putative parent pool were inspected for clusters for which at least twenty parents could be identified (Supplementary tables 2 and 3 and supplementary Figure 6). The proportions of parental contributions to the largest cluster (Cluster 13) were not statistically different from those of the remaining five clusters included in this analysis, while the proportions of all other clusters statistically different from each other at the 5% statistical threshold (Supplementary Table 3).

Discussion

The genetic dynamics of post-disturbance regeneration of the light-demanding tree, *Virola michelii*, were studied here. The results presented describe a picture of overall mixing of genotypes at the landscape level, with a slight majority (56%) of seeds having no parent in the study area or in neighbouring plots (a relevant portion of the saplings whose parents are not accounted for may, nevertheless, be the progeny of unsampled trees located in the forest surrounding the focal plots). Parents located at distances larger than 1 Km from the study plots account for 50% of the identified parental contribution: 7% from females, 11% from males, and 32% from trees with undetermined sex. Given that it is easier to determine sex for females than for males, it is likely that trees with undetermined sex are mostly males. In spite of the apparent spatial mixing, groups of closely related saplings could be found in the post-disturbance as well as in the control plot: fourteen genetic clusters were identified by automated Bayesian assignment, and those clusters were globally restricted to a limited part of the study area, although different clusters had overlapping distributions. This is confirmed by the presence of a clear autocorrelation structure in genetic relatedness in the juvenile cohorts, particularly in the treatment plot and in the high-density cohorts, which agrees with data on the adult population on the same site, though on a larger scale (Degen *et al.* 2001; Hardy *et al.* 2006). Therefore, at the sapling stage, populations are not genetically continuous and genetic subgroups are not entirely randomly distributed, although mixing occurs at the landscape scale. Moreover, based on autocorrelation structures, high-density and low-density areas differ in that high-density saplings tend to show a “local” structure, whereas low-density saplings show a “global” structure (Jombart *et al.* 2008) similar to adult populations. Differences are even deeper between the treatment and the control plot, with the former showing a very strong spatial structure and the latter almost no structure. Distant dispersal and strong spatial genetic structure seem to be contradictory, unless groups of related seeds are dispersed together and have at least a partially clumped distribution. The large imbalance of parental contributions, suggested by parentage analysis, seems to indicate that spatial structure is generated by the non-random dispersal of clumps of related seeds in areas favouring regeneration. The marked difference in spatial structure between the treatment and the control plot, and to a lesser extent between high-density and low-density areas, tends to support this view. The groups obtained by Bayesian assignment were found to have overall lower-than-average diversity. Therefore, individual genetic groups are genetically poorer than the global population. How does it translate into levels of diversity with and without disturbance? *V.*

michelii is not strictly heliophilic, and it is possible that the presence of more extremely light-demanding species reduces *virola* regeneration in forest clearings. Regeneration is – as expected – favoured by canopy gaps, and occupies small gaps as well as surrounding areas, where disturbance is weaker. This behaviour has already been observed in other tree species, i.e. *Thujopsis dolabrata* (Hoshino et al. 2001). Regeneration generates large continuous groups of saplings growing at high density, that were analysed as two subsets – high-density and low-density groups – to separately describe, and compare, saplings growing in (natural and artificially) perturbed areas and in continuous-cover portions of the forest. The RaBoT analyses show that genome-wide levels of diversity are higher in the high-density than in the low-density cohort and in the disturbed plot than in the control plot, after controlling for population size. At the same time, spatial structuring of genetic relatedness is similar between the high-density and low-density groups, at least at short spatial distances, and much stronger in the disturbed than in the control plot, unlike in other cases reported in the literature (Ueno et al. 2006; Sezen et al. 2007). Interestingly, the diversity of extant adult trees occupying the same area is higher than the diversity of low-density and high-density juvenile cohorts, but it is not statistically different than the diversity of the entire juvenile cohort. Given the way the RaBoT test is constructed, this is an indication that, if a subset of the current juveniles replaces extant adults with the same population density, the genome-wide diversity of the next adult generation will be comparable to current adult diversity; this is, however, granted only through the contribution of regeneration in both low-density and high-density areas. In other words, a mix of no disturbance and intermediate disturbance tends to favour the maintenance of genetic diversity in *virola*, thus indicating that intermediate disturbance hypotheses (Grime 1973; Connell 1978) hold for the maintenance of (neutral) genetic diversity in *virola*. Actually, given that the occurrence of high sapling density areas is favoured by the opening of canopy gaps by logging, it can be concluded that moderate logging impact maintains genetic diversity in *virola*. The present study cannot provide, on the other hand, any hint about the effect of more severe disturbance on genetic diversity.

At least two mechanisms may explain this observation. Dispersers (most likely large birds) may prefer, or have preferred at the time of disturbance, canopy gaps as resting areas. Dispersers would have fetched fruits from surrounding trees to the disturbed areas, thus contributing to increase the diversity of high-density cohorts relative to low-density cohorts. Parentage analyses performed in this study are consistent with dispersal of groups of related seeds over distances of at least 1 Km. Howe (1981), Howe and Richter (1982), Russo and

Angspurger (2004), and Holbrook and Loiselle (2007; 2009) have already observed that seeds from different *Viola* species (*V. sebifera*, *V. surinamensis*, *V. flexuosa* and *V. calophylla*) can be dispersed as far as several hundred meters, mainly by spider monkeys (Cebidae) and toucans (Ramphastidae), although dispersal distances directly observed for toucans (Holbrook 2011) are shorter than those indirectly estimated here. *V. michelii* seeds dispersal was studied by Forget *et al.* (2000) but not further than 60 meters. Because *V. michelii* seeds are similar in size to *V. calophylla*'s (Forget *et al.* 2000; Russo 2003; Russo & Augspurger 2004), it is possible that animals move large numbers of seeds over several hundred meters. Moreover, even if large amounts of non-dispersed seeds are present under mother tree crowns, it has been shown that they produce fewer surviving seedlings than dispersed seeds (Howe *et al.* 1985; Forget 1991; Russo & Augspurger 2004), in agreement with the Janzen-Connell hypothesis of density-dependent recruitment (Janzen 1970), thus reducing the importance of regeneration under closed canopy relative to disturbed areas. Alternatively, areas surrounding canopy gaps may provide more diverse habitats than the understory of closed canopy, thus increasing genetic diversity by favouring a more diverse array of genotypes (Ueno *et al.* 2006). In this case, the pattern observed for *viola* would fit the intermediate disturbance hypothesis (Grime 1973; Connell 1978) at the within-species level, and the diversity patterns observed at neutral markers would be explained by underlying selective mechanisms favouring different multi-locus genotypes in different habitat type patches. The observation that parent trees (both females and males) can be very far from the focal area, and that their offspring appears to be evenly scattered throughout the plots tends to favour the first hypothesis and to show that dispersers have been extremely effective at mixing genetic groups. However, the two hypotheses are not mutually exclusive, as selective processes may actually be facilitated by increased dispersal.

The patterns shown by *viola* regeneration indicate that, notwithstanding the very unequal contribution of parental trees to large juvenile cohorts, genetic diversity is maintained by regeneration bursts, and it can be expected that younger cohorts will give rise to adult cohorts as diversified as the current ones. This seems to contrast with patterns observed in pioneer or secondary-forest tree and palm species (Epperson & Alvarez-Buylla 1997; Sezen *et al.* 2007) but is consistent with genetic admixture in pasturelands in primary forest species (Aldrich & Hamrick 1998; Aldrich *et al.* 1998) and with models indicating no net loss of diversity in long-term regeneration with moderate logging levels (Degen *et al.* 2006). Globally, species with the same ecological and dispersal characteristics as *viola* may take advantage from

intermediate size natural and anthropogenic disturbance. Although this cannot be generalised to all species and all kinds of disturbance, this case study shows that at least some tropical rainforest species can respond positively to, and even benefit from, events with moderate ecological impact.

Acknowledgements

The Authors wish to thank Mathieu Toebosch for sexing adult trees, Saint-Omer Cazal for plant material collection and DNA extractions, and Audrey Jolivot and Jean-Yves Goret for producing plot maps. The research was funded by the “ECOFOR – Biodiversité et gestion forestière” program. We wish to thank Eric Marcon, Caroline Scotti-Saintagne, Rosane G. Collevatti, Andrea Piotti and Pierre-Michel Forget for critically reading the manuscript.

References

- Aldrich PR, Hamrick JL (1998) Reproductive dominance of pasture trees in a fragmented tropical forest mosaic. *Science* **281**, 103-105.
- Aldrich PR, Hamrick JL, Chavarriaga P, Kochert G (1998) Microsatellite analysis of demographic genetic structure in fragmented populations of the tropical tree *Symphonia globulifera*. *Molecular Ecology* **7**, 933-944.
- Biscaia de Lacerda AE, Kanashiro M, Sebbenn AM (2008) Effects of reduced impact logging on genetic diversity and spatial genetic structure of a *Hymenaea courbaril* population in the Brazilian Amazon forest. *Forest Ecology and Management* **255**, 1034-1043.
- Bizoux JP, Daïnou K, Bourland N, *et al.* (2009) Spatial genetic structure in *Milicia excelsa* (Moraceae) indicates extensive gene dispersal in a low-density wind-pollinated tropical tree. *Molecular Ecology* **18**, 4398-4408.
- Born C, Hardy OJ, Chevallier M-H, *et al.* (2008a) Small-scale spatial genetic structure in the Central African rainforest tree species *Aucoumea klaineana*: a stepwise approach to infer the impact of limited gene dispersal, population history and habitat fragmentation. *Molecular Ecology* **17**, 2041-2050.
- Born C, Kjellberg F, Chevallier M-H, *et al.* (2008b) Colonization processes and the maintenance of genetic diversity: insights from a pioneer rainforest tree, *Aucoumea klaineana*. *Proceedings of the Royal Society B: Biological Sciences* **275**, 2171-2179.
- Brunel L, Rodolphe F (1985) Genetic neighbourhood structure in a population of *Picea abies* L. *Theoretical and Applied Genetics* **71**, 101-110.
- Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* **10**, 195-205.
- Chung M, Nason J, Epperson BK, Chung M (2003a) Temporal aspects of the fine-scale genetic structure in a population of *Cinnamomum insularimontanum* (Lauraceae). *Heredity* **90**, 98-106.
- Chung M, Epperson BK, Chung M (2003b) Genetic structure of age classes in *Camellia japonica* (theaceae). *Evolution* **57**, 62-73.
- Cloutier D, Kanashiro M, Ciampi AY, Schoen DJ (2007) Impact of selective logging in inbreeding and gene dispersal in an Amazonian tree population of *Carapa guianensis* Aubl. *Molecular Ecology* **16**, 797-809.
- Connell JH (1978) Diversity in tropical rain forests and coral reefs. *Science* **199**, 1302-1310.
- Corander J, Marttinen P, Mäntyniemi S (2006) Bayesian identification of stock mixtures from molecular marker data. *Fisheries bulletin* **104**, 550-558.
- Corander J, Marttinen P, Siren J, Tang J (2008a) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* **9**, 539.
- Corander J, Sirén J, Arjas E (2008b) Bayesian spatial modeling of genetic population structure. *Computational Statistics* **23**, 111-129.
- Davies S, Cavers S, Finegan B, Navarro C, Lowe A (2010) Genetic consequences of multigenerational and landscape colonisation bottlenecks for a neotropical forest pioneer tree, *Vochysia ferruginea*. *Tropical Plant Biology* **3**, 14-27.
- Degen B, Bandou E, Caron H (2004) Limited pollen dispersal and biparental inbreeding in *Symphonia globulifera* in French Guiana. *Heredity* **93**, 585-591.
- Degen B, Blanc L, Caron H, *et al.* (2006) Impact of selective logging on genetic composition and demographic structure of four tropical tree species. *Biological Conservation* **131**, 386-401.
- Degen B, Caron H, Bandou E, *et al.* (2001) Fine-scale spatial genetic structure of eight tropical tree species as analysed by RAPDs. *Heredity* **87**, 497-507.

- Delcamp M, Gourlet-Fleury S, Flores O, Gamier E (2008) Can functional classification of tropical trees predict population dynamics after disturbance?, pp. 209-220. Blackwell Publishing Ltd.
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**, 11-15.
- Dutech C, Sork VL, Irwin A, Smouse PE, Davis FW (2005) Gene flow and fine-scale genetic structure in a wind-pollinated tree species, *Quercus lobata* (Fagaceae). *American Journal of Botany* **92**, 252-261.
- Epperson BK (1990) Spatial patterns of genetic variation within plant populations. In: *Plant Population Genetics, Breeding, and Genetic Resources* (eds. Brown AHD, Clegg MT, Kahler A, Weir B), pp. 229-253. Sinauer Associates, Sunderland, MA.
- Epperson BK (2003) *Geographical genetics* Princeton University Press, Princeton, NJ.
- Epperson BK, Alvarez-Buylla ER (1997) Limited seed dispersal and genetic structure in life stages of *Cecropia obtusifolia*. *Evolution* **51**, 275-282.
- Fayard J, Klein EK, Lefèvre F (2009) Long distance dispersal and the fate of a gene from the colonization front. *Journal of Evolutionary biology* **22**, 2171-2182.
- Forget P-M, Milleron T, Feer F, Henry O, Dubost G (2000) Effects of dispersal pattern and mammalian herbivores on seedling recruitment for *Virola michelii* (Myristicaceae) in French Guiana. *Biotropica* **32**, 452-462.
- Forget P-M (1991) Comparative recruitment patterns of two non-pioneer canopy tree species in French Guiana. *Oecologia* **85**, 434-439.
- Forget P-M, Sabatier D (1997) Dynamics of the seedling shadow of a frugivore-dispersed tree species in French Guiana. *Journal of tropical ecology* **13**, 767-773.
- Gerber S, Chabrier P, Kremer A (2003) FaMoZ: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. *Molecular Ecology Notes* **3**, 479-481.
- Goreaud F, Pélissier R (1999) On explicit formulas of edge effect correction for Ripley's K-function. *Journal of Vegetation Science* **10**, 433-438.
- Grime JP (1973) Competitive Exclusion in Herbaceous Vegetation. *Nature* **242**, 344-347.
- Hamrick JL, Murawski DA, Nason J (1993) The influence of seed dispersal mechanisms on the genetic structure of tropical tree populations. *Vegetatio* **107/108**, 281-297.
- Hardesty BD, Dick CW, Kremer A, Hubbell SP, Bermingham E (2005) Spatial genetic structure of *Simarouba amara* Aubl. (Simaroubaceae), a dioecious, animal-dispersed Neotropical tree, on Barro Colorado Island, Panama. *Heredity* **95**, 290-297.
- Hardy OJ (2003) Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers. *Molecular Ecology* **12**, 1577-1588.
- Hardy OJ, Maggia L, Bandou E, *et al.* (2006) Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. *Molecular Ecology* **15**, 559-571.
- Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity* **83**, 145-154.
- Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**, 618-620.
- Holbrook KM (2011) Home range and movement patterns of toucans: implications for seed dispersal. *Biotropica* **43**, 357-364.
- Holbrook KM, Loiselle BA (2007) Using toucan-generated dispersal models to estimate seed dispersal in Amazonia Ecuador. In: *Seed dispersal: theory and its application in a*

- changing world*. (ed. Dennis AJ, Schupp, E.W., Green, R. and Westcott, D.A.). CAB International Wallingford, UK.
- Holbrook KM, Loiselle BA (2009) Dispersal in a Neotropical tree, *Virola flexuosa* (Myristicaceae): Does hunting of large vertebrates limit seed removal? *Ecology* **90**, 1449-1455.
- Hoshino D, Nishimura N, Yamamoto S (2001) Age, size structure and spatial pattern of major tree species in an old-growth *Chamaecyparis obtusa* forest, Central Japan. *Forest Ecology and Management* **152**, 31-43.
- Howe HF (1981) Dispersal of a Neotropical nutmeg (*Virola sebifera*) by birds. *The Auk* **98**, 88-98.
- Howe HF, Richter WM (1982) Effects of seed size on seedling size in *Virola surinamensis*; a within and between tree analysis. *Oecologia* **53**, 347-351.
- Howe HF, Schupp EW, Westley LC (1985) Early consequences of seed dispersal for a neotropical tree (*Virola surinamensis*). *Ecology* **66**, 781-791.
- Hurlbert SH (1971) The Nonconcept of Species Diversity: A Critique and Alternative Parameters. *Ecology* **52**, 577-586.
- Janzen DH (1970) Herbivores and the number of tree species in tropical forests. *The American Naturalist* **104**, 501-528.
- Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity* **101**, 92-103.
- Jones FA, Hubbell SP (2006) Demographic spatial genetic structure of the Neotropical tree, *Jacaranda copaia*. *Molecular Ecology* **15**, 3205-3217.
- Jones TH, Vaillancourt RE, Potts BM (2007) Detection and visualization of spatial genetic structure in continuous Eucalyptus globulus forest. *Molecular Ecology* **16**, 697-707.
- Klein EK, Desassis N, Oddou-Muratorio S (2008) Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. IV. Whole interindividual variance of male fecundity estimated jointly with the dispersal kernel. *Molecular Ecology* **17**, 3323-3336.
- Latouche-Hallé C, Ramboer A, Bandou E, Caron H, Kremer A (2003) Nuclear and chloroplastic genetic structure indicate fine-scale spatial dynamics in a neotropical tree population. *Heredity* **91**, 181-190.
- Luna R, Epperson BK, Oyama K (2005) Spatial genetic structure of two sympatric neotropical palms with contrasting life histories. *Heredity* **95**, 298-305.
- Murawski DA, Hamrick JL (1991) The effect of the density of flowering individuals on the mating systems of nine tropical tree species. *Heredity* **67**, 167-174.
- Nason JD, Herre EA, Hamrick JL (1998) The breeding structure of a tropical keystone plant resource. *Nature* **391**, 685-687.
- Ng K, Lee S, Saw L, Plotkin J, Koh C (2006) Spatial structure and genetic diversity of three tropical tree species with different habitat preferences within a natural forest. *Tree Genetics & Genomes* **2**, 121-131.
- Pardini EA, Hamrick JL (2008) Inferring recruitment history from spatial genetic structure within populations of the colonizing tree *Albizia julibrissin* (Fabaceae). *Molecular Ecology* **17**, 2865-2879.
- Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* **12**, 844-855.
- Piotti A, Leonardi S, Piovani P, Scalfi M, Menozzi P (2009) Spruce colonization at treeline: where do those seeds come from? *Heredity* **103**, 136-145.
- Russo SE (2003) Responses of dispersal agents to tree and fruit traits in *Virola calophylla* (Myristicaceae): implications for selection. *Oecologia* **136**, 80-87.
- Russo SE, Augspurger CK (2004) Aggregated seed dispersal by spider monkeys limits recruitment to clumped patterns in *Virola calophylla*. *Ecology Letters* **7**, 1058-1067.

- Scotti I, Gugerli F, Pastorelli R, Sebastiani F, Vendramin GG (2008) Maternally and paternally inherited molecular markers elucidate population patterns and inferred dispersal processes on a small scale within a subalpine stand of Norway spruce (*Picea abies* [L.] Karst.). *Forest Ecology and Management* **255**, 3806-3812.
- Sezen UU, Chazdon RL, Holsinger KE (2005) Genetic consequences of tropical second-growth forest regeneration. *Science* **307**, 891.
- Sezen UU, Chazdon RL, Holsinger KE (2007) Multiregional genetic analysis of tropical secondary regeneration in a canopy palm. *Ecology* **88**, 3065-3075.
- Ueno S, Tomaru N, Yoshimaru H, Manabe T, Yamamoto S (2000) Genetic structure of *Camellia japonica* L. in an old-growth evergreen forest, Tsushima, Japan. *Molecular Ecology* **9**, 647-656.
- Ueno S, Tomaru N, Yoshimaru H, Manabe T, Yamamoto S (2002) Size-class differences in genetic structure and individual distribution of *Camellia japonica* L. in a Japanese old-growth evergreen forest. *Heredity* **89**, 120-126.
- Ueno S, Tomaru N, Yoshimaru H, Manabe T, Yamamoto S (2006) Effects of canopy gaps on the genetic structure of *Camellia japonica* saplings in a Japanese old-growth evergreen forest. *Heredity* **96**, 304-310.
- Valbuena-Carabana M, Gonzalez-Martinez SC, Hardy OJ, Gil L (2007) Fine-scale spatial genetic structure in mixed oak stands with different levels of hybridization. *Molecular Ecology* **16**, 1207-1219.
- Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology* **13**, 921-935.
- Vos P, Hogers R, Bleeker M, *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acid Research* **23**, 4407-4414.
- White GM, Boshier DH, Powell W (1999) Genetic variation within a fragmented population of *Swietenia humilis* Zucc. *Molecular Ecology* **8**, 1899-1909.
- White GM, Boshier DH, Powell W (2002) Increased pollen flow counteracts fragmentation in a tropical dry forest: An example from *Swietenia humilis* Zuccarini. *Proceedings of the National Academy of Sciences of the USA* **99**, 2038-2042.
- Wright S (1943) Isolation by distance. *Genetics* **28**, 114-138.

Figures and Tables

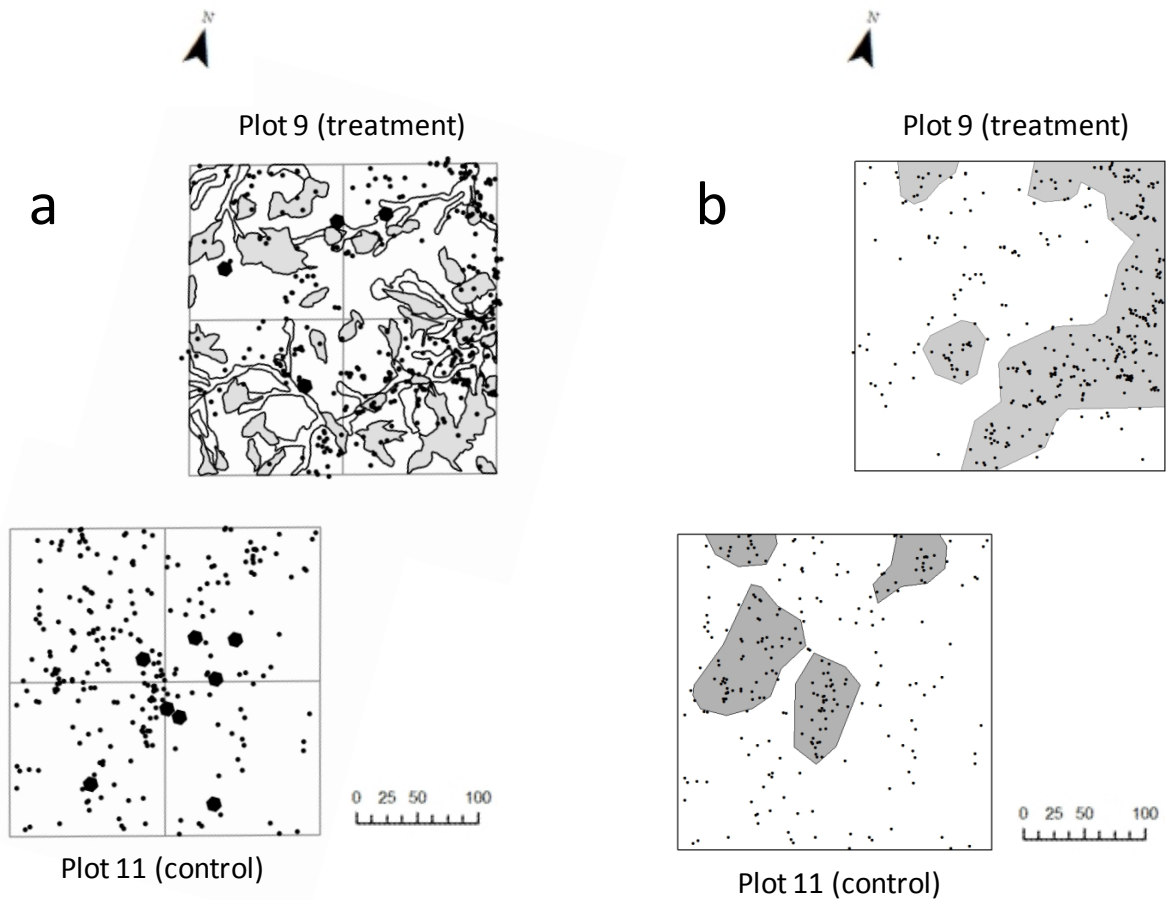


Figure 1: Distribution of mature trees and saplings in the two study plots (Plot 9, logging treatment; Plot 11, control, undisturbed plot). (a) black hexagons, mature trees; black dots, saplings; anthropic perturbation zones in plot 9 are formed by canopy gaps (grey areas) and by skidding paths (white areas). (b) black dots, saplings; grey areas: high sapling density zones. White areas: low (background) sapling density zone.

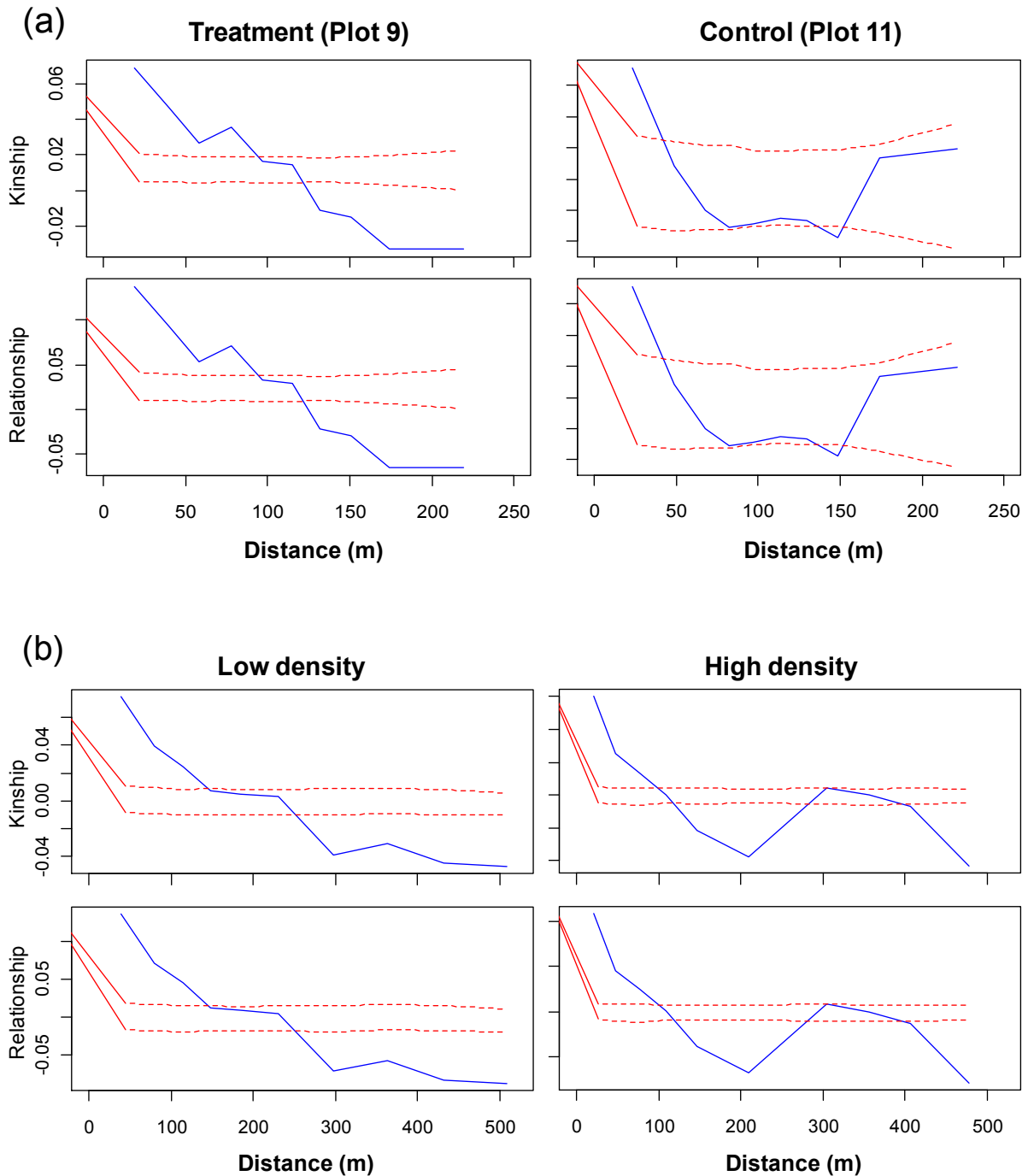


Figure 2: Spatial autocorrelation plots for sapling cohorts in Low density and High density areas and for two measures of relatedness (“Kinship” and “Relationship”, Hardy 2003). Distance: mean distance for pairs within each distance class. Dashed red lines represent the upper (97.5%) and lower (2.5%) bounds of the neutral distribution as obtained by 10000 permutations.

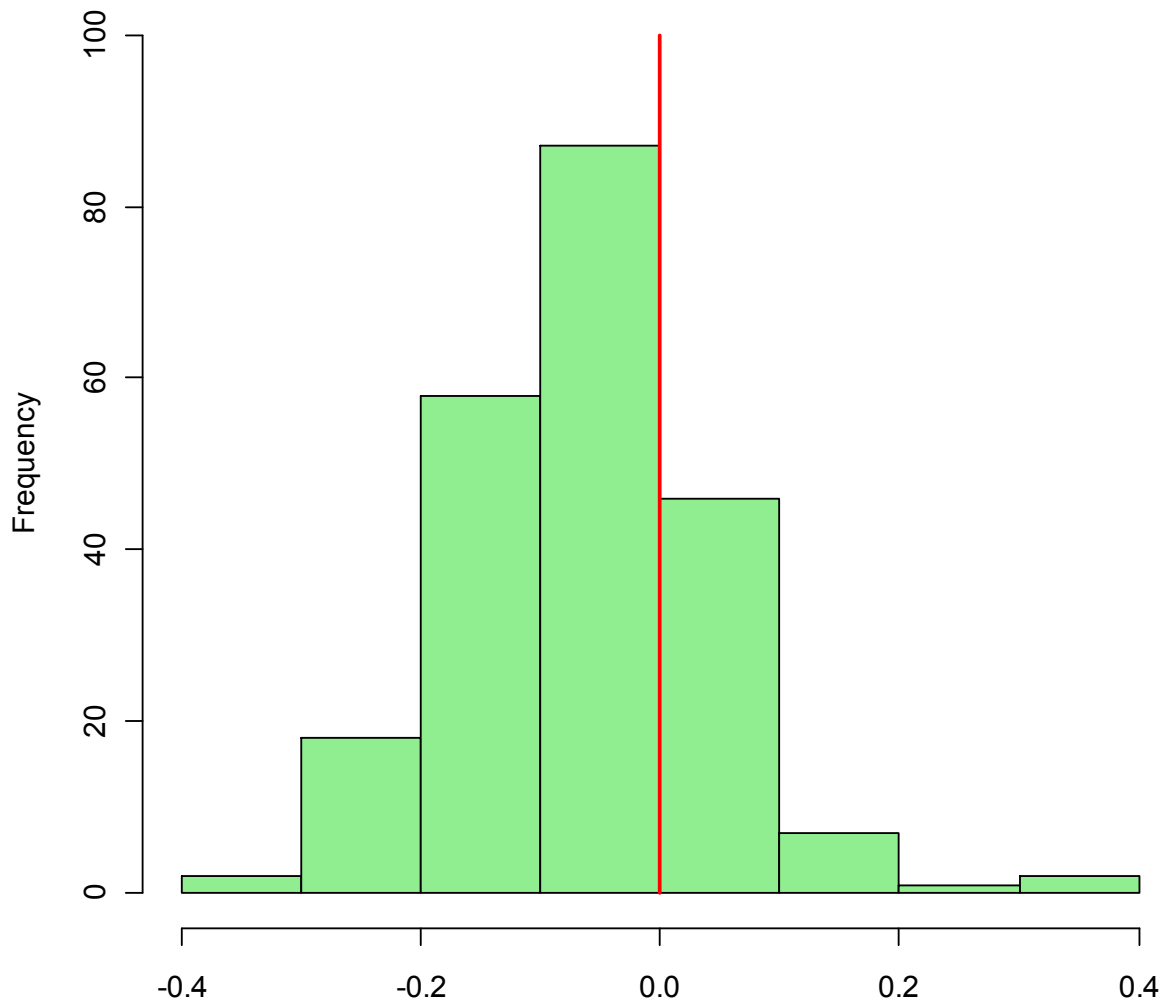


Figure 3: Distribution of ΔH between the control plot and the treatment plot for the 229 AFLP markers. The vertical bar indicates zero difference.

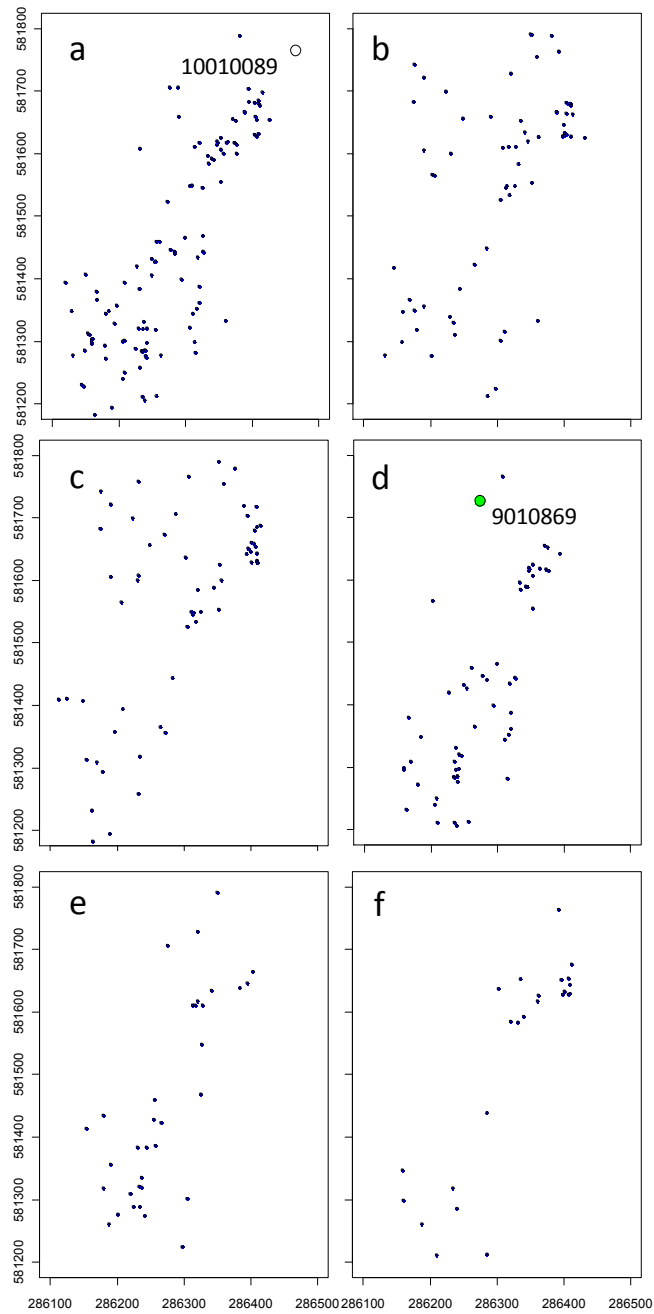


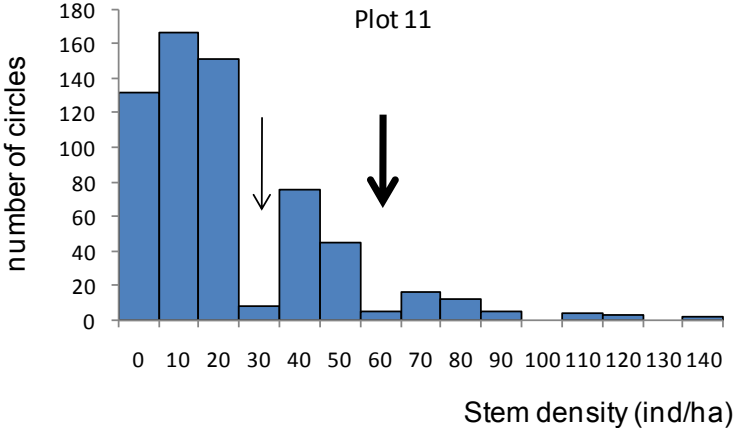
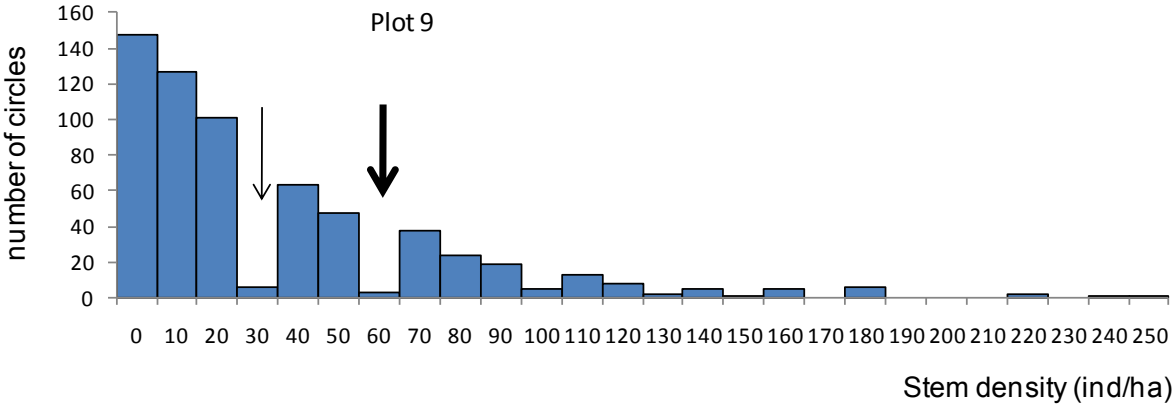
Figure 4: Distribution of seedlings derived from parent trees listed in table 1. Dots: seedlings; circles: parental trees (parental tree names are displayed next to their position; only the two trees closer than 1 Km apart from the study areas are shown). X and y axes show coordinates in the UTM system. (a) 10010089; (b) 5030500; (c) 3030733; (d) 9010869; (e) 16110208; (f) 16060457.

Table 1: Mature trees appearing in most-likely parent pairs for more than 5% of the saplings. Tree Id number: unique identification number in the Paracou reference database;

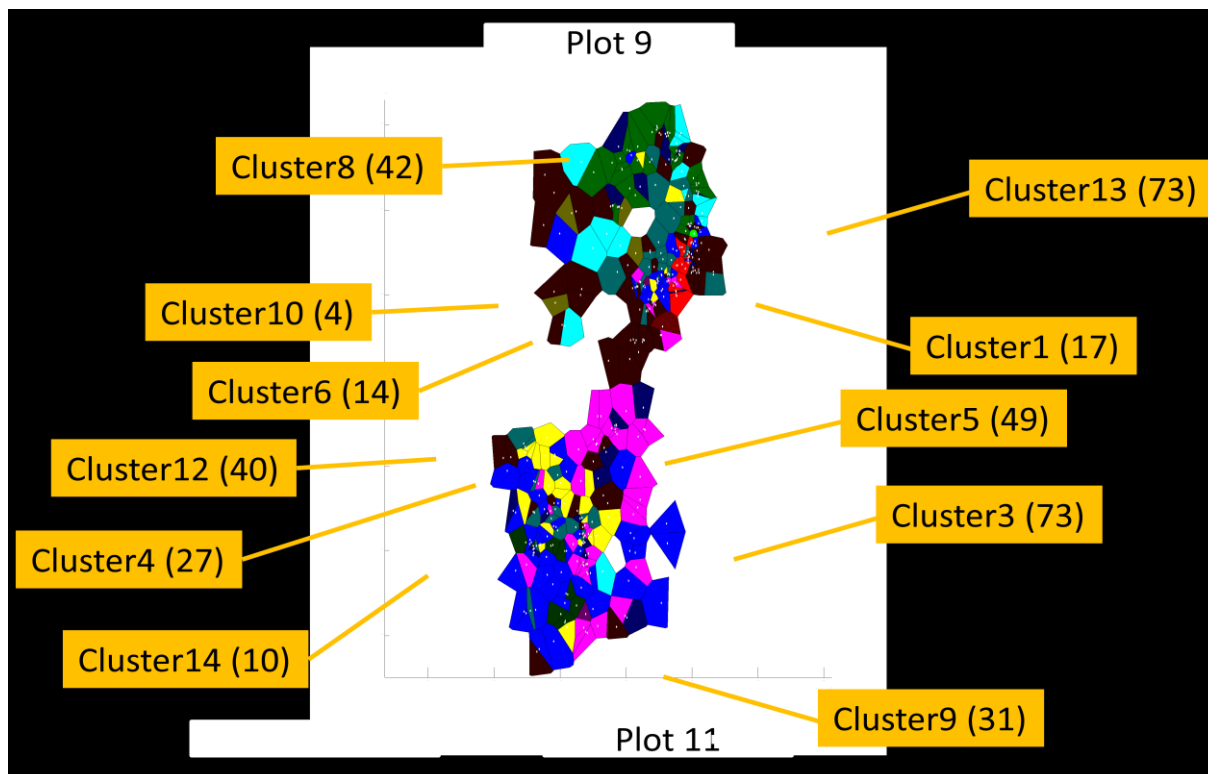
Contribution: percent of saplings for which the tree appears in the most likely parent pair;
 Sex: sex identity (nk = not known); Distance : distance from the closest border of the study area.

Tree Id number	Contribution	Sex	plot	Distance
10010089	27%	nk	10	<100 m
5030500	16%	nk	5	> 1 Km
3030733	14%	nk	3	> 1 Km
9010869	14%	Female	9	In the plots
16110208	8%	Male	16	> 1 Km
16060457	5%	Female	16	> 1 Km

Supplementary

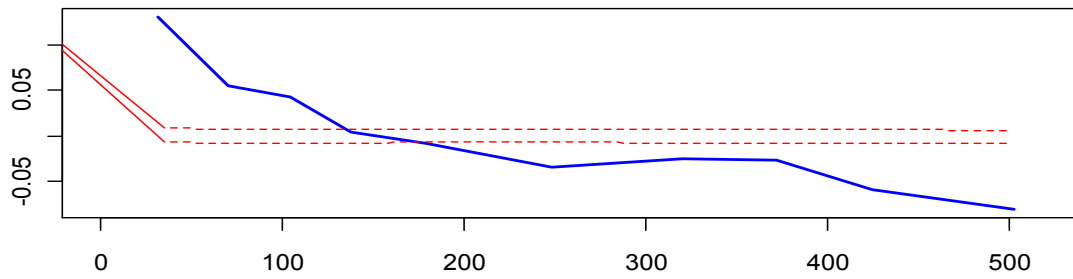


Supplementary Figure 1: Frequency histogram of areas with a given stem density for each plot for the identification of density thresholds. The thick arrow indicates the main density break in the distribution, used for the partition of low-density and high-density zones. The secondary break (not used in the partition) is indicated by the thin arrow. The first break occurs at a density of 30 stems/ha, but in this case only fewer than 50 stems are left in the low-density class; we considered this break to be an artefact.

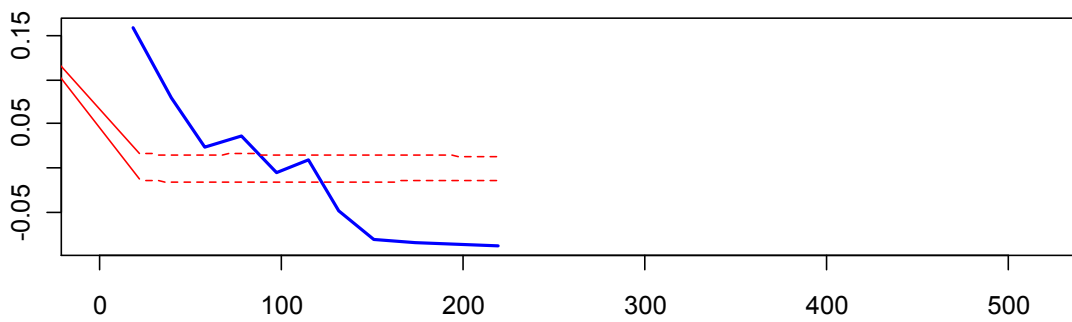


Supplementary Figure 2: Voronoi tessellation of individuals colour-coded according to genetic clusters identified by automated Bayesian assignment, obtained using BAPS with prior information on individual positions. The number of saplings belonging to each cluster is given in parentheses. Clusters 2, 7 and 11 are composed of two saplings only and are not shown.

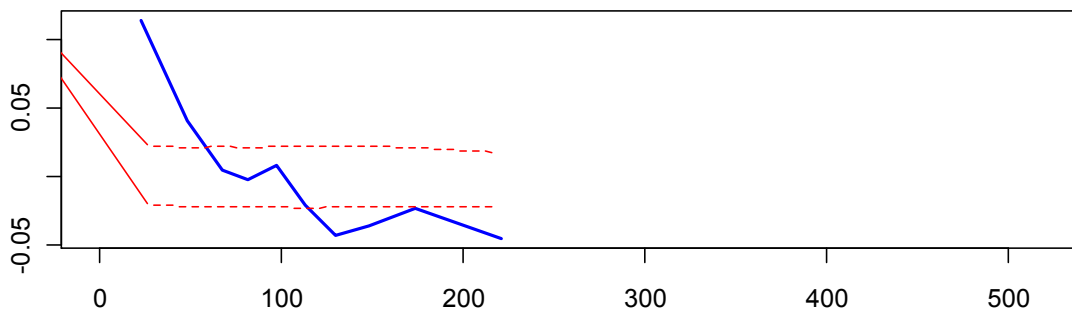
(a)



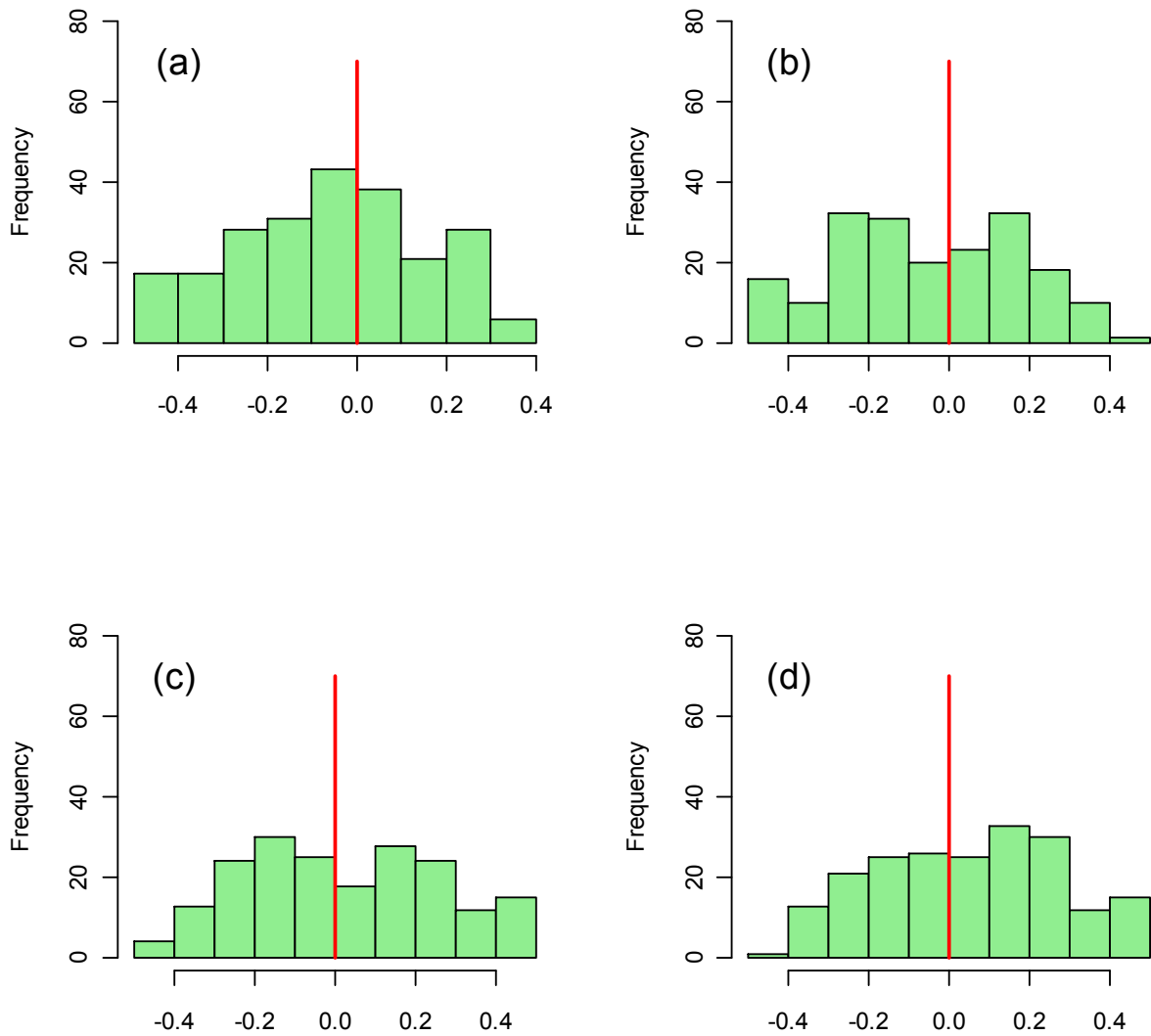
(b)



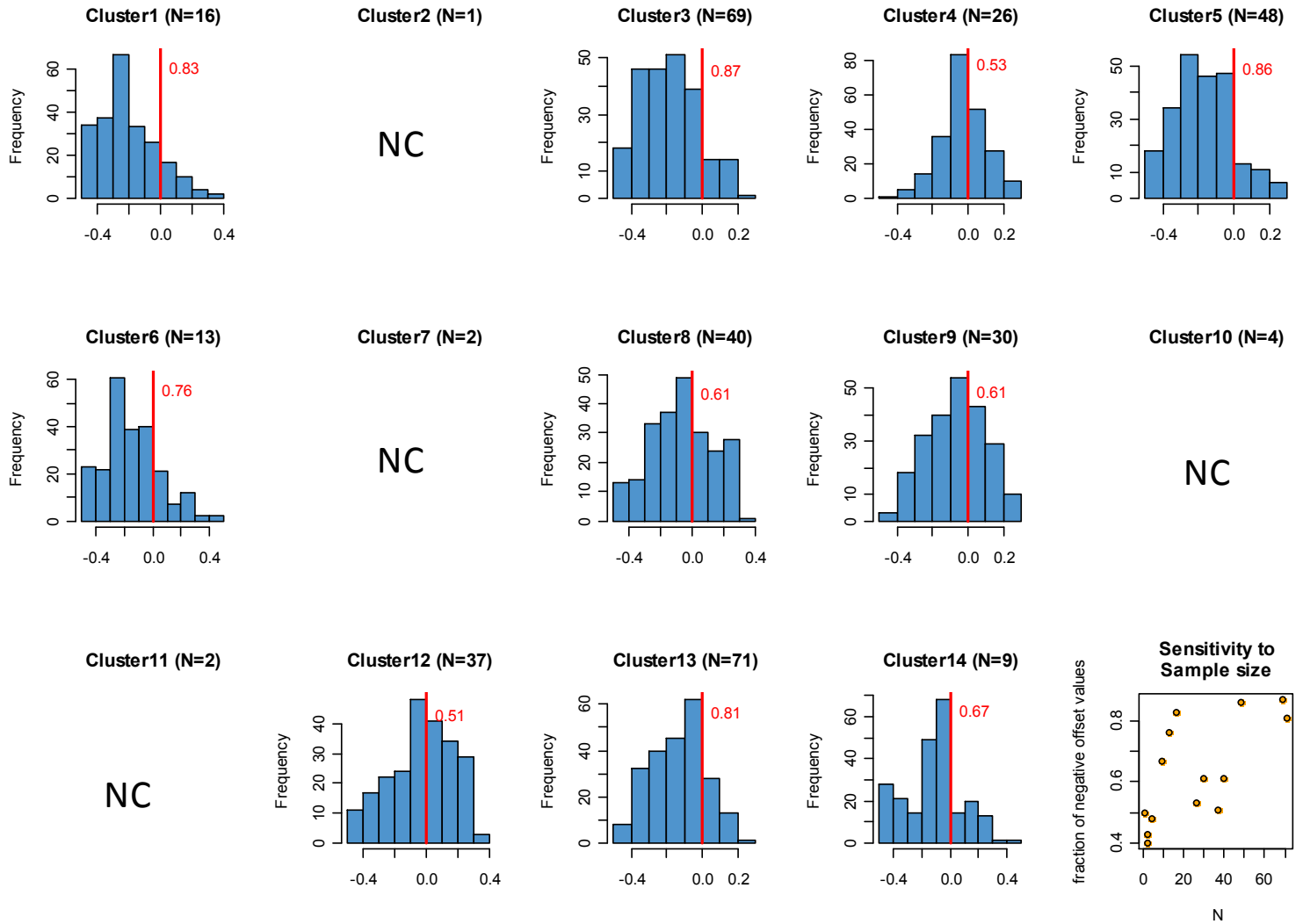
(c)



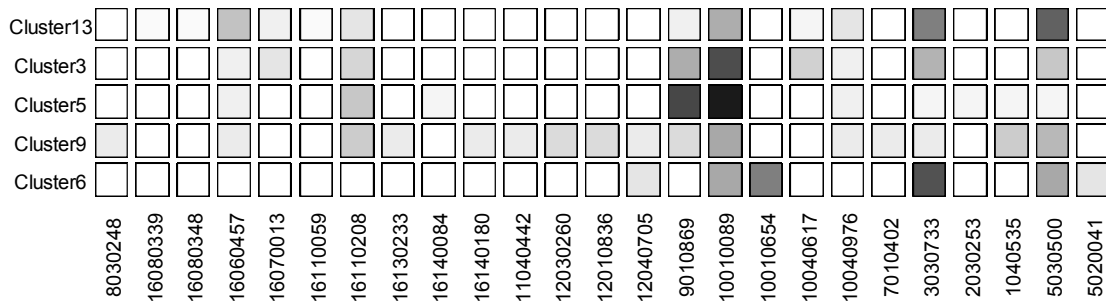
Supplementary figure 3: Spatial autocorrelation of individual trees' Bayesian cluster assignment. X axis: distance (m); Y axis: Moran's I . Patterns of spatial autocorrelation are shown for (a) the entire sapling population; (b) saplings in the treatment plot (Plot 9); (c) saplings in the control plot (Plot 11). Dashed red lines represent the upper (97.5%) and lower (2.5%) bounds of the neutral distribution as obtained by 10000 permutations.



Supplementary Figure 4: Distribution of ΔH for the 229 AFLP markers. The vertical bar indicates zero difference. Distribution of ΔH values between (a) LD and HD sapling cohorts; (b) adult and sapling cohorts; (c) adult and LD juvenile cohorts; (d) adult and HD juvenile cohorts.



Supplementary Figure 5: Distribution of ΔH between each Bayesian cluster and the global juvenile population for the 229 AFLP markers. The vertical bar indicates zero difference; the number next to the bar indicates the fraction of markers showing negative ΔH values. Lower left pane: the fraction of markers showing negative ΔH values in a given cluster as a function of cluster size. NC = not computed due to small sample size. The sensitivity of ΔH to the sample size is also given with N varying between 9 and 69.



Supplementary Figure 6: Graphical representation of the contribution of each potential parent tree within each cluster, relative to the total number of parents identified for each cluster. Grey scale corresponds to contributions from 0% (white boxes) to 50% (black boxes). Clusters are sorted from top to bottom by decreasing number of potential parents identified. Tree names are given under the form of Paracou inventory codes.

Supplementary Table 1: χ^2 values (1 degree of freedom) for RaBoT on Bayesian clusters. NC = not computed due to small sample size.

Cluster	Proportion of negative ΔH values	χ^2 value	<i>P</i> -value
1	0.83	84.4	$<10^{-4}$
2	NC	NC	NC
3	0.87	118.9	$<10^{-4}$
4	0.53	2.73	0.10
5	0.86	110.4	$<10^{-4}$
6	0.76	39.4	$<10^{-4}$
7	NC	NC	NC
8	0.61	5.35	0.02
9	0.61	5.98	0.01
10	NC	NC	NC
11	NC	NC	NC
12	0.51	0.21	0.64
13	0.81	86.8	$<10^{-4}$
14	0.67	2.31	0.13

Supplementary Table 2: Number of times each potential parent tree ('Parent') occurs for each cluster. Clusters are sorted from left to right by decreasing number of potential parents identified. Tree names are given as Paracou inventory codes. Differences in the composition of the putative parent pool were inspected for clusters for which at least twenty parents could be identified.

Parent	Cluster13	Cluster3	Cluster5	Cluster9	Cluster6
8030248	0	0	0	1	0
16080339	1	0	0	0	0
16080348	1	0	0	0	0
16060457	17	3	2	1	0
16070013	3	5	0	0	0
16110059	1	0	0	0	0
16110208	7	8	10	3	0
16130233	0	0	0	1	0
16140084	0	0	1	0	0
16140180	0	0	0	1	0
11040442	0	0	0	1	0
12030260	0	0	0	2	0
12010836	0	0	0	2	0
12040705	0	0	0	1	1
9010869	4	17	34	2	0
10010089	22	39	43	5	4
10010654	0	0	0	0	6
10040617	2	9	0	0	0
10040976	6	3	2	1	0
7010402	0	0	0	1	0
3030733	35	16	1	1	8
2030253	0	0	1	0	0
1040535	0	0	1	3	0
5030500	43	12	1	4	4

5020041 0 0 0 0 1

Supplementary Table 3: Pearson's Chi-square tests for differences in frequencies of parental contribution between Bayesian clusters of seedlings. For each pair, the first figure is the value of the test statistic, the second is the number of degrees of freedom, the third is the P-value. P-values smaller than 0.05 are in bold. Differences in the composition of the putative parent pool were inspected for clusters for which at least twenty parents could be identified.

	Cluster3	Cluster5	Cluster9	Cluster6
Cluster13	250	200	125	125
	240	192	120	120
	0.31	0.33	0.35	0.35
Cluster3		200	109	83
		80	50	50
		<0.0001	<0.0001	0.002
Cluster5			108	77
			40	40
			<0.0001	0.0003
Cluster9				55
				25
				0.0005

II. Article 2: Natural selection and within-population divergence in response to microenvironmental heterogeneity in a Neotropical nutmeg tree (Myristicaceae)

A. Résumé

Sélection divergente au niveau intraspécifique en réponse à l'hétérogénéité micro-environnementale chez *Virola michelii* (Myristicaceae)

Etudier les bases moléculaires de l'adaptation a toujours été un thème central de la biologie évolutive. L'étude de l'adaptation par la génétique chez les espèces modèles peut se faire par des approches basées sur les gènes candidats. Pour les espèces non-modèles, cette approche peut être difficile, par le manque d'information sur le génome étudié, mais il existe cependant d'autres méthodes issues des progrès de la génomique. En effet, le criblage de tout le génome (« genome scan ») peut s'avérer efficace pour la détection de loci potentiellement soumis à la sélection naturelle, sans que cela demande une caractérisation au préalable des locus détectés. Les locus neutres sont supposés avoir les mêmes niveaux de différenciation, déterminés par la dérive génétiques, les mutations et les migrations. Au contraire, on s'attend à ce que les locus soumis à la sélection (ou des loci proches de ceux-ci par l'effet auto-stop) aient des niveaux de différenciation différents de l'attendu neutre par l'effet de cette dernière. Il existe différentes méthodes pour détecter ces loci dit « outliers ». Certaines méthodes utilisent le modèle de coalescence pour simuler des loci ayant évolué sous le modèle neutre et identifier les outliers en comparant les valeurs de différenciation (F_{ST}) de chaque locus avec les valeurs simulées de F_{ST} alors que d'autres méthodes estiment directement, en utilisant une approche bayésienne, la probabilité postérieure pour un locus donnée d'être soumis à sélection. L'utilisation conjointe de plusieurs méthodes différentes peut se révéler efficace pour limiter les détections de faux positifs et des faux négatifs (erreur de type I et erreur de type II).

Virola michelii est une espèce d'arbre qui préfère les environnements plutôt secs comme le renseigne sa forte présence sur les plateaux. Cependant, nous avons remarqué que certains individus ne se situent pas dans le milieu attendu. Cette espèce constitue alors un bon modèle

pour comprendre l'adaptation, au niveau intrapopulationnel, des arbres dans des milieux contrastés comme en forêt tropicale. Ainsi, nous avons recherché si des signatures de sélection divergente, se référant à l'adaptation à l'hétérogénéité microenvironnementale, se présentaient dans le génome. De plus, si cette adaptation à fine échelle peut être signalée, il paraît intéressant de savoir à partir de quel stade de développement de l'arbre elle est détectable. En effet, les phénotypes (et donc les génotypes) inaptes à la survie dans leur milieu tendront à être éliminés par le filtrage environnemental. Ce tri des génotypes permettra de détecter les loci soumis à la sélection divergente. La détermination de la classe d'âge à partir de laquelle les filtres environnementaux ont agi peut fournir une aide précieuse pour les plans de gestion forestière.

Nous disposons pour cette étude d'un échantillonnage d'individus de différentes classes d'âge de *V. michelii*, incluant 77 adultes (d.b.h. > 200 mm) et 401 juvéniles du dispositif expérimental de Paracou. La géolocalisation de ces individus a été superposée aux différentes cartographies environnementales réalisées depuis la mise en place de ce dispositif. Ainsi, des caractéristiques microenvironnementales, liées à la disponibilité en eau et à la lumière, ont été attribuées individuellement et ont servi à la construction de sous-populations pour étudier l'adaptation locale. Des analyses par genome scan ont été effectuées par deux approches différentes avec l'approche fréquentiste DFDIST (Beaumont & Nichols 1996) et l'approche bayésienne BAYESCAN (Foll & Gaggiotti 2008). Nous avons choisis les AFLP (Amplified Length-Fragment Polymorphism) comme outil moléculaire car ils constituent un outil idéal, du fait du grand nombre de marqueurs ciblés dans tout le génome. L'adaptation au milieu a été étudiée au niveau intraspécifique à travers trois variables environnementales liées à la disponibilité en eau, (i) chez les individus adultes, (ii) chez l'ensemble des juvéniles, (iii) chez trois sous-groupes de juvéniles de différents stades de développement (classés selon leur d.b.h), et (iv) à l'intérieur de différents groupes de juvéniles regroupés selon leur proximité génétique (et donc potentiellement issus des mêmes parents) pour s'affranchir d'un éventuel effet famille. Deux autres variables, liées à la luminosité, ont également été étudiées chez les juvéniles. Trois outliers uniques ont été détectés par DFDIST, qui vont dans le sens d'une sélection divergente adaptative (selon un seuil stringent de détection à 99%) pour les trois comparaisons suivantes: topographie et engorgement du sol chez les adultes, et drainage chez les juvéniles. Bien que BAYESCAN ne détecte aucun outlier pour aucune analyse, il indique cependant que les outliers les plus probables chez les adultes sont les mêmes que ceux identifiés par DFDIST, alors qu'il n'y a pas de correspondance avec le locus identifié chez les

juvéniles. L'absence de concordance pour cette dernière comparaison pourrait résulter d'un faux positif détecté par DFDIST (la méthode bayésienne limitant le risque de détection d'erreur de type I), alors que les deux autres loci sous sélection sembleraient être de vrais outliers. Il est intéressant de remarquer que deux des trois analyses concernant les adultes ont révélé des signatures de sélection alors que ce n'est le cas pour aucune des analyses chez les juvéniles. Une des hypothèses pour expliquer cette observation est que, pour les juvéniles, le « tri » par les filtres environnementaux (qui agissent depuis le stade de graine jusqu'au stade adulte) n'a pas encore été finalisé. La présence d'individus partiellement inadaptés à leur milieu, n'ayant pas encore été éliminés par le filtrage environnemental, peut empêcher la détection de sélection naturelle. Pourtant, même pour la classe juvénile la plus âgée comprenant des individus avec un d.b.h. > 20 mm, aucun outlier n'a été détecté, laissant supposer que les filtres environnementaux peuvent agir tardivement. Cependant, il est clair que les adultes de *V. michelii* montrent des signes de sélection divergente à l'échelle locale face à des contraintes microenvironnements telles que les composantes hydriques, preuve que les espèces d'arbres s'adaptent aux caractéristiques environnementales à fine échelle. Malgré les flux de gènes au sein de la population, la force de la sélection doit être importante pour permettre la détection d'adaptation locale sur certains loci. L'adaptation locale intraspécifique s'avère un mécanisme important car il peut aboutir sur le long terme, à la spéciation écologique.

Bien que la détection de signes de sélection et d'adaptation chez des sous-populations appartenant à des milieux différents puisse être une fin en soi, il est possible d'entreprendre des études supplémentaires. En effet, les AFLP sont des marqueurs anonymes, et caractériser les loci outliers pourrait se révéler d'une grande utilité quant à l'identification de gènes candidats.

B. Article

Natural selection and within-population
divergence in response to microenvironmental
heterogeneity in a Neotropical nutmeg tree
(Myristicaceae)

William Montaigne¹, Caroline Scotti-Saintagne² & Ivan Scotti²

¹ UAG - Université des Antilles et de la Guyane, Unité Mixte de Recherche « Ecologie des Forêts de Guyane » (EcoFoG), Kourou, French Guiana

² INRA - Institut National de la Recherche Agronomique, Unité Mixte de Recherche « Ecologie des Forêts de Guyane » (EcoFoG), Kourou, French Guiana

Cet article est en phase de préparation.

Abstract

The identification of genetic loci leading to the adaptation of organisms to their environment is of great interest in the context of evolutionary biology and ecological genetics. In particular, mechanisms allowing the maintenance of within-population adaptive genetic diversity, through adaptation to local environmental variations, are still to be elucidated. We have sought signals of adaptation to microgeographic variation of environmental variables linked to water and light availability at different life stages in the tropical nutmeg tree *Virola michelii* (Myristicaceae). To do this we have performed a genome scan with 229 Amplified Length Fragment Polymorphism (AFLP) markers and two different methods for the detection of loci under divergent selection, based on coalescent simulations and on the Bayesian framework. The analyses were carried out in different cohorts of saplings, for a total of 401 individuals, and 77 mature trees, in a contiguous 15-ha area. The first method revealed, after correction for false positives, three loci under selection out of a total of 28 comparisons among levels of each environmental parameter for each life stage: one outlier for topography on adults, one for flooding on adults and one for drainage on saplings. The second approach did not detect any locus under selection, but the two loci, at which selection was detected in the adults by the coalescent-based method, turned up as marginally significant. Signals of natural selection have been therefore detected at the very fine spatial scale. Soil water logging turned out to be an important component of intraspecific local adaptation in forest trees and in particular in adults, suggesting that environmental filtering shapes patterns of heterogeneous genomic divergence reflecting local adaptation.

Keywords: Adaptive divergent selection, ecological habitat filtering, forest trees, dominant genetic markers, water microenvironmental conditions.

Introduction

There is an increasing interest in evolutionary biology in the detection of genes under natural selection, which permits to understand past evolutionary processes and to make predictions about future adaptations to environmental change. This subject is particularly interesting in long-lived sessile organisms such as forest trees, which rely on standing genetic variation to cope with selective pressure (Nielsen 2005; Neale & Kremer 2011). Recent advances in sequencing technologies, the rapid development and new statistical approaches have opened the way for the systematic search of polymorphisms of adaptive significance throughout the genome. Genome scan (Luikart *et al.* 2003) is the most popular approach to identify genomic regions under divergent selection, and is based on the simple idea that loci under divergent selection (also called outlier loci) display higher genetic differentiation than neutral loci (Lewontin & Krakauer 1973). This method is very useful to get a genomic vision of genetic divergence and to detect genomic regions experiencing selection (Barton 2000; Via 2009). In theory, all loci undergo the effects of neutral forces such as genetic drift and gene flow. Consequently, measures of divergence between populations, such as F_{ST} (Nei 1973), are expected to be the same for all neutral loci and are determined by the balance between drift, mutations and gene flow. On the contrary, loci under selection are expected to have statistical different F_{ST} (with values depending on the strength of selection); the same applies to neutral loci closely linked to loci under selection and in linkage disequilibrium with them: these loci undergo genetic hitch-hiking (Maynard Smith & Haigh 1974; Barton 2000) and thus genetic divergence. All these loci are candidates for being linked to genomic regions under positive selection, considered as divergent (high F_{ST}) or balancing (low F_{ST}) selection (see for reviews; Luikart *et al.* 2003; Beaumont 2005; Storz 2005). Genome scans present the advantage of not requiring knowledge of the DNA sequence underlying the identified loci (Beaumont & Nichols 1996; Vitalis *et al.* 2001; Beaumont & Balding 2004; Storz 2005). During the last few years, the population genomic approach via genome scanning has become accessible and particularly useful for non-models organisms (Luikart *et al.* 2003; Scotti-Saintagne *et al.* 2004; Bonin *et al.* 2006).

Genome scan approaches differ by the construction of the neutral genetic model and by the statistical approaches used to test the departure from neutral expectations. The first family of tests (Beaumont & Nichols 1996) uses a coalescence model to simulate loci under neutral evolution and identifies outlier loci by comparing the F_{ST} value of each locus with the distribution of simulated F_{ST} 's. The second family of test (Beaumont & Balding 2004; Foll *et*

al. 2008) uses Bayesian approaches to test for departures from the neutral expectation. It has been recently shown that using two (or more) outlier detection methods could be very efficient in limiting Type I and Type II errors, i.e. detection of false positives and false negatives respectively, (Oetjen & Reusch 2007; Pérez-Figueroa *et al.* 2010; Narum & Hess 2011; Tice & Carlon 2011).

Amplified fragment length polymorphisms (AFLP) are a popular type of genetic markers, ideally suited to detect loci under selection in genome scans of non-model species. The AFLP technique is a method of choice which permits to rapidly survey several hundreds of random loci covering the entire genome (Mueller & Wolfenbarger 1999; Bensch & Akesson 2005). AFLP-based genome scans can very useful in landscape genomics (Manel *et al.* 2003; Joost *et al.* 2007; Manel *et al.* 2010a; Schwartz *et al.* 2010) for the detection of correlations between genetic and landscape environmental discontinuities; such correspondences are useful to pinpoint genetic adaptation without making use of any phenotypic data. Because adaptation to variable environmental conditions can have a crucial role in speciation, interest for it is rapidly increasing (Schluter 2001; Kawecki & Ebert 2004; Schluter 2009; Via 2009). Several environmental factors have already been reported to exert selection pressures and to lead to population divergence, resulting in the detection of “outlier” loci in genome scans, for example: altitude in a common vole (Fischer *et al.* 2011) or in common frogs (Bonin *et al.* 2006), chemical pollution on teleost fish (Williams & Oleksiak 2008), freshwater versus marine habitat in a marine mammal (Chen & Yang 2009), radiation exposure in the Scots pine (Kuchma & Finkeldey 2011), insecticide treatment in mosquitos (Paris *et al.* 2010), water stress in burley (Altinkut *et al.* 2003), subtidal versus intertidal in a marine angiosperm (Oetjen & Reusch 2007) or salinity in seaweed (Coyer *et al.* 2011).

Tropical rainforests contain the highest amount of biodiversity on Earth. It is particularly true in the Amazonian Basin that is the largest and most diverse of the tropical forest wilderness areas (DaSilva *et al.* 2005) where 200–300 tree species can co-occur in a single site (Ter Steege *et al.* 2000). Several hypotheses have been proposed to explain such levels of diversity (see Leigh *et al.* 2004). In particular, the micro-geographic environmental diversity can explain tropical tree species diversity (Palmer *et al.* 2000; Phillips *et al.* 2003): influences of canopy gaps, topography and edaphic microenvironmental factors including drainage systems and soil flooding, are major components of the spatial organization of tropical trees (e.g. Sabatier *et al.* 1997; Beckage & Clark 2003; Gourlet-Fleury *et al.* 2004) because they may act as environmental filters (Poff 1997; Kotowski *et al.* 2004). Local spatial environmental

heterogeneity can therefore determine the establishment and the maintenance of intra-population diversity, and can play an important role in adaptation through diversifying selection (Lechowicz & Bell 1991; Ellis & Weis 2006; Parisod & Christin 2008). Local adaptation results from the balance between the effect of disruptive selection and the homogenizing effect of gene flow (Kawecki & Ebert 2004; Hedrick 2006; Savolainen *et al.* 2007) and can therefore allow a population to maintain itself in a contrasted environment. Natural selection has already been shown to be more effective at small geographical scales than at larger spatial scales for a mountain plant (i.e. a larger number of loci were potentially involved in adaptation at the landscape level than at the regional level), possibly showing that selective pressure can be very strong at the smaller spatial scales (Manel *et al.* 2010b). Working at the local scale has the additional advantage that populations have experienced the same history, thus limiting the effects of divergent demographic scenarios and population subdivision, which can result in the detection of numerous false-positive loci in outlier tests (Vitalis *et al.* 2001; Nielsen 2005; Butlin 2008). Despite the role of microenvironmental variation in the determination of the structure of tropical plant communities, the application of population genomic approaches (Luikart *et al.* 2003) to the detection of local adaptation is only beginning.

Early life stages (seedlings and saplings) are often the most critical phases for survival of trees (de Steven 1994; Gilbert *et al.* 2006), with mortality as large as 99% of all germinating seeds in some species (Kobe 1999). Although most saplings ultimately disappear, survivors will ultimately determine the composition of adult cohorts; seedling and sapling life-history traits are therefore major determinants of the composition of forest communities (Clark & Clark 1992). Environmental filtering may favour particular phenotypes, and underlying genotypes in a given environment through natural selection (Webb & Peart 2000; Masaki *et al.* 2005; Jabot *et al.* 2008). Ecological filtering is likely to target a subset of loci, at which subpopulations within juvenile cohorts will progressively diverge due to differential fitness in contrasting environments, thus generating heterogeneous genomic divergence (Nosil *et al.* 2008). The earliest life stage showing the same pattern of differentiation as adults in the same environment can be identified as the one having achieved the environmental filtering that shapes the genetic structure of the adult stand. Identifying this life stage is an important task to optimize forest conservation management efforts (Collins & Carson 2004).

Virola michelii (Myristicaceae) is a dioecious Neotropical canopy tree species. It is common in the Guiana Shield in primary forests as well as in secondary forests, where man-made forest disturbances generate favourable sites for its regeneration (Scotti et al, submitted.). *V. michelii* has a long history of intensive *timber exploitation* in the Guianas (Hammond *et al.* 1996). This species, classified either as a light-demanding (Delcamp *et al.* 2008) or as a shade-tolerant species (Forget 1991) has a short seed dormancy of about four to six weeks during which embryos develop prior to germination (Forget 1991). Despite the fact *V. michelii* preferentially grows on *terra firme* soils (Baraloto *et al.* 2007), it is not restricted to its preferred habitat. This species represents a good model to study local adaptation in a heterogeneous environment at the intraspecific level.

In this study, an AFLP-based population genomics approach has been used to detect outlier loci significantly correlated with environmental variables. More specifically, this landscape genomics approach aimed at detection of loci linked to genes under disruptive selection caused by ecological divergence for water and light availability conditions in different life stages (saplings and adults). Two genome scan approaches were applied (DFDIST (Beaumont & Nichols 1996) and BAYESCAN (Foll *et al.* 2008) to limit the detection of false positives.

Materials and Methods

This study was undertaken at the Paracou experimental site (French Guiana; 5°18'N, 52°55'W; <http://www.ecofog.gf/spip.php?article174>), part of the GUYAFOR plot network (<http://www.ecofog.gf/spip.php?article364>). The site's annual mean temperature is about 26°C with minor seasonal variations. An annual rainfall average of 2980 mm was recorded during the last 30-year period (Ferry *et al.* 2010). A patchwork of hills characterises the landscape (elevation is between 5 and 50 m). Soils generally have a low permeability that can lead to drainage (i.e. the removal excess water from the soil) during heavy rain. An exhaustive sampling of all *V. michelii* saplings (diameter at breast height: d.b.h. < 10 cm with a minimum height of 130 cm) in two 6.25 ha plots (plots 9 and 11) was realized in 2008. The former (plot 9) has undergone a moderate logging between 1986 and 1988 by a traditional selective felling for timber, representing about 10 trees/ha of d.b.h. > 50-60 cm, belonging to commercial species. We have also sampled all *V. michelii* with d.b.h. > 20 cm (characterized as mature adult trees) across all Paracou site plots (in a total of 120 ha). For each sapling and adult tree, a piece of cambium or leaf was collected and diameter at breast height (d.b.h.) and geographical coordinates were measured.

Total genomic DNA was extracted from biological tissues according to the hexadecetyl-trimethyl-ammonium bromide procedure (Doyle & Doyle 1987). DNA quality was checked by 1% agarose gel electrophoresis or by the 1420 Multilabel Counter Victor³_{TM} (PerkinElmer) spectrophotometre. We followed the procedure of Vos *et al.* (1995) to perform the Amplified Fragment-Length Polymorphism (AFLP) reactions: the digested DNA fragment with the restriction enzymes *EcoRI* and *MseI* were then ligated with *EcoRI* and *MseI* adapters. The pre-amplification step was performed with the *EcoRI*+A + *MseI*+A primer combination. Selective amplifications were performed using the following primer combinations: *MseI*+AAA / *EcoRI*+AAG, *MseI*+AAA / *EcoRI*+ACC, and *MseI*+AAA / *EcoRI*+ATG, with fluorescently labelled *EcoRI* primers. Amplification steps were performed using an ABI 9700 thermocycler (Applied Biosystems, USA). AFLP fragments were separated by capillary electrophoresis on an ABI 3130xl capillary sequencer (Applied Biosystems, USA) along with GeneScan-500 LIZ size standards and scored using the GENEMAPPER software as presence or absence of identified AFLP fragments. Peaks were carefully inspected by eye and only those showing a clear presence or absence signal as well as peaks with a size higher than 50 units were retained for further analyses. Individual samples were ranked according to the total number of peaks produced with all primer

combinations, and the 10% lower tail was excluded from analyses to exclude low-quality samples. The same process was applied to AFLP markers: the 10% of markers with the smallest numbers of peaks across all samples were excluded.

Data on five environmental features have been chosen from the Paracou environmental database, corresponding to soils properties, topology or silvicultural treatments, as follows:

- Topology: Three habitats can be identified relative to soil waterlogging: (i) bottomlands with a high-intensity waterlogging, (ii) hilltops with very rare water logging, and (iii) the intermediate slopes with a low-intensity waterlogging.
- Drainage (see for more details Sabatier *et al.* 1997): to describe excess or shortage of water, five classes of drainage were used to characterise soil water potential: Alt (altitude drainage), SLD (Superficial Lateral Drainage), UhS (Uphill System), DhS (Downhill System) and SH (surface hydromorphy, generally present in bottomlands).
- Soil flooding (adapted from Cantet 2004) was based on a soil colour index summarising hydromorphy levels and therefore frequency of flooding. The index ranged from 1 to 5 (with higher values indicating more frequent flooding).
- Logging disturbance: Saplings were assigned to a class according to their position relative to past logging disturbance areas (man-made canopy gaps). This feature can be related to light intensity.
- Density: To take account light intensity in natural perturbations too, we decided to partition *Virola* saplings into two groups, occurring at high or low density, with the former including putative natural and artificial disturbances (because disturbance enhance *Virola* regeneration through an increase in light intensity) and the latter including undisturbed, shaded areas. High and low density zones were created with the method of Goreaud and Pélissier (1999) (see Scotti *et al.* submitted, for more details).

Water and light availability maps were matched to samples' geographical positions using the ArcGis9 software package to characterize sample distribution relative to their environment and to define subpopulations. We grouped individuals according to their different environmental properties and according their life stages: adults or saplings (see table 1). Moreover we chose to classify saplings depending on their developmental stage, as represented by diameter at breast height (d.b.h.) (Lukaszkiwicz & Kosmala 2008). Three classes with approximately even sample sizes were generated: d.b.h. < 10 mm, 10 < d.b.h. < 20 mm and d.b.h. > 20 mm. In a previous study on the same genetic data, fourteen genetic clusters were defined by the clustering approach BAPS (see Scotti *et al.* submitted, for more

details) that represent the family structures. Furthermore, to avoid possible family effects that can confound the detection of local adaptation (Imbert *et al.* 1999), analyses were made inside the two larger genetic clusters ($n^{\circ}3$ and $n^{\circ}13$). For all analyses, we choose to merge neighbouring categories when sample size was lower than 8 for any one of these categories.

To detect outliers in our AFLP dataset, two approaches were used: a frequentist approach (Beaumont & Nichols 1996) and a Bayesian one (Foll *et al.* 2008). Both of them aimed at detecting loci with a higher-than-expected differentiation between demes. Because Type I and Type II errors are possible in each approach, it is recommended to use them jointly (Oetjen & Reusch 2007; Pérez-Figueroa *et al.* 2010; Narum & Hess 2011; Tice & Carlon 2011). First, we used MCHEZA (<http://popgen.eu/soft/mcheza/>, Antao & Beaumont 2011), a workbench that applies the F_{ST} -outlier method for detecting selection by implementing the DFDIST package's algorithm (Beaumont & Nichols 1996) and corrects common analytical errors when scanning for loci under selection. DFDIST is a modification of the FDIST software (Beaumont & Nichols 1996) for dominant markers which implements the method of Zhivotovsky (1999) to estimate allele frequencies (this software was, in addition, used to estimate F_{ST} values at AFLP loci and global F_{ST}). Firstly, Weir and Cockerham's F_{ST} (1984) is estimated at each locus. Then, the model generates coalescent simulations assuming Hardy-Weinberg equilibrium (under a null hypothesis of neutral genetic drift with migration between populations) to generate a null sampling distribution of estimates of F_{ST} based upon neutral expectations. The simulated F_{ST} data are used to identify loci that may not fit the neutral migration-drift simulations (i.e. outliers) given their exceedingly low or high F_{ST} values. To limit the detection of false positives, we chose to keep only loci with the probability $P(\text{simulated } F_{ST} < \text{sample } F_{ST}) = 1$ as true outliers for disruptive selection, and those with $P(\text{simulated } F_{ST} < \text{sample } F_{ST}) = 0$ as true outliers for balancing selection. MCHEZA was run with the following parameters (see user manual for more details): 50.000 simulations to obtain the 99% confidence level, false discovery rate = 0.1, sample size = 0 for an automatic computing. We also chose to use the force mean F_{ST} in order to approximate average simulated F_{ST} to the average value found in the real dataset, and the neutral mean F_{ST} in order to remove a first candidate subset of selected loci in the computation of the neutral F_{ST} . The infinite alleles model was applied to AFLP markers and Zhivotovsky (beta a and beta b) priors (1999) were set to 0.25 as recommended by user manual. The BAYESCAN approach (<http://cmpg.unibe.ch/software/bayescan/news.html>, Foll *et al.* 2008) is based on the multinomial-Dirichlet model and estimates directly the probability that each locus is subject

to selection using a Bayesian method. BAYESCAN estimates the probability that a locus is under selection by calculating a Bayes factor (BF), which is the ratio of the posterior probabilities of two models (selection / neutral) given the data. The posterior probabilities that a locus is subject to selection [$P(\alpha \neq 0)$] are estimated using a reversible jump Markov chain Monte Carlo. We chose to set the critical posterior odds (PO) for identifying outlier markers at $PO > 3$, which corresponds to a posterior probability $P(\alpha \neq 0) > 0.76$. According to interpretation of Jeffreys (1961), a value of P above this threshold is substantial evidence of selection. Model parameters were estimated based on 20 pilot runs with a length of 50,000 iterations. The sample size was set to 50,000 and the thinning interval to 10 resulting in a total of 500,000 iterations, with a 500,000 iterations burn-in.

Results

A total of 229 AFLP markers were retained from the three primers combinations on 478 individuals (401 saplings and 77 mature trees, figure 1). A total of 28 tests for selection were performed with both the frequentist and the Bayesian approaches (table 1): three environmental parameters were tested on mature trees and on saplings genetic clusters, five on all saplings and on three classes within the saplings group (see Methods). Global F_{ST} values ranged from 0.001 to 0.084 (table 2).

The frequentist approach identified some outlier loci with significant higher differentiation than predicted under neutrality at the 99% confidence level with the probability $P(\text{simulated } F_{ST} < \text{sample } F_{ST}) = 1$: three outliers were detected for disruptive selection out of 28 combinations of sub-populations and environmental factors and no outlier were detected for balancing selection, i.e. when $P(\text{simulated } F_{ST} < \text{sample } F_{ST}) = 0$. Outlier loci were detected: for the differentiation of adult sub-populations across topography levels; for the differentiation of adult sub-populations across soil flooding and for the differentiation of sapling sub-populations across soil drainage classes (table 3 and table 4). Outliers resulting from loci showing higher differentiation than expected at the 99% confidence level are shown in figure 2 (F_{ST} plotted against heterozygosity). The F_{ST} values for these outliers ranged from 0.145 to 0.312 (table 4), much higher than global F_{ST} (table 2). No outlier loci were detected in comparisons involving saplings 'developmental stages (table 3).

In contrast, the analyses using BAYESCAN revealed no outliers with substantial evidence of selection (posterior probability $P(\alpha \neq 0) > 0.76$). Nevertheless, the loci showing the highest posterior probabilities were the same identified by DFDIST for topography and flooding soil in adults (but not for drainage in the saplings comparison; table 4). Contingency tables (table 5) show the allelic distribution (presence or absence of AFLP bands) in each environmental class for these loci. As expected, Fisher exact tests were significant (p-values < 0.01), indicating that allelic frequencies are correlated with microenvironments.

Discussion

The tests for outliers, run on 229 loci and 28 comparisons of environmental conditions, detected three loci under divergent selection, standing out against a background of low genetic differentiation (table 2). This result is an indication of local adaptation to microenvironmental heterogeneity at the landscape scale (figure 2, table 3) occurring at a very short spatial scale, well within the range of the mixing effects of gene flow and seed dispersal. The fact that loci under disruptive selection can be identified in these very stringent conditions is an indication that environmental filters are extremely strong selective forces, capable of shaping genetic within-population diversity.

These signals of adaptive selection were revealed by DFDIST, at least for some environmental comparisons, for three loci in a total of 229 at the 99% IC (table 2). Despite the great number of analyses undertaken (28 comparisons with 229 AFLP markers), only a very small number of outlier loci was detected (only one outlier detected in only three comparisons, or 0.05% of all comparisons). This is compatible with other studies performed at the intraspecific level (less than 0.5%; Jump *et al.* 2006; Miller *et al.* 2007). Conversely, other studies reported a proportion of outlier between 1% and 20% (Wilding *et al.* 2001; Altinkut *et al.* 2003; Bonin *et al.* 2006; Oetjen & Reusch 2007; Williams & Oleksiak 2008; Chen & Yang 2009; Paris *et al.* 2010; Coyer *et al.* 2011; Kuchma & Finkeldey 2011). However, our work differs from these studies at the intraspecific level because we have investigated divergent selection at the very fine scale (i.e. microenvironmental adaptation within one population) and can not be precisely compared with others studies. Most AFLP genome scan studies applied the frequentist approach, and many of them did not implement multitest correction (with some exceptions, see Pérez-Figueroa *et al.* 2010), leading to the detection of large numbers of loci under selection, many of which may be false positives. On the contrary, our results have been obtained by the application of a multitest correction based on false discovery rates (FDR) (Benjamini & Hochberg 1995), as implemented in Chiurugwi *et al.* (2011), and under a high conservative significance level, thus minimising the risk of over-estimating the proportion of loci that are under selection (e.g. Galindo *et al.* 2009). To further reduce the risk of detecting false positives, we have combined two outlier detection methods, as recently suggested (Oetjen & Reusch 2007; Pérez-Figueroa *et al.* 2010; Narum & Hess 2011; Tice & Carlon 2011).

Bayesian approaches are generally more conservative than frequentist methods (Pérez-Figueroa *et al.* 2010; Narum & Hess 2011; Tice & Carlon 2011), but they tend to suffer from

Type II errors (Narum & Hess 2011). In this study, BAYESCAN did not reveal any outlier (posterior probability $P(\alpha \neq 0) > 0.76$), but two of the three outliers detected with DFDIST had the highest posterior probability in the corresponding BAYESCAN analyses. These loci have posterior probabilities under 0.50, and thus are considered “barely worth mentioning” according to Jeffreys (1961); however, their rank in the Bayesian analysis, combined to their statistical significance in the frequentist analysis, is an indication that these two loci (number 106 and 187) are under selection in adults. On the contrary, the most plausible outlier revealed by BAYESCAN for the test at the sapling stage and drainage is not the one detected by DFDIST, suggesting that the locus n°137 detected by DFDIST is likely to be a false positive.

Signals of disruptive selection suggest that microenvironmental variation plays a major role in shaping within-population genetic diversity, particularly in relation to water stress. Previous studies have already shown the strength of selective forces leading to plant adaptation to soil water in (Kramer & Boyer 1995; Nielsen & Jørgensen 2003; Jensen & Hansen 2010). Different outliers have been detected for different aspects of plant relationships to water: whereas topography may be related to changes in soil water content (Ferry *et al.* 2010), and therefore to drought stress in one or the other environmental types, mean soil flooding is more likely to be related to root hypoxia. It would therefore not be surprising that different loci respond to the two environmental parameters (Holmes & Rice 1996). It has to be pointed out, however, that topography is strongly correlated to other factors than soil water content, such as soil nitrogen availability (Raghubanshi 1992; Tateno & Takeda 2003), which may be an important source of adaptive divergence. Some authors have confirmed the functional role of loci under selection, by testing the association between outlier and quantitative trait loci (Rogers & Bernatchez 2005). This approach can provide a better understanding of adaptation of populations in contrasted environments.

Signals of selection were detected for the mature cohort, but not for the saplings cohort. Several explanations can be suggested to explain this difference. First, the study was undertaken at different spatial scales for saplings and for adults' cohorts. Analyses on adults were realized at a larger scale across Paracou site plots (in a total of 120 ha) whereas analyses on saplings were performed on two contiguous 6.25 ha plots. We can suppose that larger scale contains more diversified and so more contrasted habitats facilitating detection of adaptation. However, distribution of saplings covers the same different microenvironments as adults, making this hypothesis poorly suitable (Table 1). The most plausible explanation can be due to environmental filters that can occur at every life stages: from the seedling to the adult stage

(de Steven 1994; Battaglia *et al.* 2000; Webb & Peart 2000; Gilbert *et al.* 2006; John *et al.* 2007). This ecological filtering, by eliminating non-adapted individuals, will progressively shape patterns of heterogeneous genomic divergence. Absence of signatures of selection in juveniles is an indication that the filtering process is incomplete in younger cohorts. This may explain why, although light intensity is a major microenvironmental factor that may induce local adaptation (Rincon & Huante 1993; Van Der Meer *et al.* 1998; Battaglia *et al.* 2000; Goulart *et al.* 2011), no outliers were detected here; these tests are only based on saplings.

In conclusion, our results show that microenvironment-related adaptive differentiation at the very local scale can occur within *V. michelii* populations, particularly for parameters linked to soil water content. This indicates that disruptive selection can act at very small spatial scale.

References

- Altinkut A, Kazan K, Gozukirmizi N (2003) AFLP marker linked to water-stress-tolerant bulks in barley (*Hordeum vulgare* L.). *Genetics and Molecular Biology* **26**, 77-82.
- Antao T, Beaumont MA (2011) Mcheza: A workbench to detect selection using dominant markers. *Bioinformatics* **27**, 1717-1718.
- Baraloto C, Morneau F, Bonal D, Blanc L, Ferry b (2007) Seasonal water stress tolerance and habitat associations within four neotropical tree genera. *Ecology* **88**, 478-489.
- Barton NH (2000) Genetic Hitchhiking. *Philosophical Transactions: Biological Sciences* **355**, 1553-1562.
- Battaglia LL, Fore SA, Sharitz RR (2000) Seedling emergence, survival and size in relation to light and water availability in two bottomland hardwood species. *Journal of Ecology* **88**, 1041-1050.
- Beaumont M (2005) Adaptation and speciation: what can F_{ST} tell us? *Trends in Ecology and Evolution* **20**, 435 - 440.
- Beaumont M, Balding D (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**, 969 - 980.
- Beaumont M, Nichols R (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London series B: Biological Sciences* **263**, 1619 - 1626.
- Beckage B, Clark JS (2003) Seedling survival and growth of three forest tree species: the role of spatial heterogeneity. *Ecology* **84**, 1849-1861.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300.
- Bensch S, Akesson M (2005) Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology* **14**, 2899-2914.
- Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution* **23**, 773-783.
- Butlin R (2008) Population genomics and speciation. *Genetica* **138**, 409-418.
- Cantet L (2004) *Prédiction de l'engorgement hydrique de surface par les cortèges floristiques en forêt tropicale humide (Guyane française)*, Université Henri Poincaré.
- Chen L, Yang G (2009) A genomic scanning using AFLP to detect candidate loci under selection in the finless porpoise (*Neophocaena phocaenoides*). *Genes & Genetic Systems* **84**, 307-313.
- Chiurugwi T, Beaumont MA, Wilkinson MJ, Battey NH (2011) Adaptive divergence and speciation among sexual and pseudoviviparous populations of *Festuca*. *Heredity* **106**, 854-861.
- Clark DA, Clark DB (1992) Life History Diversity of Canopy and Emergent Trees in a Neotropical Rain Forest. *Ecological monographs* **62**, 315-344.
- Collins RJ, Carson WP (2004) *The effects of environment and life stage on Quercus abundance in the eastern deciduous forest, USA: are sapling densities most responsive to environmental gradients?* Elsevier, Kidlington, UK.
- Coyer JA, Hoarau G, Pearson G, et al. (2011) Genomic scans detect signatures of selection along a salinity gradient in populations of the intertidal seaweed *Fucus serratus* on a 12 km scale. *Marine Genomics* **4**, 41-49.

- DaSilva JMC, Rylands AB, da Fonseca G (2005) The Fate of the Amazonian areas of endemism. *Conservation Biology* **19**, 689-694.
- de Steven D (1994) Tropical tree seedling dynamics: recruitment patterns and their population consequences for three canopy species in Panama. *Journal of Tropical Ecology* **10**, 369-383.
- Delcamp M, Gourlet-Fleury S, Flores O, Gamier E (2008) Can functional classification of tropical trees predict population dynamics after disturbance?, pp. 209-220. Blackwell Publishing Ltd.
- Doyle J, Doyle J (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* **19**, 11 - 15.
- Ellis AG, Weis AE (2006) Coexistence and differentiation of 'flowering stones': the role of local adaptation to soil microenvironment. *Journal of Ecology* **94**, 322-335.
- Ferry B, Morneau F, Bontemps J-D, Blanc L, Freycon V (2010) Higher treefall rates on slopes and waterlogged soils result in lower stand biomass and productivity in a tropical rain forest. *Journal of Ecology* **98**, 106-116.
- Fischer MC, Foll M, Excoffier L, Heckel G (2011) Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). *Molecular Ecology* **20**, 1450-1462.
- Foll M, Beaumont MA, Gaggiotti O (2008) An approximate Bayesian computation approach to overcome biases that arise when using amplified fragment length polymorphism markers to study population structure. *Genetics* **179**, 927-939.
- Forget PM (1991) Comparative recruitment patterns of 2 non-pioneer canopy tree species in French Guiana. *Oecologia* **85**, 434-439.
- Galindo J, Moran P, Rolan-Alvarez E (2009) Comparing geographical genetic differentiation between candidate and noncandidate loci for adaptation strengthens support for parallel ecological divergence in the marine snail *Littorina saxatilis*. *Molecular Ecology* **18**, 919-930.
- Gilbert B, Wright SJ, Muller-Landau HC, Kitajima K, Hernandez A (2006) Life history trade-offs in tropical trees and lianas. *Ecology* **87**, 1281-1288.
- Goreaud F, Pélissier R (1999) On explicit formulas of edge effect correction for Ripley's K-function. *Journal of Vegetation Science* **10**, 433-438.
- Goulart MF, Lovato MB, de Vasconcellos Barros F, Valladares F, Lemos-Filho JP (2011) Which extent is plasticity to light involved in the ecotypic differentiation of a tree species from savanna and forest? *Biotropica*, doi: 10.1111/j.1744-7429.2011.00760.x.
- Gourlet-Fleury S, Guehl J-M, Laroussinie O (2004) *Experimental plots : key features. Ecology and management of a Neotropical rainforest. Lessons drawn from Paracou, a long-term experimental research site in French Guiana* Elsevier, Paris.
- Hammond DS, Gourlet-Fleury S, van der Hout P, ter Steege H, Brown VK (1996) A compilation of known Guianan timber trees and the significance of their dispersal mode, seed size and taxonomic affinity to tropical rain forest management. *Forest Ecology and Management* **83**, 99-116.
- Hedrick PW (2006) Genetic polymorphism in heterogeneous environments: The age of genomics. In: *Annual Review of Ecology, Evolution, and Systematics*, pp. 67-93.
- Holmes TH, Rice KJ (1996) Patterns of growth and soil-water utilization in some exotic annuals and native perennial bunchgrasses of California. *Annals of botany* **78**, 233-243.

- Imbert E, Escarré J, Lepart J (1999) Local adaptation and non-genetic maternal effects among three populations of *Crepis sancta*. *Ecoscience* **6**, 223-229.
- Jabot F, Etienne RS, Chave J (2008) Reconciling neutral community models and environmental filtering: theory and an empirical test. *Oikos* **117**, 1308-1320.
- Jeffreys H (1961) *Theory of Probability* Clarendon Press, Oxford.
- Jensen JS, Hansen JK (2010) Genetic variation in responses to different soil water treatments in *Quercus robur* L. *Scandinavian Journal of Forest Research* **25**, 400-411.
- John R, Dalling JW, Harms KE, *et al.* (2007) Soil nutrients influence spatial distributions of tropical tree species. *Proceedings of the National Academy of Sciences* **104**, 864-869.
- Joost S, Bonin A, Bruford MW, *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology* **16**, 3955-3969.
- Jump AS, Hunt JM, Martinez-Izquierdo JA, Penuelas J (2006) Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. *Molecular Ecology* **15**, 3469-3480.
- Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters* **7**, 1225-1241.
- Kobe RK (1999) Light gradient partitioning among tropical tree species through differential seedling mortality and growth. *Ecology* **80**, 187-201.
- Kotowski W, van Diggelen R, Rapson GL (2004) Light as an environmental filter in fen vegetation. *Journal of Vegetation Science* **15**, 583-594.
- Kramer PJ, Boyer JS (1995) *Water relations of plants and soils* Academic Press, Inc.
- Kuchma O, Finkeldey R (2011) Evidence for selection in response to radiation exposure: *Pinus sylvestris* in the Chernobyl exclusion zone. *Environmental Pollution* **159**, 1606-1612.
- Lechowicz MJ, Bell G (1991) The Ecology and genetics of fitness in forest plants. II. Microspatial heterogeneity of the edaphic environment. *Journal of Ecology* **79**, 687-696.
- Leigh EG, Davidar P, Dick CW, *et al.* (2004) Why do some tropical forests have so many species of trees? *Biotropica* **36**, 447-473.
- Lewontin R, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**, 175 - 195.
- Luikart G, England P, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature reviews* **4**, 981 - 994.
- Lukaszkiwicz J, Kosmala M (2008) Determining the Age of Streetside Trees with Diameter at Breast Height-based Multifactorial Model. *Arboriculture & Urban Forestry* **34**, 137-143.
- Manel S, Joost S, Epperson BK, *et al.* (2010a) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology* **19**, 3760-3772.
- Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010b) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Molecular Ecology* **19**, 3824-3835.
- Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**, 189-197.

- Masaki T, Osumi K, Takahashi K, Hozshizaki K (2005) Seedling dynamics of *Acer mono*; and *Fagus crenata*: an environmental filter limiting their adult distributions. *Plant Ecology* **177**, 189-199.
- Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetics Research* **23**, 23-35.
- Miller NJ, Ciosi M, Sappington TW, *et al.* (2007) Genome scan of *Diabrotica virgifera virgifera* for genetic variation associated with crop rotation tolerance. *Journal of Applied Entomology* **131**, 378-385.
- Mueller UG, Wolfenbarger LL (1999) AFLP genotyping and fingerprinting. *Trends in Ecology & Evolution* **14**, 389-394.
- Narum SR, Hess JE (2011) Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources* **11**, 184-194.
- Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. *Nat Rev Genet* **12**, 111-122.
- Nei M (1973) Analysis of gene diversity in subdivided populations. *PNAS* **70**, 3321-3323.
- Nielsen CN, Jørgensen FV (2003) Phenology and diameter increment in seedlings of European beech (*Fagus sylvatica* L.) as affected by different soil water contents: variation between and within provenances. *Forest Ecology and Management* **174**, 233-249.
- Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics* **39**, 197-218.
- Nosil P, Egan SP, Funk DJ, Hoekstra H (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: 'Isolation by Adaptation' and multiple roles for divergent selection. *Evolution* **62**, 316-336.
- Oetjen K, Reusch TBH (2007) Genome scans detect consistent divergent selection among subtidal vs. intertidal populations of the marine angiosperm *Zostera marina*. *Molecular Ecology* **16**, 5156-5157.
- Palmer MW, Clark DB, Clark DA (2000) Is the number of tree species in small tropical forest plots nonrandom? **1**, 95-101.
- Paris M, Boyer S, Bonin A, *et al.* (2010) Genome scan in the mosquito *Aedes rusticus*: population structure and detection of positive selection after insecticide treatment. *Molecular Ecology* **19**, 325-337.
- Parisod C, Christin P-A (2008) Genome-wide association to fine-scale ecological heterogeneity within a continuous population of *Biscutella laevigata* (Brassicaceae). *New Phytologist* **178**, 436-447.
- Pérez-Figueroa A, Garc-Pereira MJ, Saura M, Rolan-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology* **23**, 2267-2276.
- Phillips OL, Vargas PN, Monteagudo AL, *et al.* (2003) Habitat association among Amazonian tree species: a landscape-scale approach. *Journal of Ecology* **91**, 757-775.
- Poff NL (1997) Landscape Filters and Species Traits: Towards Mechanistic Understanding and Prediction in Stream Ecology. *Journal of the North American Benthological Society* **16**, 391-409.
- Raghubanshi AS (1992) Effect of topography on selected soil properties and nitrogen mineralization in a dry tropical forest. *Soil Biology and Biochemistry* **24**, 145-150.
- Rincon E, Huante P (1993) Growth responses of tropical deciduous tree seedlings to contrasting light conditions. *Trees - Structure and Function* **7**, 202-207.

- Rogers SM, Bernatchez L (2005) FAST-TRACK: Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (*Coregonus clupeaformis*). *Molecular Ecology* **14**, 351-361.
- Sabatier D, Grimaldi M, Prevost M-F, *et al.* (1997) The influence of soil cover organization on the floristic and structural heterogeneity of a Guianan rainforest. *Plant Ecology* **131**, 81-108.
- Savolainen O, Pyhajarvi T, Knurr T (2007) Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics* **38**, 595-619.
- Schluter D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution* **16**, 372-380.
- Schluter D (2009) Evidence for ecological speciation and its alternative. *Science* **323**, 737-741.
- Schwartz MK, Luikart G, Mckelvey K (2010) *Landscape genomics: A brief perspective [chapter 9]* Springer, New York.
- Scotti-Saintagne C, Mariette S, Porth I, *et al.* (2004) Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics* **168**, 1615-1626.
- Scotti I, Montaigne W, Cseke K, Traissac S (submitted) Regeneration dynamics of a tropical rainforest tree: does intermediate disturbance maintain genetic diversity ? *Ecology*.
- Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence, pp. 671-688. Blackwell Science Ltd.
- Tateno R, Takeda H (2003) Forest structure and tree species distribution in relation to topography-mediated heterogeneity of soil nitrogen and light at the forest floor. *Ecological Research* **18**, 559-571.
- Ter Steege H, Sabatier D, Castellanos H, *et al.* (2000) An analysis of the floristic composition and diversity of Amazonian forests including those of the Guiana Shield. *Journal of Tropical Ecology* **16**, 801-828.
- Tice KA, Carlon DB (2011) Can AFLP genome scans detect small islands of differentiation? The case of shell sculpture variation in the periwinkle *Echinolittorina hawaiiensis*. *Journal of Evolutionary Biology* **24**, 1814-1825.
- Van Der Meer P, Sterck FJ, Bongers F (1998) Tree seedling performance in canopy gaps in a tropical rain forest at Nouragues, French Guiana. *Journal of Tropical Ecology* **14**, 119-137.
- Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences* **106**, 9939-9946.
- Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**, 1811 - 1823.
- Vos P, Hogers R, Bleeker M, *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acid Research* **23**, 4407-4414.
- Webb CO, Peart DR (2000) Habitat associations of trees and seedlings in a bornean rain forest. *Journal of Ecology* **88**, 464-478.
- Weir B, Cockerham C (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358 - 1370.
- Wilding C, Butlin R, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology* **14**, 611 - 619.

- Williams L, Oleksiak M (2008) Signatures of selection in natural populations adapted to chronic pollution. *BMC Evolutionary Biology* **8**, 282.
- Zhivotovsky LA (1999) Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology* **8**, 907-913.

Tables and figures

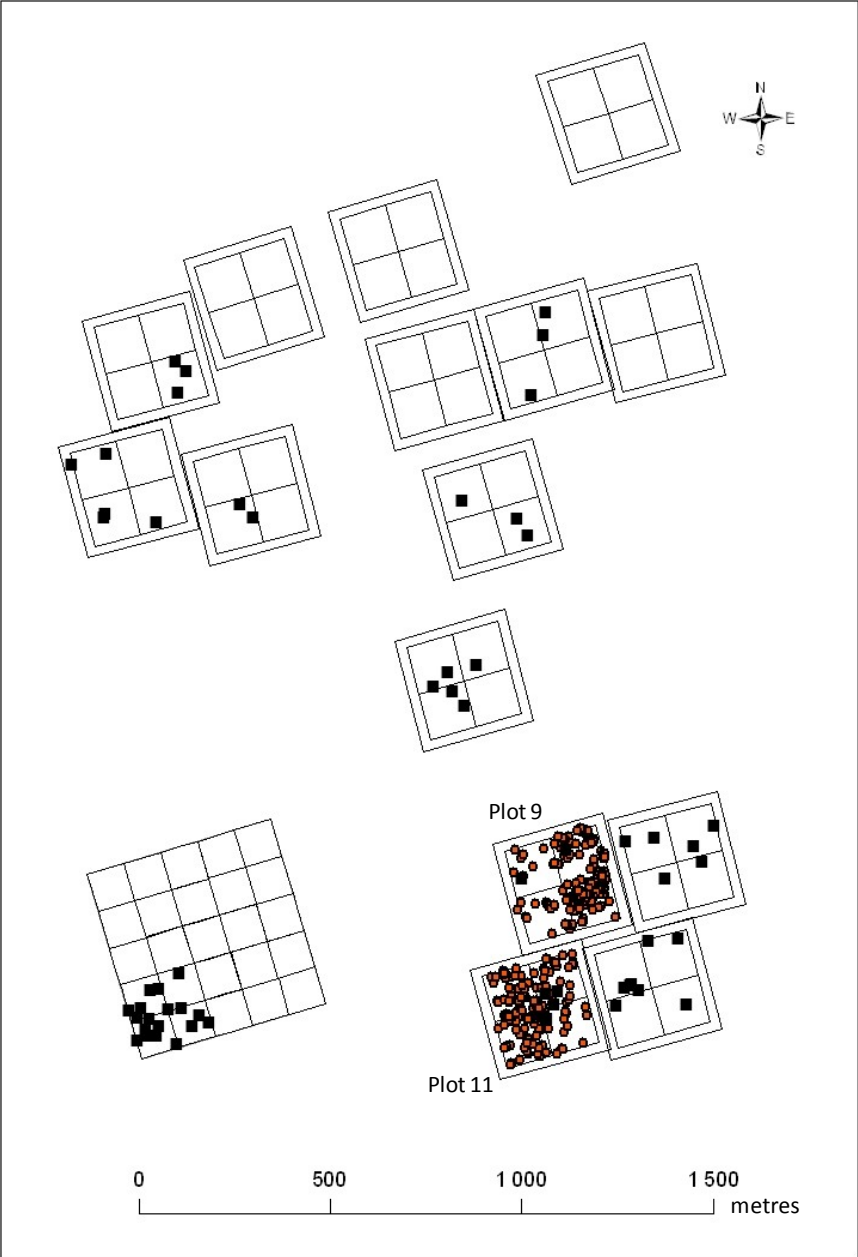


Figure 1: Distribution of *Virola michelii* samples: mature trees across all Paracou site plots are represented by black squares and saplings from plot 9 and plot 11 are represented by orange dots.

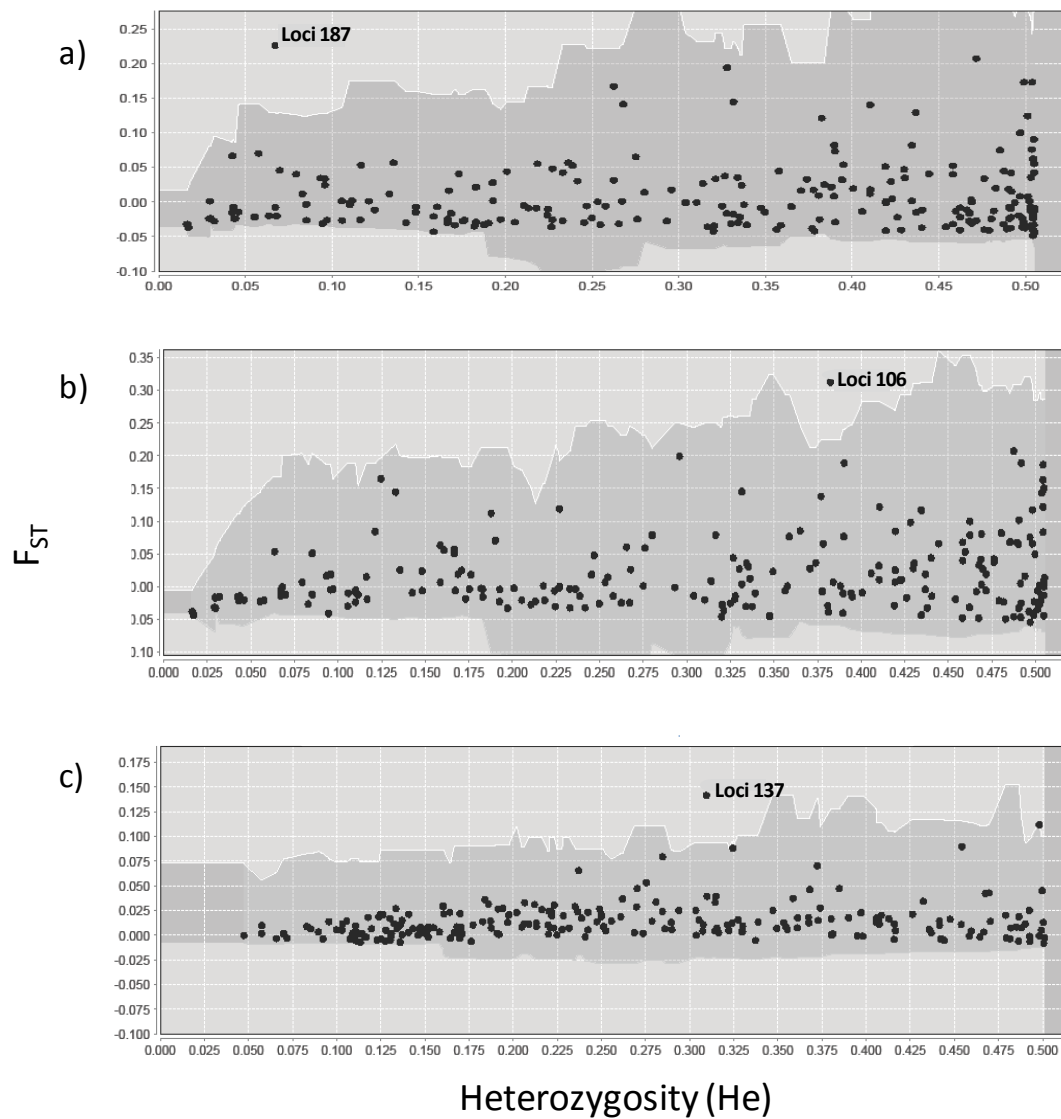


Figure 2: Outlier loci detection results from DFDIST analyses for some comparisons:

- a) adults relating to topography
- b) adults relating to flooding soil
- c) saplings relating to drainage systems

Plots show F_{ST} values against heterozygosity of the 229 AFLP loci studied. The white lines represent the 99% confidence level. Non-outlier loci are present inside the confidence level and outliers were indicated by their number identity beside their position. Only analyses revealing outlier loci were shown.

Table 1: environmental parameters (topography, flooding soil, drainage, logging canopy gaps areas and density of individuals) tested on different samples groups (adults, saplings, saplings sub-groups relating to their development stage or to their genetic cluster belonging).

A total of 28 comparisons (shown by dotted frames) are also investigated.

Numbers in table correspond to the number of individuals in each class.

		SAPLINGS			total	Cluster 3	Cluster 13	ADULTS
		< 10 mm	10 - 20mm	> 20 mm				
Topography	bottomland	2	0	1	3	2	0	8
	slope	43	40	45	128	23	19	50
	hilltop	93	95	87	275	46	54	19
Soil flooding	1	18	13	2	33	6	0	10
	2	84	97	122	303	39	65	47
	3	31	25	8	34	22	6	10
	4	4	1	1	6	2	2	2
	5	2	0	0	2	2	0	7
Drainage	Alt	14	17	8	39	4	5	12
	SLD	17	20	10	47	7	12	20
	UhS	48	62	86	196	25	44	26
	DhS	48	27	16	91	29	7	13
	SH	10	10	11	31	6	5	6
Logging canopy gaps	inside	4	10	15	29			
	outside	134	121	107	362			
Density	high	85	84	96	265			
	low	50	43	25	118			

Table 2: Global genetic differentiation F_{ST} values for each analysis.

	SAPLINGS			total	Cluster 1	Cluster 2	ADULTS
	< 10 mm	10 - 20mm	> 20 mm				
Topography	0.012	0.003	0.02	0.003	0.084	0.021	0.011
Soil flooding	0.035	0.041	0.008	0.036	0.073	0.009	0.027
Drainage	0.028	0.012	0.033	0.015	0.027	0.057	0.009
Logging canopy gaps		0.007	0.001	0.003			
Density	0.03	0.002	0.019	0.005			

Table 3: number of outliers detected for each analysis by the DFDIST approach (and the BAYESIAN approach under brackets)

	SAPLINGS			total	Cluster 1	Cluster 2	ADULTS
	< 10 mm	10 - 20mm	> 20 mm				
Topography	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0)
Soil flooding	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0)
Drainage	0 (0)	0 (0)	0 (0)	1 (0)	0 (0)	0 (0)	0 (0)
Logging canopy gaps		0 (0)	0 (0)	0 (0)			
Density	0 (0)	0 (0)	0 (0)	0 (0)			

Table 4: Identity of outlier loci revealed by DFDIST with their genetic differentiation F_{ST} values and loci with the best posterior probability $P (\alpha \neq 0)$ revealed by BAYESCAN.

$P (\text{simulated } F_{ST} < \text{sample } F_{ST}) = 1$ means a significant detection of outlier under diverging selection

$\alpha > 0$ means a diverging direction to selection but all $P (\alpha \neq 0)$ values were low (< 0.76 , from the Jeffreys' interpretation) corresponding to absence of detection of selection.

Dataset		Mcheza (Dfdist)			BayeScan		
		outlier detected	P (simulated $F_{ST} < \text{sample } F_{ST}$)	F_{ST}	"outlier" detected	P ($\alpha \neq 0$)	α
adults	Topography	187	1.00	0.226	187	0.197	0.228
	Flooding soil	106	1.00	0.312	106	0.161	0.157
saplings	Drainage	137	1.00	0.145	35	0.397	0.471

Table 5: Tests for significant association between environmental conditions and presence or absence of AFLP bands for outlier loci:

- For the outlier (n°187) detected in topography
- For the outlier (n°106) detected in soil flooding

a)

Topography	presence	absence
bottomland	4	4
slope	1	49
hilltop	0	19

b)

Soil flooding	presence	absence
1	6	3
2	10	25
3	6	0
4+5	5	4

p-value of Fisher's exact test = 0.0002 (< 0.01)

p-value of Fisher's exact test = 0.0029 (< 0.01)

C. Travaux supplémentaires hors-article

Pour cette étude, l'approche SAM ("Spatial Analysis Method", Joost *et al.* 2007) a également été testée. Le principal avantage est que cette approche se base sur les génotypes individuels. Cette méthode individu-centrée offre donc l'avantage de détecter les loci outliers adaptatifs sans avoir besoin de définir arbitrairement les populations. Cette approche est basée sur deux tests statistiques (Wald test et G-test). La présence de données manquantes dans nos jeux de données empêche la réalisation du G-test. Les individus et les loci avec des données manquantes n'ont pas été exclus des analyses car chaque individu et chaque locus contenaient au moins une donnée manquante. Des simulations aléatoires ont été testées pour remplacer ces données manquantes, mais différentes simulations pour les mêmes données aboutissaient à des résultats différents. Ne se baser que sur un seul de ces tests (Wald test) laissait apparaître beaucoup d'erreurs de Type I. Pour cette raison, cette méthode n'a pas été retenue pour l'étude de l'adaptation.

Avant ce travail à l'échelle locale chez *V. michelii*, une étude de l'adaptation avait été menée au niveau régional, avec l'utilisation d'une autre méthode : l'approche gène candidat.

Les espèces de *Virola* ont des préférences écologiques différentes notamment vis-à-vis des conditions hydriques du milieu (Baraloto *et al.* 2007). Le gradient pluviométrique constitue une variabilité hydrique environnementale spatiale qui peut entraîner l'adaptation des populations se caractérisant par des signes de divergence aux niveaux de caractères phénotypiques (Unwin & Kriedemann 1990; Cunningham *et al.* 1999; Tieleman *et al.* 2003) et par le maintien d'un niveau important de diversité génétique aux loci et pour les caractères soumis à sélection (Peleg *et al.* 2008; Eckert *et al.* 2010). L'étude de la diversité génétique d'un gène potentiellement soumis à sélection permet d'en caractériser le niveau de pression sélective et donc son importance dans le maintien des populations dans un certain milieu.

Les aquaporines sont une classe de protéines membranaires qui forment des pores perméables aux molécules d'eau dans les membranes biologiques, dont les aquaporines PiP qui sont potentiellement impliquées dans le stress hydrique (Alexandersson *et al.* 2005). Des séquences de gènes de ces aquaporines ont été isolées dans notre laboratoire, dont un gène chez *Virola* (Audigeos *et al.* 2010) : *VsePiP2.1*. L'amplification et le séquençage de ce gène ont montré du polymorphisme chez des individus de différentes espèces et différentes populations de *Virola*, échantillonnés sur différents sites guyanais aux conditions environnementales différentes par rapport à la pluviométrie (Figure a). Si ce gène montre des

signes de divergence moléculaire entre stations à l'intérieur d'une espèce, cela pourrait indiquer une adaptation des populations à l'environnement notamment par rapport au gradient pluviométrique. Au niveau interspécifique, la divergence moléculaire entre les espèces de virola pourrait résulter de l'action de la sélection divergente sur des traits impliqués dans le stress hydrique.

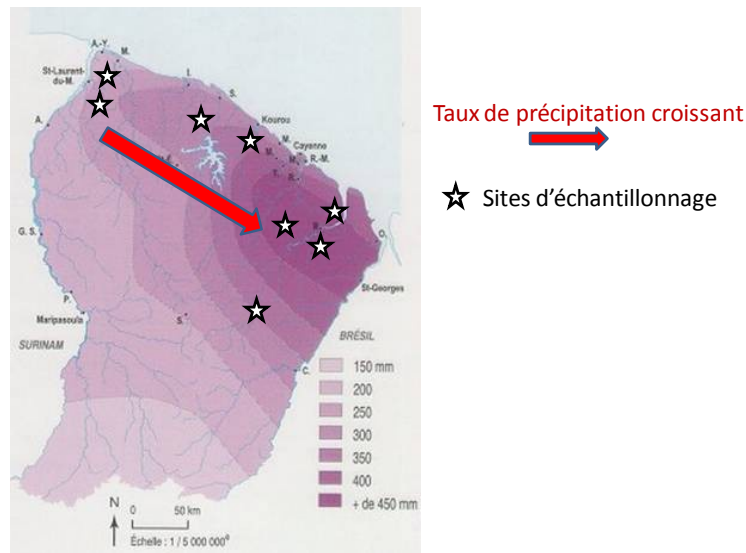


Figure a : sites d'échantillonnage et niveau de pluviométrie en Guyane française

Les résultats montrent une amplification et une lecture relativement bonnes des séquences pour ce gène chez 146 individus appartenant à 3 espèces et à 6 populations, soit 118 *V. michelii*, 25 *V. surinamensis* et 3 *V. kwatae*. L'analyse de la variance moléculaire (AMOVA, Figure b) révèle que la majeure partie de la variance génétique est expliquée par le facteur espèce mais peu par le facteur population. Il a également été observé que toutes les populations sont bien différenciées (les F_{ST} par paire de populations sont significativement différents).

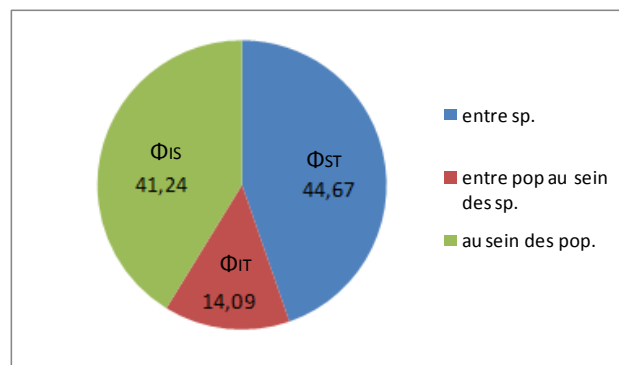


Figure b: Analyse moléculaire de la variance (AMOVA)

Le test du K_A / K_S qui compare le taux de substitution non-synonymes et le taux de substitution synonyme entre populations (ou entre espèces), au niveau des régions codantes du génome, est reconnu comme un bon indicateur de pression de sélection divergente ou purificatrice (Kimura 1980; Li *et al.* 1985; McDonald & Kreitman 1991; Yang & Nielsen 2000). Il a été entrepris au niveau interspécifique pour les trois espèces *V. michelii*, *V. surinamensis* et *V. kwatae* et également au niveau intraspécifique chez *V. michelii* (Figure c).

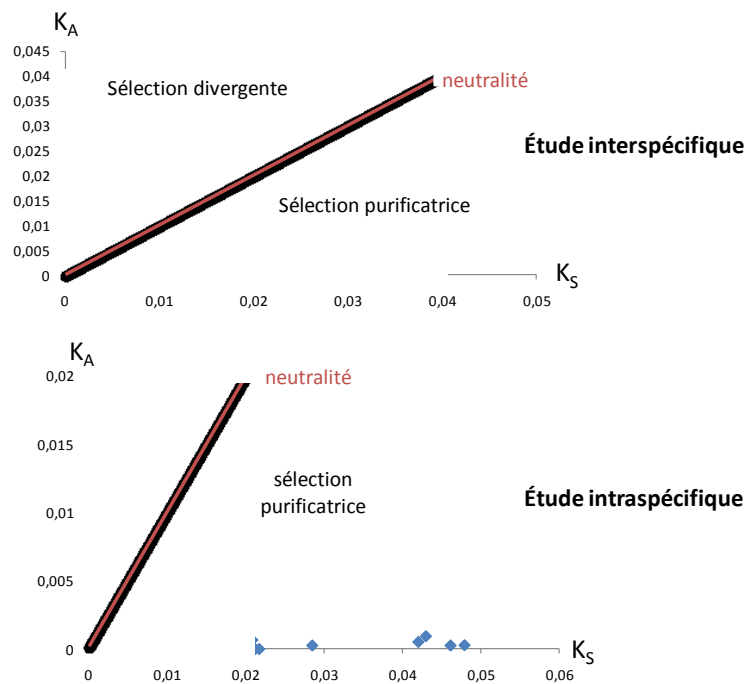
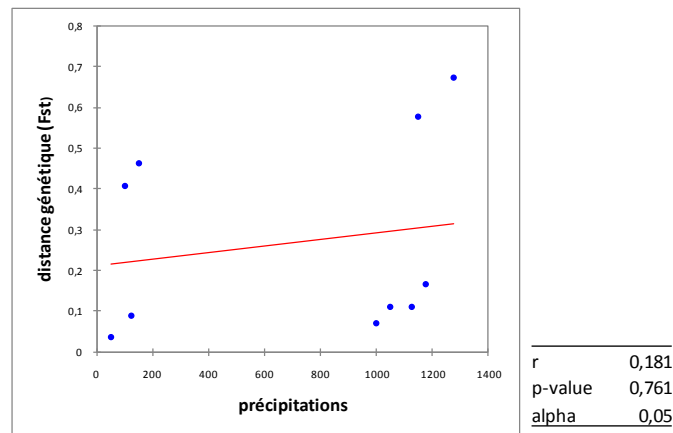


Figure c : test du K_A / K_S au niveau interspécifique (haut) et intraspécifique (bas)

Pour chaque paires d'espèces au niveau interspécifique et chaque paires de populations à l'intérieur de *V. michelii*, le ratio K_A / K_S est inférieur à 1. Ceci renseigne sur la présence d'une sélection purificatrice à la fois au niveau interspécifique et intraspécifique. Le gène est donc bien conservé, indiquant qu'il doit avoir un rôle fonctionnel important l'empêchant de trop diverger dans les parties codantes. Les allèles favorisés par la sélection sont probablement les mêmes dans toutes les populations. $K_S > K_A$ indique que la différenciation génétique s'accumulent au niveau des sites synonymes sous l'effet de la dérive génétique.

L'absence de l'effet de la sélection en réponse aux précipitations est confirmée par un test de Mantel (Mantel 1967) réalisé au niveau intraspécifique chez *V. michelii* (Figure d) qui montre l'absence de corrélation entre la différenciation génétique et les distances de précipitation.



Test non-significatif

Figure d : test de Mantel corrélant les distances génétiques aux distances environnementales des précipitations entre population de *V. michelii*

Evoquer le fait que la pluviométrie n'exerce pas de pression de sélection sur les *Virola* serait bien évidemment trop hâtif suite à ce travail préliminaire chez ce seul gène candidat. J'ai donc essayé d'isoler d'autres gènes de cette famille des aquaporines à partir des amorces universelles, mais sans succès. J'ai également voulu tester des gènes candidats dans la réponse au stress anoxique de la famille des Adh (alcool déshydrogénase, Gregerson *et al.* 1993), mais ces tests ont échoué (pas d'amplification moléculaire des amorces dégénérées). Cette étude par l'approche gènes candidats a finalement été mise de côté de part le manque d'outils moléculaires à disposition.

Par ailleurs, nous avons privilégié l'étude de l'adaptation au stress hydrique à un niveau local plutôt qu'au niveau régional. En effet, en forêt tropicale humide, il existe localement des zones partiellement voir totalement inondées lors des saisons des pluies (bas-fond), alors que d'autres zones, relativement proches, restent toujours émergées (plateau). Ces deux zones extrêmes sont affectées par des systèmes de drainage et d'hydromorphie des sols différents (Sabatier *et al.* 1997; Cantet 2004) et des individus de la même espèce peuvent être distribués dans différents environnement. Afin d'éviter une confusion d'effet entre le gradient pluviométrique à l'échelle régional et le niveau d'hydromorphie à l'échelle locale, nous avons réalisé l'étude de l'adaptation au stress hydrique à fine échelle géographique.

III. Article 3: Sisters, but not alike: ecological divergence with gene flow in a trio of Neotropical nutmeg tree species (Myristicaceae)

A. Résumé

Spéciation écologique rapide avec flux de gènes interspécifiques au sein du genre *Virola* (Myristicaceae)

La sélection naturelle peut engendrer l'accumulation de divergence génétique entre populations qui peut aboutir au processus de spéciation. En effet, selon la définition donnée au terme « espèce », l'action de la sélection naturelle, mais également celles des autres forces évolutives peut engendrer l'apparition de barrières géniques qui aura comme conséquence la création d'un isolement reproducteur aboutissant à l'apparition de nouvelles espèces. Il semble raisonnable de penser que des espèces différentes montrent un certain degré de divergence génétique. Cependant, il n'est pas toujours aisé de visualiser ce clivage génétique entre espèces, et ce, pour plusieurs raisons. D'une part, comme indiqué par Darwin dans l'Origine des espèces (1859), le processus de spéciation est un événement lent, continu dans le temps : deux espèces ayant divergé récemment montreraient alors une plus grande ressemblance génétique que deux espèces anciennes. D'autre part, deux espèces distinctes peuvent échanger leur matériel génique via des flux de gènes interspécifiques. Ces derniers auront alors la tendance à homogénéiser les pools génétiques d'espèces différentes. Il paraît alors difficile de différencier les effets des flux de gènes (s'ils existent) d'une divergence récente.

Outre les F-statistiques (indice de fixation), les méthodes de « clustering » (regroupement d'individus en « populations ») ou l'organisation de réseaux haplotypiques qui renseignent de la proximité génétique entre des populations (espèces), la théorie de coalescence apporte une aide précieuse pour comprendre les forces évolutives sous-jacentes. En effet, le coalescent permet de déterminer les mécanismes évolutifs qui sont les plus compatibles avec la

généalogie étudiée pour ainsi retracer l'histoire évolutive de ces populations. Ils existent des méthodes efficaces de calculs pour l'estimation de ces paramètres évolutifs tels que la taille efficace des populations, le temps de divergence entre populations ou les événements de migrations (flux de gènes), dont deux types d'approches seront utilisées : l'approche bayésienne d'« Isolation with Migration » (IMa2, Hey 2010) et l'approche d'ABC par simulations (SIMCOAL, Excoffier *et al.* 2000).

La famille des Myristicaceae est ancienne et est composée par une vingtaine de genres comprenant plus de 500 espèces se distribuant dans diverses niches écologiques. Malgré cette diversité spécifique et cette diversité écologique, la résolution phylogénétique au sein de cette famille est difficile et toujours incomplète du fait d'un très faible niveau de divergence génétique entre ses membres. Ces derniers constituent alors un choix idéal pour étudier les mécanismes évolutifs mis en jeu dans la distribution de tels patrons génétiques et écologiques. Plus particulièrement et préliminairement, cette étude s'est portée sur trois espèces de *Virola* du plateau des Guyanes, ayant pour certaines des préférences écologiques et des distributions géographiques différentes. En effet, *V. surinamensis* et *V. michelii* sont largement répandus (présents au moins sur tout le plateau des Guyanes) alors que *V. kwatae* ne semble être présent que localement en Guyane française et au Surinam. *V. michelii* et *V. kwatae* partagent cependant le même optimum écologique dans les zones les plus sèches de forêt (plateau), contrairement à *V. surinamensis* qui affectionne particulièrement les bas-fonds.

Trois questions sont posées dans cette étude. Les trois espèces de *Virola* ont-elles un faible niveau de divergence génétique, comme cela semble être reflété à l'échelle de la famille ? Ont-elles divergé récemment ? Leurs différences morphologiques et écologiques sont-elles suffisantes pour empêcher des flux de gènes ? Pour ce travail, des marqueurs génétiques nucléaires (10 microsatellites) et des séquences du génome chloroplastique (2 régions intergéniques) ont été employés pour une trentaine d'individus par espèce, récoltés en Guyane française. Différentes analyses ont été entreprises pour caractériser la structure génétique au sein de ce genre pour ces trois espèces (F-statistiques, « clustering » bayésien, réseau des haplotypes) et des approches de coalescence (bayésien « complet » et ABC) ont été menées pour évaluer l'impact de la démographie, des temps de divergence et des possibles flux de gènes sur cette structure génétique.

Les F-statistiques montrent des différenciations génétiques significatives entre les trois espèces, à la fois pour les marqueurs nucléaires et pour les marqueurs chloroplastiques.

Cependant, les niveaux de différenciation sont bien plus faibles entre *V. surinamensis* et *V. kwatae* que chez les deux autres comparaisons de paires d'espèces. Ce résultat se confirme, d'une part, par la méthode de clustering (à partir des données des microsatellites nucléaires) issue du logiciel STRUCTURE (Pritchard *et al.* 2000) qui indique que le nombre de groupes le plus probable $K = 3$, avec dans chaque groupe les représentants de chaque espèce (avec l'exception de trois *V. kwatae* classés dans le groupe des *V. surinamensis*). Quand on choisit de répartir les individus dans deux groupes ($K = 2$), les individus *V. kwatae* et *V. surinamensis* sont assignés dans le même groupe, alors que les *V. michelii* forme un groupe qui lui est propre. D'autre part, le réseau haplotypique dessiné, parcimonieusement, à partir du polymorphisme des séquences chloroplastiques (à l'aide du logiciel TCS, Clement *et al.* 2000), montre également la présence de deux groupes distincts : le premier est constitué des individus de *V. michelii*, et le deuxième des deux autres espèces (séparées en sous-groupes). Il ressort donc de ces analyses que *V. michelii* est bien différencié génétiquement des deux autres espèces de *Virola*, qui elles, semblent très proches. Ce dernier résultat nous interroge sur les facteurs évolutifs qui pourraient expliquer cette faible divergence génétique entre *V. surinamensis* et *V. kwatae* alors qu'elles sont fortement différenciées écologiquement et morphologiquement. Les approches par la coalescence renseignent d'un temps de divergence relativement récent (entre 0.9 et 2.7 millions d'années) pour ces trois espèces par rapport à l'âge de la famille des Myristicaceae (début de l'Eocène), la divergence la plus récente étant attribuée au couple *V. surinamensis* / *V. kwatae*. Le coalescent indique également des flux de gènes réciproques et significatifs entre *V. kwatae* et *V. surinamensis* qui tendent à homogénéiser les patrons de différenciation génétiques chez ces deux espèces formant alors un complexe d'espèces.

Les espèces, en gardant leur intégrité morphologique et écologique malgré les flux de gènes, montrent que l'action de la sélection naturelle est relativement forte pour contrer les effets antagonistes des migrations interspécifiques et qu'elle a du être un composant décisif dans le processus de spéciation. Les préférences écologiques différentes de ses espèces-sœurs associées et une divergence récente suggèrent une récente spéciation écologique. Cette notion, relativement récente et encore peu démontrée chez les arbres forestiers (sauf chez quelques espèces modèles tels que les chênes ou les peupliers), va à l'encontre du concept de la conservation des niches écologiques, déjà largement admis. Celui-ci stipule que des espèces phylogénétiquement proches auront tendance à garder les mêmes préférences écologiques. En

effet, celles-ci seraient déjà prédisposées à occuper l'habitat déjà occupé par ses ancêtres par la conservation de traits déterminant leur tolérance environnementale.

Les types de spéciation basés sur la répartition géographique des espèces (allopatrique, sympatrique et parapatricque) peuvent à la fois inclure des espèces ayant des préférences écologiques différentes et également inclure la présence de flux de gènes (à des degrés et à des temps de migration différents, voir Figure e). Il semblerait que *V. kwatae* ait divergé à partir d'une (ou plusieurs) population de *V. surinamensis* ; d'une part, les résultats sur les mesures de F_{ST} montrent un niveau de différenciation plus faible entre *V. surinamensis* et *V. michelii* qu'entre *V. kwatae* et *V. michelii* et d'autre part cette proximité accentuée peut aussi être visualisée par le réseau des chlorotypes. De plus, la répartition géographique de *V. kwatae* est restreinte ; cette espèce ne semble être présente qu'à l'échelle régionale (à l'est du plateau des Guyanes) contrairement à *V. surinamensis* qui est bien distribué en Amérique du Sud (Pérou, Brésil et jusqu'en Amérique centrale). Nos résultats sont compatibles avec un mode de spéciation sympatrique entre ces deux espèces, sans exclure la possibilité d'une spéciation allopatrique suivie d'un contact secondaire. Des analyses supplémentaires telles que des modèles de coalescent plus précis (par exemple avec une estimation du temps des flux de gènes) ou des analyses phylogénétiques, seraient nécessaires pour renseigner du type de spéciation écologique.

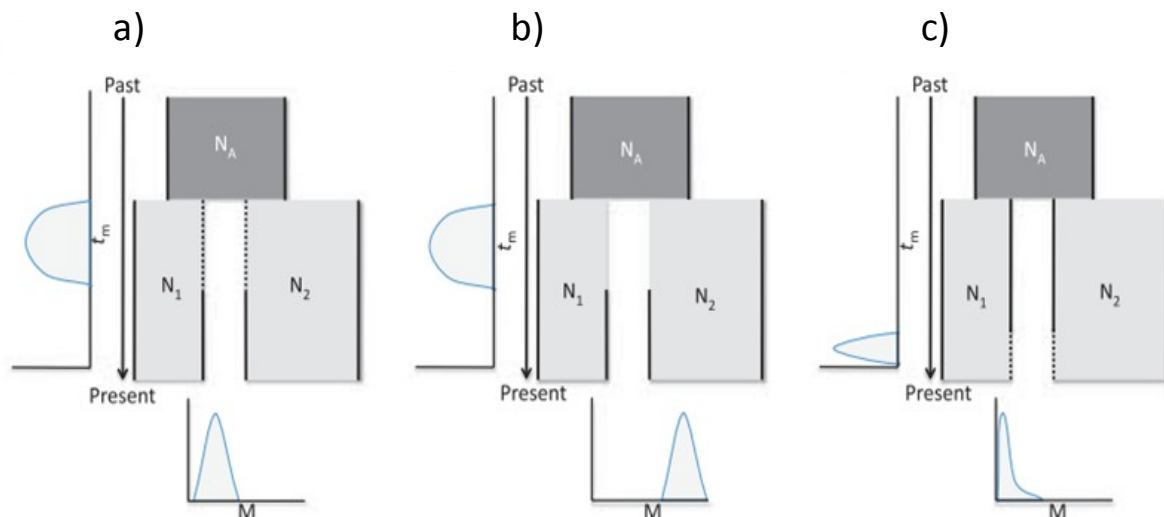


Figure e: différents modèles de spéciation autorisant les flux de gènes (d'après Gaggiotti 2011) : le gris foncé indique l'espèce ancestrale et en gris clair les espèces actuelles ayant divergées après le processus de spéciation. Les traits pleins correspondent à une barrière au flux de gènes, l'absence de traits pleins montre une forte perméabilité aux flux de gènes alors que les traits pointillés schématisent des flux de gènes occasionnels. t_m indique le temps pendant lequel il y a eu flux de gènes entre espèces (d'intensité M). a) spéciation parapatricque, b) spéciation sympatrique et c) spéciation allopatrique suivie d'un contact secondaire.

La raison du faible niveau de divergence génétique observé au sein de la famille des Myristicacées (comprenant des espèces bien diversifiées autant morphologiquement qu'écologiquement), empêchant une bonne résolution phylogénétique de la famille, vient de trouver au moins quelques éléments de réponse à travers cette étude. En effet, des spéciations récentes agrémentées de flux de gènes interspécifiques, même chez des espèces occupant des niches écologiques différentes (faisant référence à une rapide radiation écologique) vont tendre à homogénéiser les patterns génétiques de ces espèces.

B. Article

**Sisters, but not alike: ecological divergence
with gene flow in a trio of Neotropical nutmeg
tree species (Myristicaceae).**

William Montaigne¹, Daniel Sabatier², Caroline Scotti-Saintagne³ & Ivan Scotti³

¹ UAG - Université des Antilles et de la Guyane, Unité Mixte de Recherche « Ecologie des Forêts de Guyane » (EcoFoG), Kourou, French Guiana

² IRD - Institut de Recherche pour le Développement, Unité Mixte de Recherche « Botanique et Bioinformatique de l'Architecture des Plantes » (AMAP), Montpellier, France

³ INRA - Institut National de la Recherche Agronomique, Unité Mixte de Recherche « Ecologie des Forêts de Guyane » (EcoFoG), Kourou, French Guiana

Cet article est en phase de finalisation.

Abstract

Ecological divergence is an important source of functional biological diversity. Species may arise from divergent ecotypes and their phylogenetic separation may occur in spite of recent or ongoing gene flow. We have studied the genetics of ecological divergence and speciation in a trio of two widespread and one regionally common species of the *Viola* genus (Myristicaceae, the nutmeg family). Ten nuclear microsatellites and two chloroplast DNA sequences were used to infer interspecific evolutionary relationships. We have assessed the amount of divergence among the three species (widespread *V. surinamensis* and *V. michelii* and regionally distributed *V. kwatae*) and estimated their divergence time and the extent of interspecific gene flow by a mix of approaches based on conventional F-statistics, Bayesian assignment and coalescent modelling. Despite strong differences in morphology and ecological preferences, genetic data suggest that *V. surinamensis* and *V. kwatae* are very close for both nuclear and chloroplast markers and that speciation between sister-species has occurred with sustained gene flow. On the other hand, *V. michelii*, which is ecologically close to *V. kwatae*, is phylogenetically divergent from the *V. kwatae* / *V. surinamensis* pair. The large amount of shared polymorphism in this pair could be explained by recent ecological speciation, with the recently described *V. kwatae* species having derived from its widespread ancestor *V. surinamensis*.

Keywords: Interspecific gene flow, ecological speciation, *Viola* species complex, coalescent modeling

Introduction

Genetic divergence is the main mechanism underlying speciation and can be defined as the process in which two or more populations of an ancestral species accumulate independent genetic changes through time. Because different evolutionary factors can interact, studying the causes of genetic dissimilarity or similarity in natural populations may be difficult.

Ecologists are increasingly adopting in recent years, the widespread idea that closely related species are ecologically similar (Wiens et al. 2010). The tendency of species to retain ancestral ecological characteristics niche is largely predicted to occur because species inherits traits that determine their environmental tolerances from their ancestors (Wiens & Graham 2005; Losos 2008). Consequently closely related species are expected to share ecological niches, because shared evolutionary history predisposes organisms to share habitat tolerances and preferences (Webb *et al.* 2002; Cavender-Bares *et al.* 2009; Harvey & Pagel 1991). Phylogenetic niche conservatism plays the role of an “evolutionary inertia” factor that can be pitted against forces favouring ecological divergence. Indeed, ecological divergence between populations is driven by divergent natural selection between contrasting environment and is increasingly expected to play a prominent role in the process of speciation (Coyne & Orr 2004; Schluter 2009; Givnish 2010). This process is suspected to have a major role in the evolution of biological diversity and can lead to a rapid adaptive radiation (Raine & Travisano 1998; Rieseberg & Willis 2007; Futuyma 2008). Although the niche conservatism theory and the ecological divergence theory are completely opposed, both of them accept the idea that speciation events can be accompanied by gene flow (Morjan & Rieseberg 2004; Nosil *et al.* 2008). Gene flow has an important role in the niche conservatism concept, because, as mentioned by Wiens and Graham (2005) small populations at the edge of the geographic range may be flooded by individuals from the center, which may prevent these populations from adapting to environmental conditions outside the range; gene flow has then a homogenizing effect between populations. But despite that, ecological speciation can occur if natural selection is strong enough to overcome this slowing down effect (Schluter 2009). Relatively few studies have been undertaken for implication of gene flow during this ecological divergence and its consequences to adaptation and speciation (e.g. Zheng & Ge 2010)

Gene flow is an important evolutionary force that can determine the potential for genetic differentiation among populations and as well for local adaptation (Slatkin 1985; Keyghobadi

et al. 2005; Hey 2006). Natural interspecific gene flow occurs commonly in flowering plants (Mallet 2005), notably in tropical trees (Leigh *et al.* 2004), and an exceptional maintenance of species integrity is currently kept in the face it (Petit & Hampe 2006). Gene flow is an important evolutionary process contributing to the adaptive variation and speciation of organisms (Mallet 2008), to the point that genera including closely related species sharing genes through gene flow – as opposed to sharing genes inherited from their common ancestor – are identified by the special name of “species complexes”. Although, species complexes seem common in trees because they may represent a means to colonize new habitats, further works are clearly needed to estimate frequency of forests trees species complex (Petit & Hampe 2006). Molecular methods are useful for the detection of interspecific gene flow and more generally to recount the evolutionary history of species. The availability of highly polymorphic genetic markers (*e.g.* microsatellites) and powerful statistical analyses (*e.g.* Bayesian clustering) are greatly appreciated (Field *et al.* 2010). One of the most widely used approaches revealing gene flow is to combine the study of neutral nuclear and chloroplast genetic markers (McCauley et al. 1995). Chloroplast DNA has often been used for phylogenetic and phylogeographic studies because sequences can be obtained easily using universal markers and because of their relative evolutionary conservation. Plastid DNA data can be used simultaneously as a tool for inferring the population history of related taxa. Nuclear microsatellites are hypervariable and widely distributed in the genome; these properties are of particular interest for species delimitation (Hausdorf & Hennig 2010), for the detection of population genetic structure (Jarne & Lagoda 1996) and for inferring phylogenies (Richard & Thorpe 2001). In theory, because they have different evolutionary dynamics (Powell 1983), chloroplast and nuclear DNA should be able to complement each other in diverse genetic fields. Especially, genomics studies at different loci can help to describe gene flow and speciation processes (Lexer & Widmer 2008) and can help to detect hybrid individuals (Rieseberg & Soltis 1991).

Gene flow is difficult to distinguish from shared ancestral variation as both scenarios produce a similar pattern of allele sharing (Muir & Schlötterer 2005; Won & Hey 2005). In this context, the theoretical framework provided by the coalescent theory plays a major role in evolutionary investigations (Kingman 1982). The gene-genealogical inferences it produces are thought to be the best way to break apart the role of different forces having led to the currently observed patterns of molecular diversity (Rosenberg & Nordborg 2002; Bowie *et al.* 2006). Coalescent theory has enabled powerful methods to describe mechanisms of genetic

divergence between species, allowing the estimation of effective population size, population divergence time (Hey & Nielsen 2004) and population size changes (Cornuet & Luikart 1997; Beaumont 1999; Heled & Drummond 2010) as well as migration events. More particularly, an increasingly common test involving bayesian coalescent-based analyses is used to identify the role of divergence time and gene flow on levels of genetic divergence, in an ‘isolation with migration’ model (the IM approach; Hey 2006). But because evolutionary processes may be too complicated to be dealt with a full-likelihood method using bayesian MCMC as the IM approach, other techniques permitting inferences in complex demographic models, in approximating them by simulations, as Approximate Bayesian Computation (ABC), are greatly required to facilitate modelling in acting on parameters (Bertorelle et al. 2010; Csilléry et al. 2010). Contrary to the IM approach, it is possible to detect past changes in population size with ABC: it is an important fact because demographic changes in acting on effective population size can have an influence on genetic differentiation and migration measures (Lande & Barrowclough 1987).

Myristicaceae, the nutmeg family, is an ancient, widely distributed pantropical plant family, member of Magnoliales, one of the most archaic orders of flowering plants (Cronquist 1981; Stevens 2001 onwards; Doyle et al. 2004; Doyle et al. 2008). The family consists of 21 genera and up to 500 species of monoecious and dioecious trees, shrubs and lianas that hold significant ecological importance throughout lowland wet tropical forests of the world (Janovec & Neill 2002; Doyle *et al.* 2008) that exhibit a variety of habitat niches distribution (Queenborough et al. 2007). Despite its high ecological and species diversity, the resolution of relationships within this family is particularly difficult and still incomplete due to very low levels of molecular divergence (Sauquet et al. 2003). This family seems therefore a model of choice to address mechanisms underlying this pattern of diversification and to test hypothesis the role of the different evolutionary forces leading to such diversity. For that, we have studied a trio of species of the 45 species-strong Neotropical *Virola* genus (Wilson 2004): *Virola michelii* Heckel, *V. surinamensis* Warburg and *V. kwatae* Sabatier that can present differences in their biogeographical distribution, in their local environment preference and in their morphology. The two former species are widespread over Amazonia and the Guianas, whereas the latter is endemic to the eastern Guianas (Sabatier 1997). The three species show contrasting morphology (Sabatier 1997) and contrasting habitat preferences (Sabatier 1997; Baraloto et al. 2007; Macía 2011) : *V. michelii* and *V. kwatae* being *terra firme* forests specialists while *V. surinamensis* is quite exclusively found on seasonally flooded forests.

Studying mechanisms leading to such a rich taxonomical and ecological diversification in the absence of strong genetic divergence (at the family scale, as evidenced in previous studies) is an important theoretical issue, notably to understand the genesis of such diversity. Have the *Virola* species diverged recently? Are differences in morphology and habitat sufficient to maintain species boundaries? For this work, we employed different chloroplast and nuclear DNA markers to evaluate the genetic divergence of three *Virola* species present in the Guiana Shield. More specifically, we investigated the genetic parameters underlying the levels of genetic diversity and divergence of *V. michelii*, *V. surinamensis* and *V. kwatae* to evaluate the impact of demography, divergence and migration on the genus' genetic structure.

Material and methods

Samples of leaf or cambium were collected from 93 individuals of *V. michelii*, *V. surinamensis* and *V. kwatae* (32, 32 and 27 samples respectively) belonging to several populations located in the eastern Guiana Shield (7 to 11 depending on species). The botanical identification was performed using the presence or absence of several specific characters (see Sabatier 1997). No recombined character assemblages or intermediate morphology was observed. For all samples, total genomic DNA was extracted according to the hexadecyltrimethyl ammonium bromide (CTAB) procedure from Doyle & Doyle (1987) or a DNA Plant HTS© 96 kit (Invisorb) according to the manufacturer's instructions. DNA quality was analysed by spectrophotometry or by agarose gel electrophoresis.

We used two universal chloroplast sequences markers: *trnH-psbA* and *trnC-ycf6* (see e.g. Shaw *et al.* 2005). PCR reactions were performed in a 15µL reaction volume containing 10 ng of DNA, 1× ThermoPolBuffer© (New England Biolabs; 100Mm KCl, 10mM Tris-HCl at Ph 7.4, 0.1mM EDTA, 1Mm dithiothreitol), 2.5 mM MgCl₂, 0.2 mM each dNTP, 1µM of each primer and 0.5 U New England Biolabs ThermoPol© Taq DNA polymerase, using the following thermal profile: 94°C for 4 min, 35 cycles of [45 s at 94°C, 45 s at 56°C, 1.5 min s at 72°C] and a final extension at 72°C for 8 min. PCR products were purified with ExoSAP-IT (USB Corp.). Sequencing reactions was performed with BigDye Terminator© v3.1 cycle sequencing kit (Applied Biosystems). Sequencing reactions were then purified on Sephadex© columns (Millipore). Electrophoresis and detection were performed on an ABI 3130xl© capillary sequencer (Applied Biosystems). Both forward and reverse sequences were checked carefully by eye for all mutations. DNA sequences were aligned with CODONCODE ALIGNER© (CodonCode Corp.) and BIOEDIT (HALL 1999). Alignments were adjusted manually to minimize the number of gaps. Forward and reverse sequences were assembled using an R script available from the Authors upon request. A microsatellite region from the *trnH-psbA* sequence was excluded for analyses because the mutation rate of SSR is not the same as for other parts of DNA sequences.

We tested ten nuclear microsatellites (nSSRs): Vsur34, Vsur56, Vmul2-65, Vseb3, Vsur58, Vseb21, Vsur2-35, Vsur2-41, Vmul2-66 and Vmul68 drawn from Draheim *et al.* (2009). PCR reactions were performed using 20 ng DNA in a 15 mL volume containing 1× ThermoPolBuffer© (New England Biolabs; 100Mm KCl, 10mM Tris-HCl at pH 7.4, 0.1mM EDTA, 1Mm dithiothreitol), 2.5 mM MgCl₂, 0.2 mM each dNTP, 0.2µM of each primer and 1U New England Biolabs ThermoPol© Taq DNA polymerase. The PCR amplifications were

carried out on an Applied Biosystems GeneAmp PCR System 9700© or a BioRad Peltier Tetrad DNA Engine© thermal cycler, using the following protocol: 94°C for 5 min, followed by 35 cycles of [30 s at 94°C, 40 s at 52°C, 1 min at 72°C] and a final extension at 72°C for 10 min. For each primer pair, the forward primer was labeled with fluorescent dyes (NED, PET, VIC, 6-FAM). PRC products were separated and detected on an ABI 3130xl© capillary sequencer (Applied Biosystems) with Liz 500 size standard. Genotypes were scored using GENEMAPPER software.

Because no recombination occurs in chloroplast DNA, sequences of the two loci (*trnH-psbA* and *trnC-ycf6*) were assembled and treated as a single locus. Only polymorphic sites were kept for analysis. For both nuclear microsatellites and chloroplast DNA markers, allelic richness (R), expected heterozygosity (H_e) and haplotype diversity (H) were computed using ARLEQUIN 3.1 software (Excoffier et al. 2005). Pairwise F_{ST} values among the three species were calculated by ARLEQUIN 3.1 software. Species differentiation was tested by permutation ($N = 1000$) and probability values were determined according to the approach described by Excoffier et al. (1992).

To further characterize the genetic structure among species, we used the Bayesian clustering method implemented in STRUCTURE version 2.0 (Pritchard et al. 2000; <http://pritch.bsd.uchicago.edu>), to probabilistically assign individuals to K clusters, without prior on samples identity, based on the nSSR data set. Ten independent runs were performed for each K with an admixture model, 200,000 MCMC repetitions and a 50,000 burn-in period. We then selected the most plausible number of clusters K using both the approach based on $\ln P(D)$ from STRUCTURE output file (Pritchard et al. 2000) and the approach based on ΔK developed by Evanno et al. (2005). Two *V. michelii* and three *V. surinamensis* samples were discarded for clustering analyses because they had data for fewer than 6 nSSRs. We assigned individuals to a cluster if their value in the admixture coefficient was above the 0.90 threshold.

To investigate relationships among chloroplast haplotypes, we performed TCS version 1.16 (Clement et al. 2000) for the construction of the phylogenetic network. This software is based on the method of statistical parsimony described by Templeton et al. (1992).

We used coalescent approaches to determine which demographic parameters best fitted summary diversity statistics drawn from our empirical data. Inferences of population

parameters were obtained using two alternative strategies. We used the two different following methods not to compare the values of posteriors values, but by thought their differences in their method, to see if results are going to the same way as recent or ancient divergence, or presence or absence of interspecific gene flow in particular. The tree topology inferred from genetic differentiation and clustering results, was based on results of genetic differentiation between species and on clustering and networking methods.

The IMA2 program (Hey 2010) applies the Isolation with Migration model to genetic data drawn from several closely related populations or species. It implements a Markov chain Monte Carlo method (MCMC) for generating relative likelihoods/posterior probabilities for complex demographic population genetic models described by Hey and Nielsen (2007), including divergence time, migration rate, and effective population sizes. The model assumes that an ancestral population splits into two descendant populations with gene flow possibly continuing between the diverging populations. Because chloroplast capture can be common in plants, the use of cpDNA may result in biased estimates of the evolutionary history (Chan & Levin 2005), so we choose to use, for these analyses, multiple SSRs data. Since SSR data must be used as repeat numbers, to avoid errors in the estimation of repeat numbers from PCR fragment sizes only dinucleotide SSRs ($n = 7$) were used. The stepwise mutation model (SSM) and an inheritance scalar 1.0 were used for SSR loci. The following priors were used, after being adjusted by previous preliminary tests, (see Hey and Nielsen (2007) for parameters conversions): mutation rate parameter per year, uniform [10^{-9} ; 10^{-2}]; population size (in $4Nu$, where N is the effective population size and u is the mutation rate per generation): uniform [0; 500]; time of population splitting (in tu where t is the time since splitting): uniform [0; 50]; migration (in m/u , where m is migration rate per generation): exponential, mean = 1. Bayesian analyses were performed using 70 Metropolis-coupled chains with a 5 million-step burn-in followed by 3 million steps of the Markov chain. Autocorrelation values, effective sample size (ESS) and inspection of parameter trend plots by eye indicated adequate convergence of the Markov chain. For each demographic parameter, the marginal density was recorded. The peaks of the resulting distributions were taken as estimates of the parameters. For credible intervals, we assessed for each parameter the 95% highest posterior density (HPD) interval, which are the boundaries of the shortest span that includes 95% of the probability density of a parameter. The estimated divergent times (t) and effective population sizes (N) were converted to years and effective number of individuals respectively. The migration parameter corresponding to the effective migration rate at which genes are introduced into a population

per generation ($M = 2Nm$) is the product of the effective number of gene copies and the gene copy migration rate (Hey & Nielsen 2004).

In parallel, an Approximate Bayesian Computation (ABC; (for recent reviews see Bertorelle et al. 2010; Csilléry et al. 2010) was used to analyse a model with changes in population size, such as contraction or population expansion (demographic events are not implemented in IMA). Coalescent simulations were performed using the SIMCOAL2 program (Excoffier *et al.* 2000; Laval & Excoffier 2004). We simulated seven nuclear microsatellites to tally with IMA2 results. Microsatellite data were generated under a generalized stepwise mutation model (SMM). We choose to test demographic events parameters to inquire for expansion or bottleneck events. After being adjusted by previous preliminary tests, the following priors were used: mutation rate per generation, uniform [10^{-6} ; 10^{-3}]; effective population size: uniform [100; 50000]; time of population splitting in generations: uniform [5000; 300000] for the first split ($t1$) and [$t1$; 500000] for the second one; migration rate per generation: exponential, mean = 10^{-6} ; time of changes in population size: uniform [1; $t1$]; ratio between ancient and recent effective population size (related to signals of expansion or bottleneck): uniform [1; 1500]. Two kinds of summary statistics, that were calculated using ARLEQUIN 3.1: population-level heterozygosity H_e per locus and the multi-locus genetic differentiation measure F_{ST} , were chosen for the calculations of posteriors parameters because they are sensitive to variation in the migration rate between populations (Bohonak 1999). Calculations of posterior distributions of demographic events were performed using the program ABCEST2 (Excoffier et al. 2005). We performed 250,000 simulations with values randomly drawn from the priors. Parameters were estimated on the 1000 (0.4%) best scenarios by weighted linear regression and were transformed using a $\log(\tan)$ transformation to keep posteriors within the boundaries of the priors.

Results

All the ten nuclear microsatellites were polymorphic for all species. Allelic richness was quite similar for all species (from 14.4 for *V. kwatae* to 16.5 for the other two species). Expected heterozygosity values ranged from 0.74 for *V. michelii* to 0.87 for *V. surinamensis* (table 1). For chloroplast DNA markers, allelic richness and chloroplast diversity were quite similar, $R = 12, 9$ and 8 , and $H = 0.27, 0.25$ and 0.24 (table 1) for *V. surinamensis*, *V. kwatae* and *V. michelii* respectively.

Pairwise differentiation comparisons among species reveal that all species are significantly differentiated (table 2) for nuclear and chloroplast markers. In both cases, a larger genetic differentiation for both kinds of markers occurs between *V. michelii* and *V. surinamensis* ($F_{ST} = 0.164$ and 0.793 respectively for nSSR and cpDNA) and between *V. michelii* and *V. kwatae* (0.224 and 0.809) than *V. surinamensis* and *V. kwatae* (0.108 and 0.431).

In total, 93 *trnH-psbA + trnC-ycf6* sequences of about 600 bp in length were obtained from individuals of *V. michelii* ($n = 32$), *V. surinamensis* ($n = 32$) and *V. kwatae* ($n = 29$). Within these, 22 polymorphic sites were observed. These variants combined into 29 haplotypes (figure 1): 8 haplotypes for *V. michelii* (haplotypes 1-8), 12 for *V. surinamensis* (haplotypes 9-20) and 9 for *V. kwatae* (haplotypes 21-27). The haplotype network shows that haplotypes of each species are grouped together. No shared haplotypes are observed between these species using this combination of two cpDNA markers. Haplotypes are generally separated by one or two mutation except haplotype h8 counting 8 mutations from haplotype h9. This step clearly separates two haplogroups, one comprising haplotypes of *V. michelii* and the other one comprising haplotypes of the other two species.

The highest posterior probability is obtained for three genetic clusters according to the ΔK statistics of Evanno et al. (2005) with the highest ΔK value for $K = 3$ (figure 2a; $\Delta K=1664.65$). This result is also confirmed by the values of $\ln P(D)$ (figure 2b, with highest value = $-3681,78$). Because $K=2$ also shows a strong ΔK value, we report admixture results for $K=3$ and $K=2$ (figure 3). For $K=2$ (figure 3a), one cluster corresponds to *Virola michelii* ($Q1$) and the second to the other two species ($Q2$). When a third cluster is added ($K=3$; figure 3b), *V. surinamensis* and *V. kwatae* are grouped into two separate clusters ($Q2$ and $Q3$ respectively), with some exceptions: three individuals of *V. kwatae* being assigned to the *V. surinamensis* cluster.

The tree topology inferred for coalescent analyses was based on results of genetic differentiation between species and on clustering and networking methods is shown in figure 4. For the ABC approach, the two kinds of summary statistics used for calculations of posteriors (population-level heterozygosity H_e and multi-locus genetic differentiation measures F_{ST} , for a total of 24 summary statistics) are indicated in table 2 and in supplementary table 1.

The divergence time between *V. kwatae* and *V. surinamensis* is estimated to be 1.509 MYA (millions of years ago) with 95% HPD (highest posterior density) interval of 0.916 to 2.206 MYA with the IMA2 approach while the ABC method dates this around 22299.9 generations (95% HPD: 5316.51 – 108976). The divergence time between *V. michelii* and the most recent common ancestor of the two other species is estimated to be 1.796 MYA with 95% HPD interval of 1.132 to 2.704 (figure 5, table 3) by IMA2 and around 45389.3 generations (95% HPD: 7597.2 – 159017) with ABC approach.

Significant gene flow occurs between *V. surinamensis* and *V. kwatae* for both approaches: (i) from *V. kwatae* to *V. surinamensis*: $M = 0.9480$ with 95% HDP from 0.3846 to 1.617, (ii) from *V. surinamensis* to *V. kwatae*: $M = 0.4779$ with 95% HDP from 0.02752 to 1.051 (table 3, table 4, figure 5). No significant measure of gene flow was detected between *V. michelii* and the ancestor of *V. kwatae* and *V. surinamensis* by IMA although ABC approach tends to detect it (table 3, table 4, figure 5).

Contrary to pure Bayesian approach, one advantage of the ABC computations is to allow dealing with parameters as past demographic events. The results inform that bottlenecks were detected (ratio $N_{ancient} / N_{current} \gg 1$) and this reduction of population size for all species occurred between 8.14 to 73030.4 generations with wide values intervals for each species (table 4).

Discussion

Few years ago, specimens of *V. kwatae* were identified as *V. michelii* in the herbaria and floras. Although these two species occupy the same *terra firme* forest environment, some morphological characters (like leaf shape, lamina abaxial indumentum, fruit size and indumentums or tree and buttresses statures) allowed to distinguish them leading to the description of *V. kwatae* as a novel species (Sabatier 1997). A soon deciduous indumentum at the abaxial surface of the leaf lamina made of arbuscular trichomes versus a persistent indumentum made of stellate trichomes showed to be key features segregating *V. kwatae* from *V. michelii* and *V. surinamensis*. Moreover *V. kwatae* fruits and seeds are always significantly bigger than those of its two congeners (size coefficients of seed length: 1.3 and 1.5 compared to *V. michelii* and *V. surinamensis* respectively); a feature probably linked to its high dispersal rate by spider monkeys (Sabatier 1997). Our genetic study confirms that *V. kwatae* and *V. michelii* are two well differentiated species (significant genetic differentiation, no shared and distant chloroplast haplotypes, and distinct nuclear groups' assignation). As *V. kwatae*, *V. surinamensis* shows a genetic dissimilarity with *V. michelii*. This is not so surprising given that they differ clearly from each other in ecological distribution and in morphological characters supposing a different life history. Interestingly, despite a clear contrast in their environmental optima, *V. surinamensis* and *V. kwatae* are genetically and evolutionarily close both at nuclear and chloroplast markers. Firstly, F_{ST} values between these two species are the lower for both chloroplast and nuclear markers ($F_{ST} = 0.108$ for nSSR and 0.431 for cpDNA; table 2), comparing to other observations. As expected, because of their uniparently maternally inherited information (only dispersed by seed), chloroplast markers show greater differentiation than nuclear markers, inherited biparentally through seed and pollen (Ennos 1994; McCauley et al. 1995). This level of interspecific genetic differentiation is comparable to other species, as the well-known *Quercus* species that form a species complex, with F_{ST} values from 0.011 to 0.378 depending the nuclear microsatellites loci between *Q. petraea* and *Q. robur* (with a mean of $F_{ST} = 0.101$ on 14 nSSR) according to Neophytou (2010) showing some specific genome regions with high differentiation and so underlying the fact that results may depend on genetic markers. Focusing on the nutmeg family, Sauquet and his collaborators (2003) have found a global low genetic divergence between taxa within Myristicaceae: less than 5% of variation along non-coding sequences, but no interspecific genetic differentiation level has been reported for any genus to compare precisely our results at genus level. Nevertheless, it has already been shown that it is difficult to differentiate some

species within *Compsonera* genus (Newmaster et al. 2008) with based on some genetic markers, for an example, no variation of the *trnH-psbA* intergenic spacer was detected between species. Secondly, nSSR clustering have grouped these two species together when $K=2$, and when $K=3$ few *V. kwatae* trees are assigned in the *V. surinamensis* group (figure 3). Moreover, we observe for these few *V. kwatae* samples only two or three mutations which differentiate them from the closest *V. surinamensis* individuals based on two chloroplast sequence markers (haplotypes h27 and h21, figure2). More generally, individuals from these different species can be closer than two individuals from a same species. The observation of this genetic proximity between *V. surinamensis* and *V. kwatae* could be a consequence of recent divergence and or/gene flow between species.

Recently, coalescent-based analyses such as the full-likelihood method using bayesian MCMC ‘isolation with migration model’ or the SIMCOAL ABC approach, have emerged and have been successfully used to distinguish effects of time divergence and gene flow, as well as past demographic events for ABC methods, from levels of molecular differentiation (Hey 2006; Csilléry et al. 2010). Our study indicated divergence among the three species is relatively very recent compared to the age of the Myristicaceae family (early Eocene; Doyle et al. 2008), with *V. kwatae* and *V. surinamensis* having split most recently (figure 5, table 3, table 4; 0.916 – 2.20 MYA and 5316.51 – 108976 generations with IMA2 and SIMCOAL2 approaches respectively), and soon after the divergence of their ancestor from *V. michelii* (1.132 – 2.704 MYA and 7597.2 – 159017 generations with IMA2 and SIMCOAL2 approaches respectively). Both the chlorotype network (figure 1) and the genetic differentiation (table 2) support the idea that *V. michelii* is closer to *V. surinamensis* than to *V. kwatae*, suggesting that *V. kwatae* has diverged from a subset of *V. surinamensis* populations. The restricted distribution range shown by *V. kwatae* relative to *V. surinamensis* is consistent with this scenario. Ecological differences between these two species suggest that an ecological transition was associated to the speciation process. There is therefore the possibility that the ecological divergence was one of the forces driving the speciation process. Ecological speciation is a recent but important speciation concept that applies to cases of divergent selection between populations exploiting different resources or environments (Schluter 2001; Wu 2001; Levin 2004; Rundle & Nosil 2005; Schluter 2009). Nonetheless, this process is not well documented because of the lack of observation in nature, notably in forest trees (Lexer & Fay 2005; Roda 2011), even if some observations have already been shown in some plant model species as for oaks (Cavender-Bares & Pahlisch 2009) or poplars species (Lexer et al. 2005). In the case of *V. kwatae*, a shift ecological evolution would have

occurred from the lowland, seasonally flooded habitats, typical of *V. surinamensis*, to the *terra firme* forest environment, probably accompanied by a shift from bird to mammal seed dispersal.

Ecological speciation is likely to be of key importance in facilitating speciation with gene flow (Savolainen et al. 2006; Nosil 2008; Gaggiotti 2011) that can help populations to adapt to newly invaded environments (Seehausen 2004). No gene flow was detected by the ‘isolation with migration’ method between *V. michelii* and the ancestor of *V. surinamensis* and *V. kwatae* contrary to the ABC approach. More tests are also needed to understand this misleading. We may suppose that the chosen model used by the ABC approach was not exactly the true model. If we based these results on the promising IMA method (Marko & Hart 2011), this absence of detection may have several reasons. The first and most obvious is that the two species may not be a sister species pair (a complete phylogeny of the genus is lacking). Secondly, their divergence may have been allopatric; third, detecting migration when this happens early on in the divergence process can be difficult (Gaggiotti 2011). But, interestingly, coalescent-based methods reveal interspecific gene flow between the morphological and ecological contrasting species *V. kwatae* and *V. surinamensis* (figure 5, table 3) strongly rejecting a model of divergence with zero gene flow. Different elements can support this scenario. The flowering time occurs approximately between April and September for these three species (Loubry & Puig 1994; Sabatier 1997). Although *V. kwatae* and *V. surinamensis* occupy different habitats (*terra firme* vs flooded area), there are no geographic barriers to gene flow between them. In fact, habitat turnover occurs over short distances (few tens to few hundred meters) relative to dispersal distances: pollen can be dispersed over long distances by insects in tropical trees (Dick et al. 2003) notably in some *Virola* species (Scotti et al., submitted). There is also, space and time compatibilities favouring this gene flow which may lead to a *Virola* species complex. Moreover, even if there are morphological differences between these species, hybridisation cannot be excluded. Different cases of significant phenotypic differentiation despite homogenising effects of gene flow have already been reported at intraspecific level (Nosil & Crespi 2004; Lin et al. 2008) and interspecific level, as it is well documented for oak species (e.g. Whittemore & Schaal 1991; Kremer et al. 2002). Regardless, no morphological signals of hybridization have been reported in these *Virola* species yet. But, it has already been shown the existence of ‘cryptic hybrids’ in trees, i.e. individuals that are phenotypically, but not genetically, indistinguishable from one of their parental phenotypes (e.g. Jasińska et al. 2010). The three *V. kwatae* samples, grouped with the *V. surinamensis* group (figure 3b) may be cryptic individuals with a *V. kwatae* chlorotype in a

V. surinamensis nuclear background (corresponding to the borderline chlorotypes h21 and h27; figure 1). The absence of allele-sharing can suggest ancient gene flow. In fact, the patterns of genetic and ecological divergence observed here and by the detection of significant levels of interspecific gene flow raised some questions: the role and timing of gene flow during and after speciation, as well as its current extent. We have tested a simple model assuming that some gene flow has been going on without interruption since the beginning of interspecific divergence. Thus, if our results are compatible with a prolonged exchange after the start of speciation, it is hard to tell whether it has stopped between the start of divergence and today. Additionally, if speciation has occurred allopatrically, gene flow may have been caused by secondary contact after the two sister species have reached their currently observed sympatric distribution (Hewitt 2001; Gay et al. 2007). Regardless, we cannot exclude that an allopatric speciation occurred in absence of knowledge about past colonisations. More complex coalescent models or comparative phylogeographical approaches (Bermingham & Moritz 1998) may help to elucidate this.

The possibility of sympatric speciation of regionally distributed *V. kwatae* from widespread *V. surinamensis* (parenthetically, actual N_e results are in agreement with the geographical distribution of species, table 4) cannot be ruled out based on their current ranges. In this scenario, the former would have diverged from local populations of the latter without geographical separation, in the patchy environmental context described above, and we would be observing a case of sympatric speciation with continued, albeit restricted, gene flow. The alternative, allopatric scenario implies the divergence of local populations of *V. surinamensis* from the rest of the populations – perhaps during a phase of disjunction caused by hostile environmental conditions – followed by a secondary invasion of the range of newly emerged *V. kwatae* by the surrounding populations of its ancestor *V. surinamensis*. Changes in population size associated with local extinction followed by recolonization events (e.g. the refugia hypothesis; Haffer 1969) affect genetic properties (Wade & McCauley 1988; Tajima 1989; Dutech *et al.* 2003; Petit *et al.* 2003; Lowe *et al.* 2010), notably in acting on biogeography, effective population size, DNA polymorphism and on levels of differentiation. Especially, for this study, levels of demographic changes and the times of these events can be very helpful in speciation events knowledge and are too often ignored for this kind of study. We observed a contraction event (i.e. bottleneck with an higher ancient N_e for all species) for the three *Virola* species even if the credible intervals on shrinkage ratios and times are quite wide (table 4) (lack of precision may be caused by an insufficient sampling of loci (Peter et al. 2010) or by the fact that the demographic history of the populations is more complex than our

simple model). The complete lack of fossil record, as well as the structure of the current data set, do not allow to part these two hypotheses; nevertheless, our model indicates that all species have undergone population contraction (see below), whereas expansion would be expected for *V. surinamensis* if it had invaded the Guiana shield soon after its divergence from *V. kwatae*.

Conclusion

The data obtained on nuclear and chloroplast markers for three relatively recent species of *Virola* show that two of these species (*V. surinamensis* and *V. kwatae*) are genetically close although they occupy contrasted environments. On the contrary, the third species (*V. michelii*) is more genetically distant from its ecological equivalent (*V. kwatae*). Coalescent modelling helped us to understand that the current genetic patterns were probably caused by a relatively recent ecological divergence between sister species, a moderate reduction in population size and interspecific migration events. Our study highlights an example of rapid species radiation by divergent ecological selection in tropical trees, in contrast with the niche conservatism concept. The recent divergence between *V. surinamensis* and *V. kwatae* in the presence of homogenizing effects of significant gene flow strongly suggests that natural selection has played a very important role in the process of speciation. If the recent divergence time and interspecific gene flow inside the *Virola* genus, limiting molecular divergence, can be extrapolated at different genus of the Myristicaceae family, the difficulties of the resolution of relationships within this family can be explained (Stevens 2001 onwards; Sauquet et al. 2003; Doyle et al. 2004). For further improvement in understanding this scheme, it would be interesting to consider other regionally distributed *Virola* species (such as the rarer *V. multicostata*) and to extend comparisons to the whole genus, which includes species with a wide variety of ecological preferences (e.g. *V. sebifera* is a continentally common species that occupies open woodlands and savannas), to see whether other species could have originated from a rapid ecological radiation.

The loss of shared polymorphisms and the accumulation of fixed differences between taxa are supposed to occur more rapidly at loci involved in adaptive divergence or linked loci (i.e. by hitchhiking effect; Thomson 1977), compared to background genetic variation that can show little differentiation over a very large evolutionary time scale (Via 2009). It is true because selection at some loci may contribute to the development of some phenotypic difference (Wilding et al. 2001). Our case represents a part of this genetic background because of expected neutrality of the used loci. Audigeos et al. (2010) have already succeeded to isolate a gene (from PIP aquaporins family) that is potentially implicated in water stress in a Guianan *Virola* species (*V. sebifera*). Because *Virola* species have different environmental preferences, notably in water gradient (Baraloto et al. 2007), it is expected that evidence for selection can be detected. Genetic divergence based on candidate genes between species could comfort this underlying ecological speciation.

References

- Audigeos D, Buonamici A, Belkadi L, *et al.* (2010) Aquaporins in the wild: Natural genetic diversity and selective pressure in the PIP gene family in five Neotropical tree species. *BMC Evolutionary Biology* **10**.
- Baraloto C, Morneau F, Bonal D, Blanc L, Ferry b (2007) Seasonal water stress tolerance and habitat associations within four neotropical tree genera. *Ecology* **88**, 478-489.
- Beaumont MA (1999) Detecting Population Expansion and Decline Using Microsatellites. *Genetics* **153**, 2013-2029.
- Bermingham E, Moritz C (1998) Comparative phylogeography: concepts and applications. *Molecular Ecology* **7**, 367-369.
- Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology* **19**, 2609-2625.
- Bohonak AJ (1999) Dispersal, gene flow, and population structure. *The Quarterly Review of Biology* **74**, 21-45.
- Bowie RCK, Fjeldså J, Hackett SJ, Bates JM, Crowe TM (2006) Coalescent models reveal the relative roles of ancestral polymorphism, vicariance, and dispersal in shaping phylogeographical structure of an African montane forest robin. *Molecular Phylogenetics and Evolution* **38**, 171-188.
- Cavender-Bares J, Kozak KH, Fine PVA, Kembel SW (2009) The merging of community ecology and phylogenetic biology. *Ecology Letters* **12**, 693-715.
- Cavender-Bares J, Pahlich A (2009) Molecular, morphological, and ecological niche differentiation of sympatric sister oak species, *Quercus virginiana* and *Q. geminata* (Fagaceae). *American Journal of Botany* **96**, 1690-1702.
- Chan KMA, Levin SA (2005) Leaky prezygotic isolation and porous genomes: Rapid introgression of maternally inherited DNA. *Evolution* **59**, 720-729.
- Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657-1659.
- Cornuet J, Luikart G (1997) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**, 2001 - 2014.
- Coyne JA, Orr HA (2004) *Speciation* Sinauer Associates Inc, Sunderland, Massachusetts.
- Cronquist A (1981) *An integrated system of classification of flowering plants* The New York Botanical Garden, New York.
- Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution* **25**, 410-418.
- Dick CW, Etchelecu G, Austerlitz F (2003) Pollen dispersal of tropical trees (*Dinizia excelsa*: Fabaceae) by native insects and African honeybees in pristine and fragmented Amazonian rainforest. *Molecular Ecology* **12**, 753-764.
- Doyle JA, Manchester SR, Sauquet H (2008) A seed related to Myristicaceae in the Early Eocene of southern England. *Systematic Botany* **33**, 636-646.
- Doyle JA, Sauquet H, Scharaschkin T, Thomas AL (2004) Phylogeny, molecular and fossil dating, and biogeographic history of Annonaceae and Myristicaceae (Magnoliales). *International Journal of Plant Sciences* **165**, S55-S67.
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**, 11-15.
- Draheim H, Cui M, Dick CW (2009) Characterization of 14 microsatellite DNA markers for the tropical forest tree *Virola surinamensis* (Rol.) Warb. (Myristicaceae). *Molecular Ecology Resources* **9**, 1386-1388.
- Dutech C, Maggia L, Tardy C, Joly HI, Jarne P (2003) Tracking a genetic signal of extinction-recolonisation events in a neotropical tree species : *Vouacapoua americana* Aublet in French Guiana. *Evolution* **57**, 2753-2764.
- Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* **72**, 250-259.

- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**, 2611 - 2620.
- Excoffier L, Estoup A, Cornuet J-M (2005) Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* **169**, 1727-1738.
- Excoffier L, Novembre J, Schneider S (2000) Computer note. SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *Journal of Heredity* **91**, 506-509.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.
- Field DL, Ayre DJ, Whelan RJ, Young AG (2010) Patterns of hybridization and asymmetrical gene flow in hybrid zones of the rare *Eucalyptus aggregata* and common *E. rubida*. *Heredity*.
- Futuyma DJ (2008) Ecology, speciation and adaptive radiation: the long view. *Evolution* **62**, 2446-2449.
- Gaggiotti OE (2011) Making inferences about speciation using sophisticated statistical genetics methods: look before you leap. *Molecular Ecology* **20**, 2229-2232.
- Gay L, Neubauer G, Zagalska-Neubauer M, et al. (2007) Molecular and morphological patterns of introgression between two large white-headed gull species in a zone of recent secondary contact. *Molecular Ecology* **16**, 3215-3227.
- Givnish TJ (2010) Ecology of plant speciation. *Taxon* **59**, 1329-1366.
- Haffer Jr (1969) Speciation in Amazonian forest birds. *Science* **165**, 131-137.
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95-98.
- Harvey PH, Pagel MD (1991) *The comparative method in evolutionary biology* Oxford University Press, Oxford, UK.
- Hausdorf B, Hennig C (2010) Species Delimitation Using Dominant and Codominant Multilocus Markers. *Systematic Biology* **59**, 491-503.
- Heled J, Drummond AJ (2010) Bayesian Inference of Species Trees from Multilocus Data. *Molecular Biology and Evolution* **27**, 570-580.
- Hewitt GM (2001) Speciation, hybrid zones and phylogeography — or seeing genes in space and time. *Molecular Ecology* **10**, 537-549.
- Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. *Current Opinion in Genetics & Development* **16**, 592-596.
- Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution* **27**, 905-920.
- Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**, 747-760.
- Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences* **104**, 2785-2790.
- Hollingsworth PM, Forrest LL, Spouge JL, et al. (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences* **106**, 12794-12797.
- Janovec JP, Neill AK (2002) Studies of the Myristicaceae: an overview of the *Compsoeura atopa* complex, with descriptions of new species from Colombia. *Brittonia* **54**, 251-261.
- Jarne P, Lagoda P (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution* **11**, 424-429.
- Jasińska AK, Wachowiak W, Muchewicz E, et al. (2010) Cryptic hybrids between *Pinus uncinata* and *P. sylvestris*. *Botanical Journal of the Linnean Society* **163**, 473-485.
- Keyghobadi N, Roland J, Strobeck C (2005) Genetic differentiation and gene flow among populations of the alpine butterfly, *Parnassius smintheus*, vary with landscape connectivity. *Molecular Ecology* **14**, 1897-1909.
- Kingman JFC (1982) On the genealogy of large populations. *Journal of Applied Probability* **19**, 27-43.

- Kremer A, Dupouey JL, Deans JD, *et al.* (2002) *Leaf morphological differentiation between Quercus robur and Quercus petraea is stable across western European mixed oak stands* EDP Sciences, Les Ulis, France.
- Lande R, Barrowclough GF (1987) *Effective population size, genetic variation and their use in population management* Cambridge University Press.
- Laval G, Excoffier L (2004) SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics* **20**, 2485-2487.
- Leigh EG, Davidar P, Dick CW, *et al.* (2004) Why Do Some Tropical Forests Have So Many Species of Trees? *Biotropica* **36**, 447-473.
- Levin DA (2004) The ecological transition in speciation. *New Phytologist* **161**, 91-96.
- Lexer C, Fay M (2005) Adaptation to environmental stress: a rare or frequent driver of speciation? *Journal of Evolutionary Biology* **18**, 893-900.
- 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. *Molecular Ecology* **14**, 1045-1057.
- Lexer C, Widmer A (2008) The genic view of plant speciation: recent progress and emerging questions. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 3023-3036.
- Lin J, Ziegler E, Quinn TP, Hauser L (2008) Contrasting patterns of morphological and neutral genetic divergence among geographically proximate populations of sockeye salmon *Oncorhynchus nerka* in Lake Aleknagik, Alaska. *Journal of Fish Biology* **73**, 1993-2004.
- Losos JB (2008) Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters* **11**, 995-1003.
- Loubry D, Puig H (1994) *Déterminismes du comportement phénologique des arbres en forêt tropicale humide de Guyane française*, Université de Paris VI.
- Lowe A, Harris D, Dormontt E, Dawson I (2010) Testing putative african tropical forest refugia using chloroplast and nuclear DNA phylogeography. *Tropical Plant Biology* **3**, 50-58.
- Macía M (2011) Spatial distribution and floristic composition of trees and lianas in different forest types of an Amazonian rainforest. *Plant Ecology* **212**, 1159-1177.
- Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology and Evolution* **20**, 229-237.
- Mallet J (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society: Biological Sciences* **363**, 2971-2986.
- Marko PB, Hart MW (2011) Retrospective coalescent methods and the reconstruction of metapopulation histories in the sea. *Evolutionary Ecology Online First*, 25.
- McCauley DE, Raveill J, Antonovics J (1995) Local founding events as determinants of genetic structure in plant metapopulations. *Heredity* **75**, 630-636.
- Morjan CL, Rieseberg LH (2004) How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology* **13**, 1341-1356.
- Muir G, Schlötterer C (2005) Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus spp.*). *Molecular Ecology* **14**, 549-561.
- Neophytou C, Aravanopoulos FA, Fink S, Dounavi A (2010) Detecting interspecific and geographic differentiation patterns in two interfertile oak species (*Quercus petraea* (Matt.) Liebl. and *Q. robur* L.) using small sets of microsatellite markers. *Forest Ecology and Management* **259**, 2026-2035.
- Newmaster SG, Fazekas AJ, Steeves RAD, Janovec J (2008) Testing candidate plant barcode regions in the Myristicaceae. *Molecular Ecology Resources* **8**, 480-490.
- Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology* **17**, 2103-2106.

- Nosil P, Crespi BJ (2004) Does gene flow constrain adaptive divergence or vice versa ? A test using ecomorphology and sexual isolation in *Timera cristinae* walking-sticks. *Evolution* **58**, 102-112.
- Nosil P, Egan SP, Funk DJ, Hoekstra H (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: 'Isolation by Adaptation' and multiple roles for divergent selection. *Evolution* **62**, 316-336.
- Peter BM, Wegmann D, Excoffier L (2010) Distinguishing between population bottleneck and population subdivision by a Bayesian model choice procedure. *Molecular Ecology* **19**, 4648-4660.
- Petit RJ, Aguinalalde I, de Beaulieu J-L, *et al.* (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**, 1563-1565.
- Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. In: *Annual Review of Ecology Evolution and Systematics*, pp. 187-214. Annual Reviews, Palo Alto.
- Powell JR (1983) Interspecific cytoplasmic gene flow in the absence of nuclear gene flow: evidence from *Drosophila*. *Proceedings of the National Academy of Sciences* **80**, 492-495.
- Pritchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945 - 959.
- Queenborough S, Burslem D, Garwood N, Valencia R (2007) Habitat niche partitioning by 16 species of Myristicaceae in Amazonian Ecuador. *Plant Ecology* **192**, 193-207.
- Rainey PB, Travisano M (1998) Adaptive radiation in a heterogeneous environment. *Nature* **394**, 69-72.
- Richard M, Thorpe RS (2001) Can microsatellites be used to infer phylogenies? Evidence from population affinities of the Western Canary Island lizard (*Gallotia galloti*). *Molecular Phylogenetics and Evolution* **20**, 351-360.
- Rieseberg L, Willis J (2007) Plant speciation. *Science* **317**, 910 - 914.
- Rieseberg L, Soltis DE (1991) Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* **5**, 65-84.
- Roda F (2011) *The genomic basis of ecological speciation in plants*, University of Queensland.
- Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics* **3**, 380-390.
- Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters* **8**, 336-352.
- Sabatier D (1997) Description and biology of a new species of *Virola* (Myristicaceae) from Guyana. *Adansonia (Paris)* **19**, 273-278.
- Sauquet H, Doyle JA, Scharaschkin T, *et al.* (2003) Phylogenetic analysis of Magnoliales and Myristicaceae based on multiple data sets: implications for character evolution. *Botanical Journal of the Linnean Society* **142**, 125-186.
- Savolainen V, Anstett M-C, Lexer C, *et al.* (2006) Sympatric speciation in palms on an oceanic island. *Nature* **441**, 210-213.
- Schluter D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution* **16**, 372-380.
- Schluter D (2009) Evidence for Ecological Speciation and Its Alternative. *Science* **323**, 737-741.
- Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution* **19**, 198-207.
- Shaw J, Lickey E, Beck J, *et al.* (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* **92**, 142-166.
- Slatkin M (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics* **16**, 393-430.
- Stevens PF (2001 onwards) Angiosperm phylogeny website. Version 9, June 2008. <<http://www.mobot.org/mobot/research/apweb/>>.
- Tajima F (1989) The Effect of Change in Population Size on DNA Polymorphism. *Genetics* **123**, 597-601.
- Templeton A, Crandall K, Sing C (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**, 619 - 633.

- Thomson G (1977) The effect of a selected locus on linked neutral loci. *Genetics* **85**, 753-788.
- Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences* **106**, 9939-9946.
- Wade MJ, McCauley DE (1988) Extinction and Recolonization: Their Effects on the Genetic Differentiation of Local Populations. *Evolution* **42**, 995-1005.
- Webb CO, Ackerly DD, McPeck MA, Donoghue MJ (2002) Phylogenies and community ecology. *Annual Review of Ecology and Systematics* **33**, 475-505.
- Whittemore AT, Schaal BA (1991) Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences* **88**, 2540-2544.
- Wiens JJ, Ackerly DD, Allen AP, *et al.* (2010) Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters* **13**, 1310-1324.
- Wiens JJ, Graham CH (2005) Niche conservatism: Integrating evolution, ecology, and conservation biology. In: *Annual Review of Ecology Evolution and Systematics*, pp. 519-539. Annual Reviews, Palo Alto.
- Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology* **14**, 611-619.
- Wilson TK (2004) *Myristicaceae*. Pp 261 - 262. In: N. Smith, S.A. Mori, A. Henderson, D.W. Stevenson & S.V. Heald (eds) : *Flowering Plants of the Neotropics* New York Botanical Garden & Princeton University Press, Princeton.
- Won Y-J, Hey J (2005) Divergence population genetics of chimpanzees. *Molecular Biology and Evolution* **22**, 297-307.
- Wu C-I (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**, 851-865.
- Zheng X-M, Ge S (2010) Ecological divergence in the presence of gene flow in two closely related *Oryza* species (*Oryza rufipogon* and *O. nivara*). *Molecular Ecology* **19**, 2439-2454.

Figures and tables

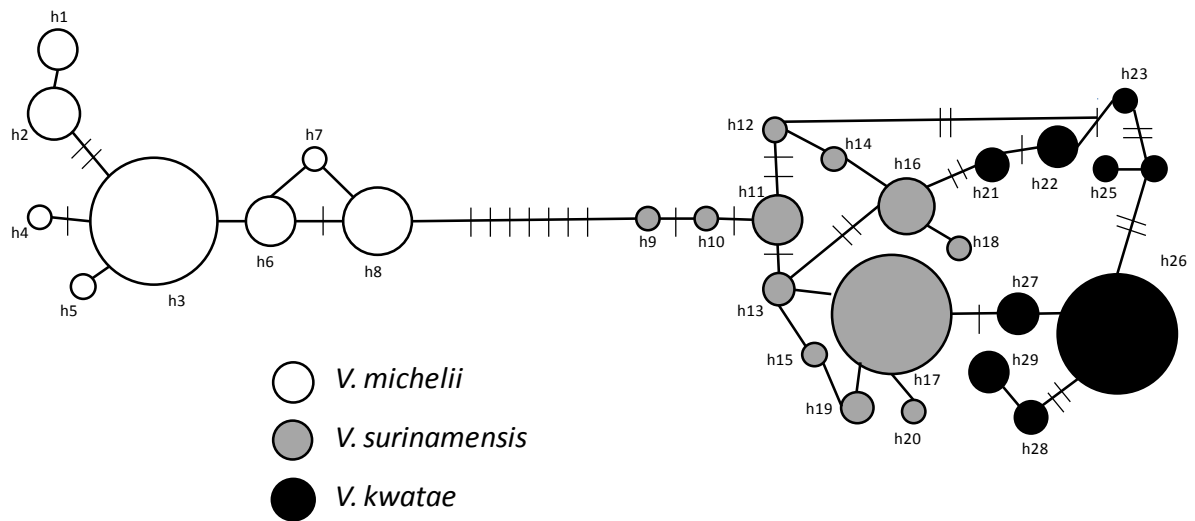


Figure 1: Network of the 29 cpDNA (*trnH-psbA* + *trnC-ycf6*) haplotypes found in *Virola michelii* (white), *V. surinamensis* (grey) and *V. kwatae* (black) obtained from the matrix of pairwise distances calculated between all pairs of haplotypes.

Each haplotype is represented by a circle whose size is proportional to its frequency. Lines connecting haplotypes correspond to one mutation distinguishing the haplotypes (+ one mutation added for each acrossed small line).

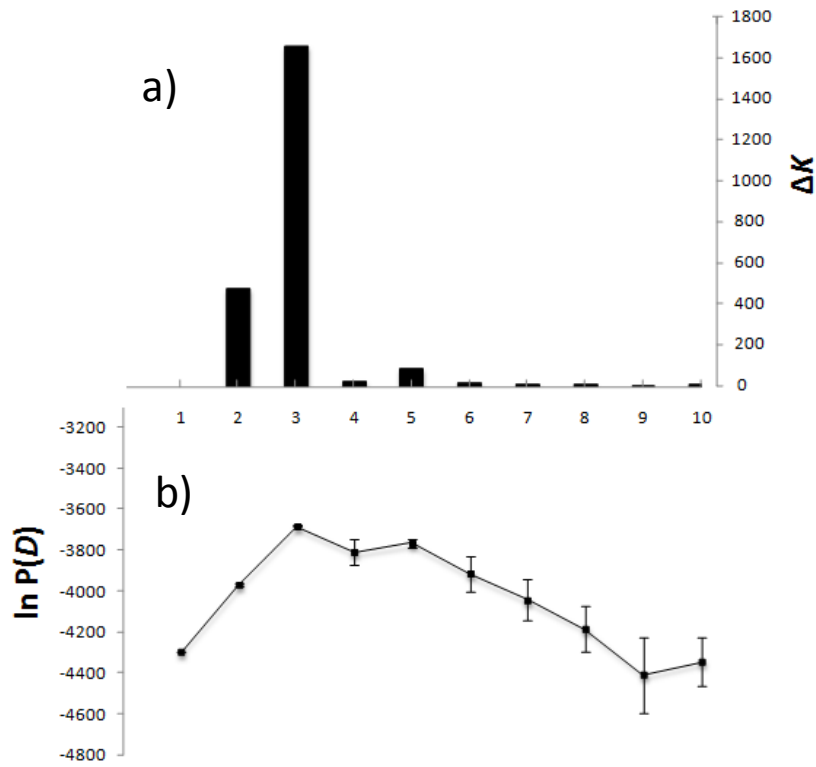


Figure 2: Estimation of number of genetic groups (K).

- a) Values of ΔK (above) plotted against K (for K from 1 to 10).
- b) Mean and standard deviation of $\ln P(D)$ for 10 independent runs plotted against K used in the structure clustering analysis.

In both cases, the peak indicates the most probable number of genetic groups given the data.

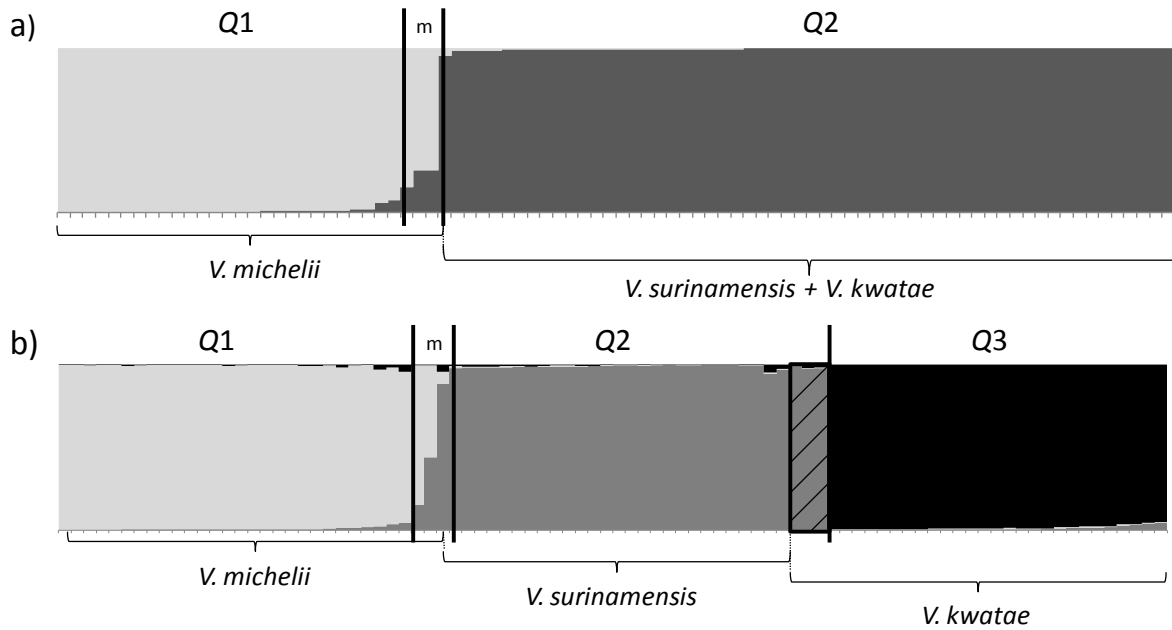


Figure 3: Genetic assignment of individuals according to the Bayesian method implemented in the program STRUCTURE at a) $K = 2$ and b) $K = 3$.

Individuals are represented by vertical colored columns. Same color in different individuals indicates that they belong to the same genetic group. Different colors in the same individual indicate the posterior probability to belong to the different genetic clusters ($Q1$, $Q2$ or $Q3$). We consider that individuals belong to a cluster if their threshold value in the admixture coefficient are upper than 0.90. If not, individuals are ranged to the group of non-assigned individuals (m). Hatched area in part b represents the three *V. kwatae* grouped with *V. surinamensis*.

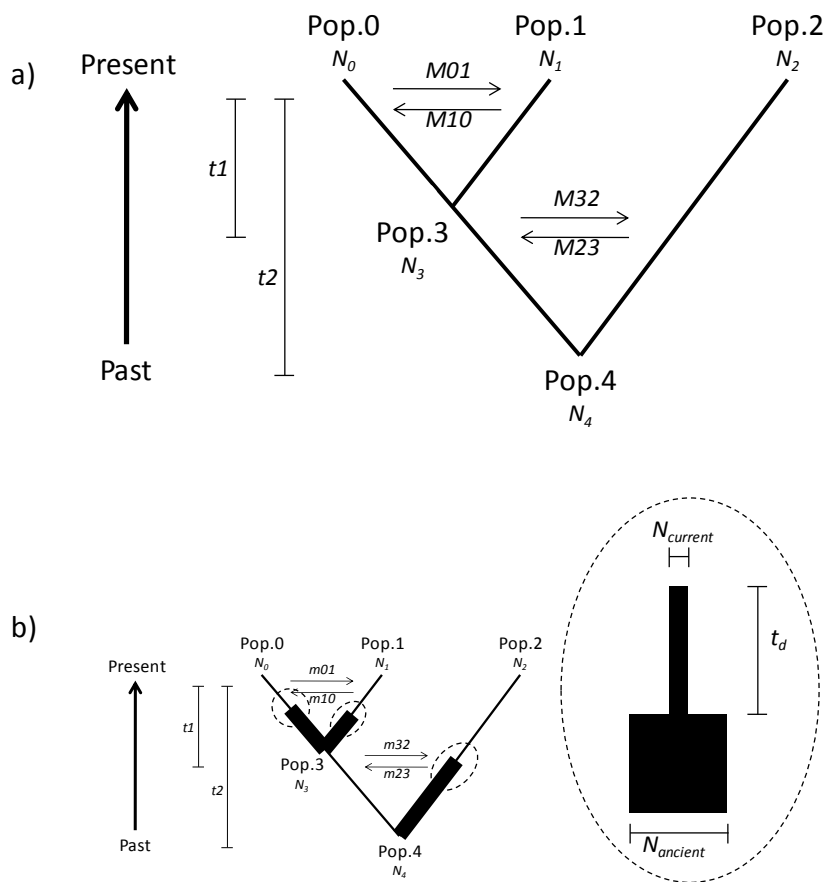


Figure 4: tree typology and demographic parameters.

a) This typology was chosen in regard to the results of the clustering analysis and haplotypes network. Parameters are issued from the isolation with migration model (IMa2).

b) The same typology with the same parameters was chosen for the SIMCOAL2 approach, with the different it allowed changes in demographic size of populations.

Pop.0: *Virola surinamensis*, Pop.1: *V. kwatae*, Pop.2: *V. michelii*. Pop.3 and Pop.4 are ancestral populations.

N_1 , N_2 , N_3 , N_4 and N_5 : effective population sizes.

M_{01} , M_{10} , M_{32} and M_{23} : gene-flow between population (direction: from first population to the second) estimated by IMa2 with $M = 2Nm$.

m_{01} , m_{10} , m_{32} and m_{23} : : gene-flow between population (direction: from first population to the second) estimated by SIMCOAL2.

t_1 and t_2 : times of population splitting.

$N_{current}$: current population size, $N_{ancient}$: ancient population size and t_d : time since population size change. The ratio $N_{ancient} / N_{current} > 1$ will mean a bottleneck event, and inversely, $N_{ancient} / N_{current} < 1$ an expansion event.

t_d : time of the demographic event in number of generations.

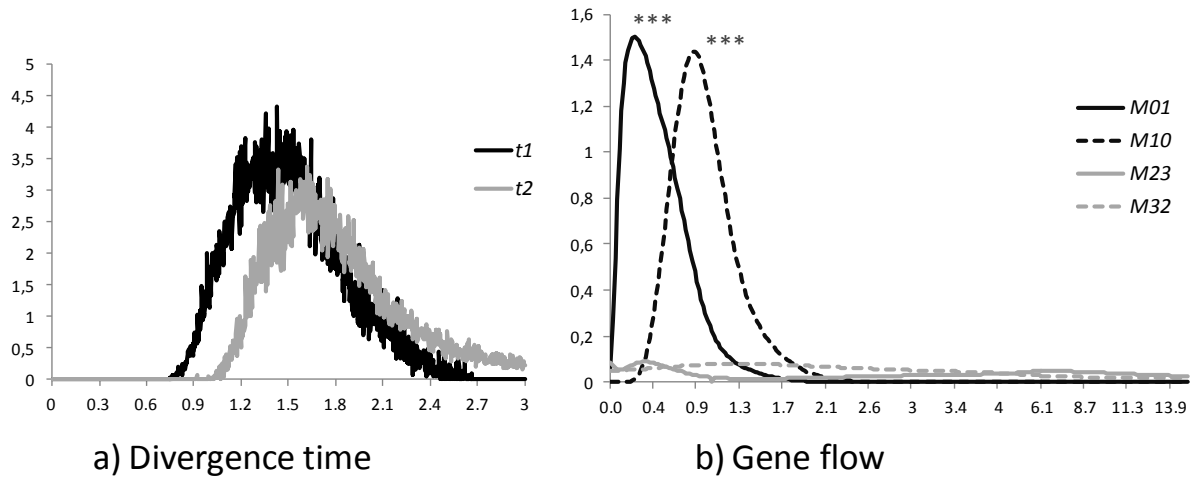


Figure 5: The marginal posterior probability distributions for demographic parameters from IMA approach.

a) Divergence time in millions of years ago (MYA).

b) Population migration rate estimate (i.e. effective migration rate at which genes are introduced into a population per generation ($M = 2Nm$)).

For migration parameter, statistical significance as assessed by log-likelihood ratio (LLR) tests is indicated: ***(<0.001).

$M01$, $M10$, $M32$ and $M23$: gene-flow between population (direction: from first population to the second) estimated with $M = 2Nm$ and 0: *Virola surinamensis*, 1: *V. kwatae*, 2: *V. michelii*, 3: ancestral population (see figure 1)

$t1$ and $t2$: times of population splitting (see figure 4)

Table 1: Diversity parameters assessed with nuclear microsatellites and chloroplast DNA markers in *Virola* species.

species	nuclear microsatellites (10 loci)				chloroplast DNA		
	<i>N</i>	<i>R</i>	<i>Ho</i>	<i>He</i>	<i>N</i>	<i>R</i>	<i>H</i>
<i>V.michellii</i>	46.40 (8.98)	16.5 (8.16)	0.59 (0.29)	0.74 (0.29)	32	8	0.24 (0.16)
<i>V.surinamensis</i>	46.20 (6.82)	16.5 (7.13)	0.73 (0.17)	0.87 (0.13)	32	12	0.27 (0.16)
<i>V.kwatae</i>	45.60 (17.50)	14.4 (9.27)	0.69 (0.28)	0.80 (0.27)	29	9	0.25 (0.09)
all species	138.2 (28.91)	34.4 (17.17)	0.66 (0.21)	0.89 (0.11)	93	29	0.34 (0.17)

N: number of genes sampled, *R*: related allelic richness, *Ho*: observed heterozygosity, *He*: expected heterozygosity, *H chloroplast diversity*. Numbers in parenthesis correspond to standard deviation values.

Table 2: Pairwise *Fst* values of genetic differentiation among the three species for chloroplast DNA markers (below) and nuclear microsatellites (above).

	<i>V. michellii</i>	<i>V. surinamensis</i>	<i>V. kwatae</i>
<i>V. michellii</i>	-	0.164	0.224
<i>V. surinamensis</i>	0.793	-	0.108
<i>V. kwatae</i>	0.809	0.431	-

Number in bold indicate statistically significant values ($P < 0.05$) based on 1000 random permutations.

Table 3: Maximum-Likelihood estimates (MLE) and the 95% Highest Posterior Density (HPD) intervals of demographic parameters from ‘Isolation with Migration’ approach.

		MLE	lower 95% HDP	upper 95% HPD
mutation	θ_0	0.9046	0.4656	1.629
	θ_1	0.9561	0.5200	1.722
	θ_2	11.54	5.702	21.09
	θ_3	0.4129	0.2109	0.7112
	θ_4	0.6890	0.3802	1.127
	θ_5	2.200	1.213	3.597
	θ_6	0.2288	0.09550	0.4742
effective population size	N_0	707175	443750	993750
	N_1	495057	281250	718750
	N_2	1135724	643750	1693750
	N_3	6923184	1481250	12493750
	N_4	7049486	3756250	10681250
gene flow	M_{01}	0.4779***	0.02752	1.051
	M_{10}	0.9480***	0.3846	1.617
	M_{23}	8.845	0.0	18.86
	M_{32}	6.955	0.0	18.00
divergence time	t_1	1509912	916500	2206500
	t_2	1796998	1132500	2704500

θ : mutation rate in $4N\mu$ (where N is the effective population size and μ the mutation rate per base pair), N : effective population size, M : effective migration rate ($M = 2Nm$) with direction from first population to the second, t : divergence time in years.

Populations: 0: *Virola surinamensis*, 1: *V. kwatae*, 2: *V. michelii*, 3 & 4: ancestral population (see figure 4).

migration significance levels (likelihood ratio test; LLT): ns non-significative; * $p < 0.05$; ** $p < 0.01$, *** $p < 0.001$

Table 4: Maximum-Likelihood estimates (MLE) and the 95% Highest Posterior Density (HPD) intervals of demographic parameters with the ABC approach.

		Mean	lower 95% HDP	upper 95% HPD
mutation	μ_0	0.000849633	0.000589441	0.000906034
	μ_1	0.000944944	0.000622215	0.000999388
	μ_2	0.000109332	1.58803e-06	0.000975261
	μ_3	6.86752e-06	1.04938e-06	2.14057e-05
	μ_4	3.40204e-05	1.40268e-06	0.000114234
	μ_5	0.000676063	0.000131062	0.000992314
	μ_6	0.000461298	3.50609e-06	0.000982094
effective population size	N_0	35449.6	11839.6	49884.4
	N_1	1168.57	288.707	2713.02
	N_2	30609.8	13328.4	47658.7
gene flow	m_{01}	7.88515e-07	1.23986e-08	3.26453e-06
	m_{10}	8.78604e-09	4.35938e-11	2.48565e-08
	m_{23}	2.656e-06	2.0109e-08	6.92159e-06
	m_{32}	1.06559e-08	1.14628e-10	4.05438e-08
divergence time	t_1	22299.9	5316.51	108976
	t_2	45389.3	7597.2	159017
demographic event time	t_{d0}	19895.1	620.569	73030.4
	t_{d1}	535.537	8.14301	1514.82
	t_{d2}	97248.5	1123.17	247224
demographic event	$N_{ancient0} / N_{current0}$	610.046	5.59363	1387.75
	$N_{ancient1} / N_{current1}$	163.936	2.03766	1358.2
	$N_{ancient2} / N_{current2}$	1199.65	142.29	1499.82

μ : mutation rate, N : current population size, m : migration rate per generation with direction from first population to the second, t : divergence time in generations, t_d : time of the in generations, the ratio $N_{ancient} / N_{current} > 1$ will mean a bottleneck event, and inversely, $N_{ancient} / N_{current} < 1$ an expansion event.

Populations: 0: *Virola surinamensis*, 1: *V. kwatae*, 2: *V. michelii*, 3 & 4: ancestral population (see figure 4).

Supplementary

Supplementary table 1: Population-level heterozygosity H_e per locus.

	Locus 1	Locus 2	Locus 3	Locus 4	Locus 5	Locus 6	Locus 7
<i>V. kwatae</i>	0.93501	0.81609	0.96794	0.83673	0.87624	0.95187	1.00000
<i>V. surinamensis</i>	0.90939	0.94686	0.96305	0.82694	0.86154	0.95122	0.95968
<i>V. michelii</i>	0.92857	0.92609	0.95755	0.95749	0.82308	0.88205	0.67143

CONCLUSION

Cette étude de la diversité génétique et de l'adaptation des arbres forestiers tropicaux a été réalisée à des échelles taxonomiques et géographiques différentes. Le principal objectif de cette thèse était de caractériser, et surtout d'expliquer la diversité et la variabilité génétique chez les arbres forestiers tropicaux, et plus particulièrement à l'intérieur du genre *Viola* (Myristicaceae) au niveau interspécifique et intraspécifique. La diversité génétique est à l'origine de la biodiversité et permet de décrire différents processus évolutifs tels que la sélection naturelle qui permet l'adaptation des individus à leur milieu et qui peut aboutir à la spéciation.

Les perturbations anthropiques, telles que l'exploitation forestière, peuvent avoir des conséquences sur la diversité génétique et sur la structuration génétique de la régénération. L'étude du premier article informe que cette modification rapide de l'environnement tend à favoriser la diversité génétique au cours de la régénération de *Viola michelii*. En effet, des individus se situant dans la parcelle exploitée et aux alentours des trouées d'exploitation possèdent globalement une plus grande diversité génétique que les individus issus d'un processus de régénération entièrement naturelle. Plusieurs hypothèses peuvent expliquer ce résultat. L'ouverture du milieu permet une régénération plus importante pour les espèces héliophiles ou semi-héliophiles et une moindre compétition pour l'espace entre plantules qui peuvent provenir de plusieurs parents. Les études de parentalité ont révélé que les adultes reproducteurs participant à la régénération, bien que peu nombreux, dispersent largement leur gamètes sur tout le site d'étude et jusqu'à des distances supérieures à un kilomètre. Cela indique d'une part, que *V. michelii* possède une bonne dispersion génique à travers le vecteur animal (insectes pour le pollen, singes et toucans principalement pour les graines), et d'autre part que les animaux peuvent assurer un bon mixage des génotypes qui se répercute sur les niveaux de diversité mesurés, notamment en périphérie des trouées qui pourraient s'avérer être des zones favorisées pour les phases d'alimentation des disséminateurs. De plus, ces zones présenteraient une plus grande variété de conditions environnementales (notamment par rapport au gradient lumineux) et donc pourraient favoriser l'établissement d'une plus grande diversité de génotypes. Ce travail préliminaire n'a été effectué que sur une relativement faible

surface d'exploitation (6.5 ha) avec une perturbation considérée comme assez faible (prélèvement d'environ 10 arbres adultes par hectare). Il serait intéressant de comparer ces résultats sur de plus grande zones avec des perturbations différentes afin d'optimiser des plans de gestion de conservation, notamment chez les espèces d'intérêt commerciale, comme *Virola*.

Bien qu'une forte diversité génétique soit garante de prospérité pour le devenir des populations, un suivi de la survie des génotypes serait nécessaire pour comprendre si la diversité observée chez les juvéniles est maintenue chez la population adulte qui en dérive ; en effet une minorité des locus pourrait déterminer le potentiel adaptatif de ces cohortes et déterminer ainsi le destin de la population. Il est donc indispensable d'estimer quelle proportion du génome est impliquée dans l'adaptation à la variabilité locale des conditions du milieu.

Une étude de l'adaptation locale des individus sur un milieu comprenant des caractéristiques microenvironnementales bien définies, comme sur le dispositif expérimental de Paracou, est une bonne stratégie pour mettre en évidence l'action de la sélection naturelle au niveau intraspécifique (article 2). En effet, ce travail indique bien que les espèces d'arbres forestiers tropicaux, ici avec l'exemple d'une population de *V. michelii* (espèce abondante du bouclier guyanais pouvant montrer une hétérogénéité dans sa distribution spatiale locale), montrent des signes de divergence adaptative pour différents microenvironnements. Des marqueurs moléculaires AFLP (N = 229) ont été utilisés dans une démarche de criblage génomique et ont montré que deux régions génomiques sont soumises à une sélection naturelle divergente notamment par rapport aux contraintes hydriques du milieu. Cela confirme l'idée reçue que la majorité des loci AFLP utilisés sont neutres, et que la variation génétique associée à ces marqueurs est effectivement neutre vis-à-vis des contraintes du milieu analysées. Néanmoins, cette étude ne révèle que deux loci outliers chez les individus adultes sur un total de trois comparaisons pour trois variables environnementales différentes (donc sur un total de $3 \times 229 = 687$ loci, soit 0.3%). Ce faible niveau de détection peut s'expliquer, d'une part par la stringence choisie pour la significativité de nos tests statistiques ($p = 0.99$), mais également par la juxtaposition des résultats de deux approches différentes qui peuvent s'avérer complémentaires pour limiter les erreurs de détection de faux positifs, comme indiqué très récemment par différents articles méthodologiques (Pérez-Figueroa *et al.* 2010; Narum & Hess 2011). Cet extrême conservatisme dans le choix des critères de détection de la sélection permet d'affirmer que les locus identifiés ont des très faibles chances

d'être des artefacts. Je peux donc affirmer avoir détecté la présence de signatures sélectives au niveau moléculaire, prouvant l'adaptation locale chez les arbres forestiers tropicaux. Il serait cependant nécessaire de valider ces résultats sur d'autres sites d'étude. Cette étude aura, de plus, permis de constater que la détection de la sélection naturelle ne se fait pas sur tous les stades de vie des individus. En effet, pour toutes les nombreuses analyses concernant les juvéniles, aucun signe de sélection n'a été détecté. Cela pourrait s'expliquer par le fait que le tri des génotypes par le filtrage environnemental n'a pas encore été effectué et donc que ce stade de vie est trop précoce pour que l'on puisse y détecter des signes de sélection naturelle. Par une approche différente, cette étude vient compléter et conforter le travail initié par D. Audigeos (2010; in prep) sur l'adaptation locale des arbres forestiers tropicaux qui montre une distribution non aléatoire des allèles le long d'un gradient environnemental pour certains gènes candidats. L'absence d'autres études sur l'adaptation locale des arbres tropicaux donne une importance toute particulière à ce travail. Pour mieux comprendre les bases génétiques du processus d'adaptation, outre la nécessité d'isoler des gènes candidats (le séquençage du transcriptome d'une espèce de *Virola* (*V. surinamensis*) a été réalisé au sein du laboratoire, mais les analyses ne sont pas encore achevées pour permettre l'identification des gènes potentiellement impliqués dans l'adaptation à l'environnement), une prochaine étape serait d'analyser quantitativement l'expression des gènes sur des traits phénotypiques adaptatifs d'individus suivis en milieu environnemental contrasté contrôlé.

La divergence adaptative des populations pour des environnements contrastés peut aboutir sur le long terme au processus de spéciation écologique.

La spéciation écologique, bien que rarement décrite dans la littérature pour les arbres forestiers, sauf pour quelques espèces modèles (Lexer & Fay 2005; Cavender-Bares & Pahlisch 2009) a été observée, ici, au sein du genre *Virola* avec *V. kwatae* ayant divergé à partir de *V. surinamensis* (article 3). En effet, ces deux espèces occupent des milieux contrastés, notamment concernant les caractéristiques hydriques du milieu. Bien que l'observation de la distribution géographique actuelle des deux espèces puisse laisser supposer un processus de spéciation sympatrique, celui-ci ne peut pas être affirmé avec certitude. En effet, des analyses phylogéographiques (Avice 2000) prenant en compte les déplacements géographiques passés des populations seraient nécessaires pour éclaircir cette zone d'ombre. Ces deux espèces ont divergé il y a entre 0.9 et 2.3 millions d'années. Cette divergence paraît alors très récente vu l'âge de la famille des Myristicaceae (début de l'Eocène, Doyle *et al.* 2008). Pourtant, celle-ci n'explique pas entièrement le faible niveau de divergence génétique

estimé entre ces deux espèces sœurs sur la base de marqueurs neutres, car leur ancêtre a également divergé très récemment de l'espèce *V. michelii* (entre 1.1 et 2.7 millions d'années) qui, elle, montre un niveau de divergence génétique plus prononcé par rapport aux deux autres espèces. Les goulets d'étranglements détectés pour ces trois espèces ne semblent pas non plus expliquer en majorité ces patrons de diversité. Ce sont les flux de gènes interspécifiques, mis en évidence par les approches de coalescence entre *V. kwatae* et *V. surinamensis*, qui en sont la cause la plus probable en homogénéisant les pools génétiques de ces deux espèces.

Pour ce type d'étude, il est bien évidemment nécessaire d'être certain de l'identité biologique des échantillons. Deux marqueurs de séquences chloroplastiques (*trnH – psbA* et *trnC – Ycf6*) et un lot de 10 microsatellites nucléaires (Draheim *et al.* 2009) ont été analysés sur plusieurs centaines d'individus. Certains de ces individus ont montré des génotypes bien différents de ceux de leur congénères et ont été assignés dans les groupes génétiques d'autres espèces à la fois au niveau nucléaire et au niveau chloroplastique ; cela a été en particulier le cas de certains individus de *V. michelii* assignés dans les groupes génétiques de *V. kwatae*. Une vérification botanique sur le terrain a révélé des erreurs d'identification, ou de saisie dans la base de données, montrant que les marqueurs utilisés pouvaient être utiles pour discriminer les espèces ; cette approche « à l'aveugle » a été mise en avant par Duminil et ses collaborateurs (2006).

Une thèse récente sur les relations phylogénétiques du genre *Viola* concorde avec mes résultats (Steeves 2011). En effet, des résultats de cette thèse montrent que les relations phylogénétiques au sein de ce genre sont difficiles à déterminer pour certaines espèces à cause d'un manque de divergence moléculaire (tout comme l'avait déjà mis en évidence (Sauquet *et al.* 2003)). Une étude menée dans le cadre de cette thèse vise à comparer génétiquement et morphologiquement deux espèces de *Viola* (*V. sebifera* et *V. loretensis*). Bien qu'estimé sur la base d'un seul marqueur chloroplastique (*trnH – psbA*), le faible niveau de divergence génétique et l'architecture ambiguë du réseau des haplotypes laissent supposer à l'auteur que d'éventuels flux de gènes entre les espèces existent (Figure 5) malgré une certaine intégrité des espèces.

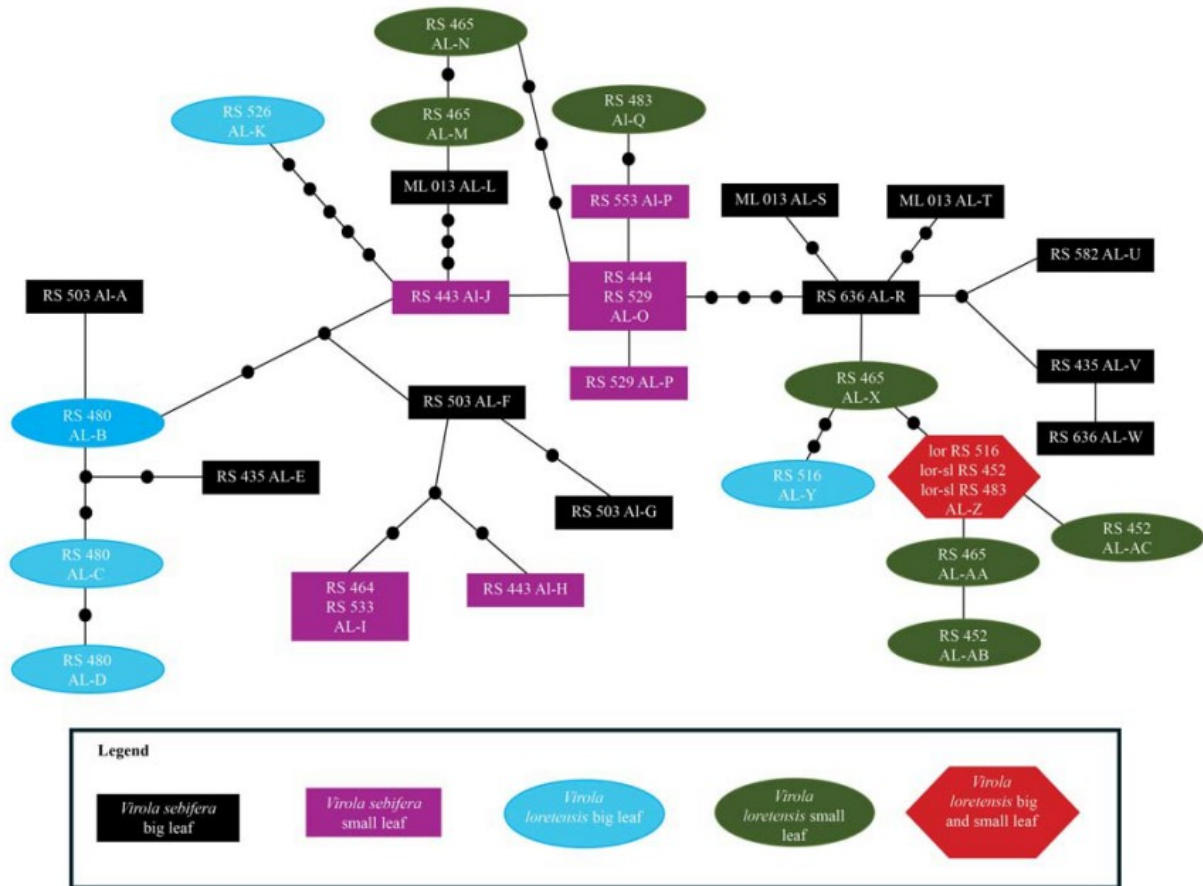


Figure 5: réseau haplotypique formé à partir de 35 séquences chez *Virola sebifera* et *V. lorentensis* (d'après Steeves 2011).

De part mon étude, je peux confirmer qu'il existe des flux de gènes au moins entre certaines espèces du genre *Virola*, et cette affirmation semblerait se confirmer pour d'autres espèces que celles qui m'ont servi de modèles.

Malgré la présence des flux de gènes interspécifiques, les espèces de *V. kwatae* et *V. surinamensis* semblent maintenir leur intégrité à la fois au niveau morphologique mais aussi environnemental. Cela s'expliquerait par le fait que certaines régions du génome, comme celles contenant des gènes impliqués dans l'adaptation au milieu, seraient certainement différenciées entre espèces et imperméables aux flux de gènes (Rieseberg & Burke 2001; Wu 2001). Une forte sélection naturelle doit alors certainement agir à ce niveau. Les autres régions, perméables aux flux géniques interspécifiques montreraient alors des faibles niveaux de différenciation comme cela semble être le cas chez les microsatellites et les marqueurs chloroplastiques utilisés pour cette étude, car ils sont supposés neutres et donc n'interviennent probablement pas dans l'expression des caractères adaptatifs différenciant ces espèces.

A travers ces différents résultats sur le genre *Virola*, cette thèse aura permis l'étude de différentes forces structurant la diversité génétique qui ont un rôle fondamental dans l'évolution des espèces forestières tropicales. Si les flux de gènes tendent à homogénéiser les patrons de diversité génétique neutre, la sélection naturelle se montre assez forte pour contrebalancer ces migrations pour certains loci. En effet, des signatures de divergence moléculaire sont détectées au sein des populations de par l'hétérogénéité du milieu, notamment au milieu local par rapport aux conditions hydriques. Ces signes d'adaptation sont très importants en génétique écologique car des barrières reproductives pourront apparaître avec le temps entre ces populations par le processus continu de la spéciation, et plus particulièrement dans ce cas, de la spéciation écologique, rarement documentée chez les arbres forestiers tropicaux. Pourtant, ce processus a été mis en avant ici entre *V. surinamensis* et *V. kwatae*. Des tests de transplantation réciproques aussi bien que des analyses sur des gènes candidats ou par criblage génomique plus systématique pourraient être envisageables pour conforter ces résultats. La notion de spéciation par l'apparition de barrières reproductives peut être débattue. Effectivement, une faible divergence moléculaire entre espèces sœurs peut être le résultat de flux de gènes interspécifiques comme c'est le cas chez certaines espèces *Virola*. En plus de caractériser les processus évolutifs passés, l'étude de la diversité génétique et plus particulièrement de l'adaptation, dans le contexte des changements globaux, a une grande importance quant au devenir des populations d'arbres qui constituent la composante dominante en termes de biomasse et surtout de capacité d'influencer la structure de l'écosystème forestier. L'étude d'un changement immédiat du milieu, par exploitation forestière de faible intensité, n'indique pas ici de conséquences dramatiques sur la diversité génétique de la population de *V. michelii*, du moins à court terme. Bien évidemment, comme il est difficile de prédire les conséquences des changements environnementaux et leur intensité, on ne peut pas tirer de conclusions, qui seraient trop hâtives, à partir de ces résultats préliminaires.

La diversité actuelle des organismes résulte de multiples interactions entre les espèces et leur environnement au cours du temps. Comprendre les différents processus évolutifs passés et actuels menant à cette diversité permet de mieux aborder son devenir. En particulier, l'étude de l'adaptation fournit une aide précieuse pour la conservation des écosystèmes, des espèces, des populations, ou même de gènes afin de préserver le maximum de diversité dans la nature.

BIBLIOGRAPHIE

- Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* **1**, 95-111.
- Aizen MA, Feinsinger P (1994) Habitat fragmentation, native insect pollinators, and feral honey bees in Argentine 'Chaco Serrano'. *Ecological Applications* **4**, 378-392.
- Aldrich PR, Hamrick JL (1998) Reproductive dominance of pasture trees in a fragmented tropical forest mosaic. *Science* **281**, 103-105.
- Alexandersson E, Fraysse L, Sjövall-Larsen S, *et al.* (2005) Whole gene family expression and drought stress regulation of aquaporins. *Plant Molecular Biology* **59**, 469-484.
- Altinkut A, Kazan K, Gozukirmizi N (2003) AFLP marker linked to water-stress-tolerant bulks in barley (*Hordeum vulgare L.*). *Genetics and Molecular Biology* **26**, 77-82.
- Anderson E, Hubricht L (1938) Hybridization in Tradescantia. III. The evidence for introgressive hybridization. *American Journal of Botany* **25**, 396-402.
- Anderson E, Stebbins GL, Jr. (1954) Hybridization as an evolutionary stimulus. *Evolution* **8**, 378-388.
- Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**, 1217-1229.
- Andrews C (2010) The Hardy-Weinberg principle. *Nature Education Knowledge* **1**, 65.
- Asner GP, Alencar A (2010) Drought impacts on the Amazon forest: the remote sensing perspective. *New Phytologist* **187**, 569-578.
- Audigeos D (2010) *Relations entre diversité génétique et environnement: quels sont les processus évolutifs mis en jeu ? Cas d'une espèce d'arbre tropical: Eperua falcata Aublet*, Université des Antilles et de la Guyane.
- Audigeos D, Brousseau L, Traissac S, Scotti-Saintagne C, Scotti I (in prep) Is divergent selection responsible for local adaptation to micro-environmental gradients in tropical rainforests?
- Audigeos D, Buonamici A, Belkadi L, *et al.* (2010) Aquaporins in the wild: Natural genetic diversity and selective pressure in the PIP gene family in five Neotropical tree species. *BMC Evolutionary Biology* **10**.
- Avise JC (2000) *Phylogeographie: the history and formation of species* Harvard University Press, Cambridge, MA, USA.
- Avise JC (2004) *Molecular markers, natural history, and evolution* Sinauer Associates, Sunderland, MA.
- Baack EJ, Whitney KD, Rieseberg LH (2005) Hybridization and genome size evolution: timing and magnitude of nuclear DNA content increases in Helianthus homoploid hybrid species. *New Phytologist* **167**, 623-630.
- Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology* **13**, 729-744.
- Baltzer JL, Thomas SC, Nilus R, Burslem DFRP (2005) Edaphic specialization in tropical trees: physiological correlates and responses to reciprocal transplantation. *Ecology* **86**, 3063-3077.

- Bamshad M, Wooding SP (2003) Signatures of natural selection in the human genome. *Nature Reviews Genetics* **4**, 99-111.
- Baraloto C, Morneau F, Bonal D, Blanc L, Ferry b (2007) Seasonal water stress tolerance and habitat associations within four neotropical tree genera. *Ecology* **88**, 478-489.
- Barluenga M, Stölting KN, Salzburger W, Muschick M, Meyer A (2006) Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature* **439**, 719-723.
- Barton NH (2001) The role of hybridization in evolution. *Molecular Ecology* **10**, 551-568.
- Barton NH, Mallet J (1996) Natural selection and random genetic drift as causes of evolution on islands. *Philosophical Transactions: Biological Sciences* **351**, 785-795.
- Bawa KS, Seidler R (1998) Natural forest management and conservation of biodiversity in tropical forests. *Conservation Biology* **12**, 46-55.
- Bayes M, Price M (1763) *An essay towards solving a problem in the doctrine of chances*. Philosophical Transactions (1683-1775).
- Beaumont M (2005) Adaptation and speciation: what can F_{ST} tell us? *Trends in Ecology and Evolution* **20**, 435 - 440.
- Beaumont M, Nichols R (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London series B: Biological Sciences* **263**, 1619 - 1626.
- Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics* **5**, 251-261.
- Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian Computation in Population Genetics. *Genetics* **162**, 2025-2035.
- Bensch S, Akesson M (2005) Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology* **14**, 2899-2914.
- Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology* **19**, 2609-2625.
- Birky Jr CW, Fuerst P, Maruyama T (1989) Organelle gene diversity under migration, mutation, and drift: Equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics* **121**, 613-627.
- Blanc L, Flores O, Molino JF, Gourlet-Fleury S, Sabatier D (2003) Diversity and groupings of arborescent species in the forests of French Guiana. *Revue forestière française* **55**, 131-146.
- Bohonak AJ (1999) Dispersal, gene flow, and population structure. *The Quarterly Review of Biology* **74**, 21-45.
- Bonin A (2008) Population genomics: a new generation of genome scans to bridge the gap with functional genomics. *Molecular Ecology* **17**, 3583-3584.
- Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution* **23**, 773-783.
- Brideau NJ, Flores HA, Wang J, *et al.* (2006) Two Dobzhansky-Muller genes interact to cause hybrid lethality in *drosophila*. *Science* **314**, 1292-1295.
- Bryja J, Charbonnel N, Berthier K, Galan M, Cosson JF (2007) Density-related changes in selection pattern for major histocompatibility complex genes in fluctuating populations of voles. *Molecular Ecology* **16**, 5084-5097.
- Callahan HS, Pigliucci M, Schlichting CD (1997) Developmental phenotypic plasticity: Where ecology and evolution meet molecular biology. *BioEssays* **19**, 519-525.
- Cantet L (2004) *Prédiction de l'engorgement hydrique de surface par les cortèges floristiques en forêt tropicale humide (Guyane française)*, Université Henri Poincaré.

- Caron H (2000) *Organisation et dynamique de la diversité génétique de cinq espèces arborées de la forêt guyanaise*, Université de Montpellier II.
- Cascante A, Quesada M, Lobo JJ, Fuchs EA (2002) Effects of dry tropical forest fragmentation on the reproductive success and genetic structure of the tree *Samanea saman*. *Conservation Biology* **16**, 137-147.
- Cavender-Bares J, Pahlich A (2009) Molecular, morphological, and ecological niche differentiation of sympatric sister oak species, *Quercus virginiana* and *Q. geminata* (Fagaceae). *American Journal of Botany* **96**, 1690-1702.
- Chen L, Yang G (2009) A genomic scanning using AFLP to detect candidate loci under selection in the finless porpoise (*Neophocaena phocaenoides*). *Genes & Genetic Systems* **84**, 307-313.
- Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657-1659.
- Corander J, Waldmann P, Sillanpää MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics* **163**, 367-374.
- Cordell S, Goldstein G, Mueller-Dombois D, Webb D, Vitousek PM (1998) Physiological and morphological variation in *Metrosideros polymorpha*, a dominant Hawaiian tree species, along an altitudinal gradient: the role of phenotypic plasticity. *Oecologia* **113**, 188-196.
- Cornuet J, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**, 2001 - 2014.
- Coyer JA, Hoarau G, Pearson G, et al. (2011) Genomic scans detect signatures of selection along a salinity gradient in populations of the intertidal seaweed *Fucus serratus* on a 12 km scale. *Marine Genomics* **4**, 41-49.
- Coyne JA, Orr HA (2004) *Speciation* Sinauer Associates Inc, Sunderland, Massachusetts.
- Coyne JA, Price TD (2000) Little evidence for sympatric speciation in island birds. *Evolution* **54**, 2166-2171.
- Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution* **25**, 410-418.
- Cunningham SA, Summerhayes B, Westoby M (1999) Evolutionary divergences in leaf structure and chemistry, comparing rainfall and soil nutrient gradients. *Ecological Monographs* **69**, 569-588.
- Darwin C (1859) *The Origin of Species* John Murray, London.
- Degen B, Caron H, Bandou E, et al. (2001) Fine-scale spatial genetic structure of eight tropical tree species as analysed by RAPDs. *Heredity* **87**, 497-507.
- Dick CW (2001) Genetic rescue of remnant tropical trees by an alien pollinator. *Proceedings: Biological Sciences* **268**, 2391-2396.
- Dick CW, Etchelecu G, Austerlitz F (2003) Pollen dispersal of tropical trees (*Dinizia excelsa*: Fabaceae) by native insects and African honeybees in pristine and fragmented Amazonian rainforest. *Molecular Ecology* **12**, 753-764.
- Dieckmann U, Doebeli M (1999) On the origin of species by sympatric speciation. *Nature* **400**, 354-357.
- Dobzhansky T (1937) *Genetics and the origin of species* Columbia University Press, New York.
- Doyle JA, Manchester SR, Sauquet H (2008) A seed related to Myristicaceae in the Early Eocene of southern England. *Systematic Botany* **33**, 636-646.
- Draheim H, Cui M, Dick CW (2009) Characterization of 14 microsatellite DNA markers for the tropical forest tree *Virola surinamensis* (Rol.) Warb. (Myristicaceae). *Molecular Ecology Resources* **9**, 1386-1388.

- Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* **22**, 1185-1192.
- Duminil J, Caron H, Scotti I, Cazal S-O, Petit RJ (2006) Blind population genetics survey of tropical rainforest trees. *Molecular Ecology* **15**, 3505-3513.
- Eckert AJ, Bower AD, González-Martínez SC, *et al.* (2010) Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Molecular Ecology* **19**, 3789-3805.
- Ellis AG, Weis AE (2006) Coexistence and differentiation of ‘flowering stones’: the role of local adaptation to soil microenvironment. *Journal of Ecology* **94**, 322-335.
- Ennos RA, Sinclair WT, Hu XS, Langdon A (1999) *Using organelle markers to elucidate the history, ecology and evolution of plant populations* Taylor & Francis, London.
- Ewens W (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3**, 87-112.
- Excoffier L, Estoup A, Cornuet J-M (2005) Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* **169**, 1727-1738.
- Excoffier L, Novembre J, Schneider S (2000) Computer note. SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *Journal of Heredity* **91**, 506-509.
- Ferry B, Freycon V, Paget D (2003) Genèse et fonctionnement hydrique des sols sur socle cristallin en Guyane. *Revue forestière française numéro spécial*, 37-59.
- Fischer MC, Foll M, Excoffier L, Heckel G (2011) Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). *Molecular Ecology* **20**, 1450-1462.
- Fisher RA (1930) *The genetical theory of natural selection* Clarendon Press.
- Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977-993.
- Forget P-M, Cuijpers L (2008) Survival and scatterhoarding of frugivores-dispersed seeds as a function of forest disturbance. *Biotropica* **40**, 380-385.
- Forget P-M, Milleron T, Feer F, Henry O, Dubost G (2000) Effects of dispersal pattern and mammalian herbivores on seedling recruitment for *Virola michelii* (Myristicaceae) in French Guiana. *Biotropica* **32**, 452-462.
- Frankham R (2003) Genetics and conservation biology. *Comptes Rendus Biologies* **326**, 22-29.
- Frankham R (2005) Genetics and extinction. *Biological Conservation* **126**, 131-140.
- Fredericksen TS, Mostacedo B (2000) Regeneration of timber species following selection logging in a Bolivian tropical dry forest. *Forest Ecology and Management* **131**, 47-55.
- Fu YX (1996) New statistical tests of neutrality for DNA samples from a population. *Genetics* **143**, 557-570.
- Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* **133**, 693-709.
- Fuchs EJ, Lobo JA, Quesada M (2003) Effects of forest fragmentation and flowering phenology on the reproductive success and mating patterns of the tropical dry forest tree *Pachira quinata*. *Conservation Biology* **17**, 149-157.
- Gaggiotti OE (2011) Making inferences about speciation using sophisticated statistical genetics methods: look before you leap. *Molecular Ecology* **20**, 2229-2232.
- Gandon S, Michalakis Y (2002) Local adaptation, evolutionary potential and host–parasite coevolution: interactions between migration, mutation, population size and generation time. *Journal of Evolutionary Biology* **15**, 451-462.

- Gaut BS, Le Thierry d'Ennequin M, Peek AS, Sawkins MC (2000) Maize as a model for the evolution of plant nuclear genomes. *Proceedings of the National Academy of Sciences* **97**, 7008-7015.
- Gerber S, Chabrier P, Kremer A (2003) FaMoZ: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. *Molecular Ecology Notes* **3**, 479-481.
- Gielly L, Taberlet P (1994) The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences. *Molecular Biology and Evolution* **11**, 769-777.
- Gompert Z, Alex Buerkle C (2010) introgress: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources* **10**, 378-384.
- Gould SJ, Lewontin RC (1979) The spandrels of San Marco and the panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London* **205**, 581-598.
- Grant BR, Grant PR (1979) Darwin's finches: Population variation and sympatric speciation. *Proceedings of the National Academy of Sciences* **76**, 2359-2363.
- Gregerson RG, Cameron L, McLean M, Dennis P, Strommer J (1993) Structure, Expression, Chromosomal Location and Product of the Gene Encoding ADH2 in *Petunia*. *Genetics* **133**, 999-1007.
- Hahn MW, Rausher MD, Cunningham CW (2002) Distinguishing between selection and population expansion in an experimental lineage of bacteriophage T7. *Genetics* **161**, 11-20.
- Hamrick JL, Murawski DA, Nason JD (1993) The influence of seed dispersal mechanisms on the genetic structure of tropical tree populations. *Plant Ecology* **107-108**, 281-297.
- Hardy GH (1908) Mendelian proportions in a mixed population. *Science* **28**, 49-50.
- Hardy OJ (2003) Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers. *Molecular Ecology* **12**, 1577-1588.
- Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity* **83**, 145-154.
- Hedrick PW (1986) Genetic polymorphism in heterogeneous environments: a decade later. *Annual Review of Ecology and Systematics* **17**, 535-566.
- Hey J (2001) The mind of the species problem. *Trends in ecology & evolution (Personal edition)* **16**, 326-329.
- Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. *Current Opinion in Genetics & Development* **16**, 592-596.
- Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution* **27**, 905-920.
- Holbrook KM, Loiselle BA (2009) Dispersal in a Neotropical tree, *Virola flexuosa* (Myristicaceae): Does hunting of large vertebrates limit seed removal? *Ecology* **90**, 1449-1455.
- Holbrook KM, Loiselle BA, Clark AM (2006) Eight polymorphic microsatellite loci for a Neotropical nutmeg, *Virola flexuosa* (Myristicaceae). *Molecular Ecology Notes* **7**, 248-250.
- Hoorn C, Wesselingh FP, ter Steege H, *et al.* (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* **330**, 927-931.
- Howe H (1981) Dispersal of a Neotropical nutmeg (*Virola sebifera*) by birds. *The Auk* **98**, 88-98.

- Hudson RR, Bailey K, Skarecky D, Kwiatowski J, Ayala FJ (1994) Evidence for positive selection in the superoxide dismutase (Sod) region of *Drosophila melanogaster*. *Genetics* **136**, 1329-1340.
- Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153-159.
- IPCC (2007) IPCC Fourth Assessment Report: Climate Change 2007 (AR4). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Janovec JP, Neill AK (2002) Studies of the Myristicaceae: an overview of the *Compsonaura atopa* complex, with descriptions of new species from Colombia. *Brittonia* **54**, 251-261.
- Joost S, Bonin A, Bruford M, *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology* **16**, 3955 - 3969.
- Jump AS, Hunt JM, Martinez-Izquierdo JA, Penuelas J (2006) Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. *Molecular Ecology* **15**, 3469-3480.
- Kawanishi K, Hashimoto Y (1987) Long chain esters of *Viola* species. *Phytochemistry* **26**, 749-752.
- Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters* **7**, 1225-1241.
- Kimura M (1968) Evolutionary rate at the molecular level. *Nature* **217**, 624-626.
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**, 111-120.
- Kimura M (1983) *The neutral theory of molecular evolution* Cambridge University Press.
- Kingman JFC (1982) On the genealogy of large populations. *Journal of Applied Probability* **19**, 27-43.
- Kingman JFC (2000) Origins of the Coalescent: 1974-1982. *Genetics* **156**, 1461-1463.
- Kuchma O, Finkeldey R (2011) Evidence for selection in response to radiation exposure: *Pinus sylvestris* in the Chernobyl exclusion zone. *Environmental Pollution* **159**, 1606-1612.
- Lande R (1976) Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**, 314-334.
- Lechowicz MJ, Bell G (1991) The Ecology and genetics of fitness in forest plants. II. Microspatial heterogeneity of the edaphic environment. *Journal of Ecology* **79**, 687-696.
- Lee SL (2000) Mating system parameters of *Dryobalanops aromatica* Gaertn. f. (Dipterocarpaceae) in three different forest types and a seed orchard. *Heredity* **85**, 338-345.
- Leigh EG, Davidar P, Dick CW, *et al.* (2004) Why do some tropical forests have so many species of trees? *Biotropica* **36**, 447-473.
- Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution* **17**, 183-189.
- Lepais O, Petit RJ, Guichoux E, *et al.* (2009) Species relative abundance and direction of introgression in oaks. *Molecular Ecology* **18**, 2228-2242.
- Levin DA (2004) The ecological transition in speciation. *New Phytologist* **161**, 91-96.
- Lewontin R, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**, 175 - 195.
- Lexer C, Fay M (2005) Adaptation to environmental stress: a rare or frequent driver of speciation? *Journal of Evolutionary Biology* **18**, 893-900.

- 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. *Molecular Ecology* **14**, 1045-1057.
- Li W, Fu R, Juarez RIN, Fernandes K (2008) Observed change of the standardized precipitation index, its potential cause and implications to future climate change in the Amazon region. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 1767-1772.
- Li WH, Wu CI, Luo CC (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Molecular Biology and Evolution* **2**, 150-174.
- Lowe A, Harris S, Ashton P (2004) *Ecological genetics: design, analysis and application* Oxford: Blackwell Publishing.
- Luikart G, England P, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature reviews* **4**, 981 - 994.
- Ma F, Zhao C, Milne R, *et al.* (2010) Enhanced drought-tolerance in the homoploid hybrid species *Pinus densata*: implication for its habitat divergence from two progenitors. *New Phytologist* **185**, 204-216.
- Macía M (2011) Spatial distribution and floristic composition of trees and lianas in different forest types of an Amazonian rainforest. *Plant Ecology* **212**, 1159-1177.
- Malhi Y, Roberts JT, Betts RA, *et al.* (2008) Climate change, deforestation, and the fate of the Amazon. *Science* **319**, 169-172.
- Malhi Y, Wright J (2004) Spatial patterns and recent trends in the climate of tropical rainforest regions. *Philosophical Transactions of the Royal Society of London* **359**, 311-329.
- Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology and Evolution* **20**, 229-237.
- Mallet J (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society: Biological Sciences* **363**, 2971-2986.
- Manel S, Joost S, Epperson BK, *et al.* (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology* **19**, 3760-3772.
- Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**, 189-197.
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 209 - 220.
- Mátyás C (1996) Climatic adaptation of trees: rediscovering provenance tests. *Euphytica* **92**, 45-54.
- Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetics Research* **23**, 23-35.
- Mayr E (1942) *Systematics and the origin of species* Columbia University Press, New York.
- McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652-654.
- Meir P, Ian Woodward F (2010) Amazonian rainforests and drought: response and vulnerability. *New Phytologist* **187**, 553-557.
- Mes TM (2004) Purifying selection and demographic expansion affect sequence diversity of the ligand-binding domain of a glutamate-gated chloride channel gene of *Haemonchus placei*. *Journal of Molecular Evolution* **58**, 466-478.

- Miller NJ, Ciosi M, Sappington TW, *et al.* (2007) Genome scan of *Diabrotica virgifera virgifera* for genetic variation associated with crop rotation tolerance. *Journal of Applied Entomology* **131**, 378-385.
- Miller TR (1999) *The use of the genus Virola as a hallucinogen In South America* Illinois University Carbondale.
- Minin VN, Bloomquist EW, Suchard MA (2008) Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution* **25**, 1459-1471.
- Moran PAP (1950) Notes on continuous stochastic phenomena. *Biometrika* **37**, 17-23.
- Morjan CL, Rieseberg LH (2004) How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology* **13**, 1341-1356.
- Morneau F (2007) *Effets d'un gradient d'engorgement sur la structure et la dynamique d'une forêt tropicale humide (Paracou, Guyane française)*, ENGREF Nancy.
- Mueller UG, Wolfenbarger LL (1999) AFLP genotyping and fingerprinting. *Trends in Ecology & Evolution* **14**, 389-394.
- Mullen LM, Hoekstra HE (2008) Natural selection along an environmental gradient: a classic cline in mouse pigmentation. *Evolution* **62**, 1555-1570.
- Narum SR, Hess JE (2011) Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources* **11**, 184-194.
- Nei M (1973) Analysis of gene diversity in subdivided populations. *PNAS* **70**, 3321-3323.
- Nei M, Chesser R (1983) Estimation of fixation indices and gene diversities. *Annals of human genetics* **47**, 253 - 259.
- Neigel JE (1997) *A comparison of alternative strategies for estimating gene flow from genetic markers* Annual Reviews, Palo Alto, CA, United States of America.
- Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics* **39**, 197-218.
- Nordborg M (2004) Coalescent Theory. In: *Handbook of Statistical Genetics*. John Wiley & Sons, Ltd.
- Nosil P, Egan SP, Funk DJ, Hoekstra H (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: 'Isolation by Adaptation' and multiple roles for divergent selection. *Evolution* **62**, 316-336.
- Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* **18**, 375-402.
- Oetjen K, Reusch TBH (2007) Genome scans detect consistent divergent selection among subtidal vs. intertidal populations of the marine angiosperm *Zostera marina*. *Molecular Ecology* **16**, 5156-5157.
- Orr HA (2005) The genetic theory of adaptation: A brief history. *Nature Reviews Genetics* **6**, 119-127.
- Palmé AE (2002) Chloroplast DNA variation, postglacial recolonization and hybridization in hazel, *Corylus avellana*. *Molecular Ecology* **11**, 1769-1779.
- Palmer JD, Jansen RK, Michaels HJ, Chase MW, Manhart JR (1988) Chloroplast DNA Variation and Plant Phylogeny. *Annals of the Missouri Botanical Garden* **75**, 1180-1206.
- Paris M, Boyer S, Bonin A, *et al.* (2010) Genome scan in the mosquito *Aedes rusticus*: population structure and detection of positive selection after insecticide treatment. *Molecular Ecology* **19**, 325-337.
- Parisod C, Christin P-A (2008) Genome-wide association to fine-scale ecological heterogeneity within a continuous population of *Biscutella laevigata* (Brassicaceae). *New Phytologist* **178**, 436-447.

- Peleg ZVI, Saranga Y, Krugman T, *et al.* (2008) Allelic diversity associated with aridity gradient in wild emmer wheat populations. *Plant, Cell & Environment* **31**, 39-49.
- Pérez-Figueroa A, Garc-Pereira MJ, Saura M, Rolan-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology* **23**, 2267-2276.
- Pernès J, Lourd M (1984) *Organisation des complexes d'espèces* Agence de coopération et technique, Paris.
- Petit RJ (2004) Biological invasions at the gene level. *Diversity and Distributions* **10**, 159-165.
- Petit RJ, Duminil J, Fineschi S, *et al.* (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology* **14**, 689-701.
- Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. In: *Annual Review of Ecology Evolution and Systematics*, pp. 187-214. Annual Reviews, Palo Alto.
- Phillips OL, Aragão LE, Lewis SL, *et al.* (2009) Drought Sensitivity of the Amazon Rainforest. *Science* **323**, 1344-1347.
- Pritchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945 - 959.
- Puig H (2001) *La forêt tropicale humide* Belin, Paris.
- Queenborough S, Burslem D, Garwood N, Valencia R (2007) Habitat niche partitioning by 16 species of Myristicaceae in Amazonian Ecuador. *Plant Ecology* **192**, 193-207.
- Quesada H, Posada D, Caballero A, Morán P, Rolán-Alvarez E (2007) Phylogenetic evidence for multiple sympatric ecological diversification in a marine snail. *Evolution* **61**, 1600-1612.
- Quesada M, Stoner KE, Rosas-Guerrero V, Palacios-Guevara C, Lobo JA (2003) Effects of habitat disruption on the activity of nectarivorous bats (Chiroptera: Phyllostomidae) in a dry tropical forest: implications for the reproductive success of the neotropical tree *Ceiba grandiflora*. *Oecologia* **135**, 400-406.
- Rambal S, Leterme J (1987) *Changes in aboveground structure and resistances to water uptake in Quercus coccifera along a rainfall gradient* Plenum, New York,.
- Rieseberg LH, Archer MA, Wayne RK (1999) Transgressive segregation, adaptation and speciation. *Heredity* **83**, 363-372.
- Rieseberg LH, Burke JM (2001) A genic view of species integration. *Journal of Evolutionary Biology* **14**, 883-886.
- Rieseberg LH, Carney SE (1998) Plant hybridization. *New Phytologist* **140**, 599-624.
- Rieseberg LH, Soltis DE (1991) Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* **5**, 65-84.
- Rieseberg LH, Willis J (2007) Plant speciation. *Science* **317**, 910 - 914.
- Rieseberg LH, Wood TE, Baack EJ (2006) The nature of plant species. *Nature* **440**, 524-527.
- Roda F (2011) *The genomic basis of ecological speciation in plants*, University of Queensland.
- Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics* **3**, 380-390.
- Rundle HD, Nagel L, Boughman JW, Schluter D (2000) Natural selection and parallel speciation in sympatric Sticklebacks. *Science* **287**, 306-308.
- Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters* **8**, 336-352.
- Russo SE, Augspurger CK (2004) Aggregated seed dispersal by spider monkeys limits recruitment to clumped patterns in *Virola calophylla*. *Ecology Letters* **7**, 1058-1067.

- Sabatier D (1997) Description and biology of a new species of *Virola* (Myristicaceae) from Guyana. *Adansonia (Paris)* **19**, 273-278.
- Sabatier D, Grimaldi M, Prevost M-F, *et al.* (1997) The influence of soil cover organization on the floristic and structural heterogeneity of a Guianan rainforest. *Plant Ecology* **131**, 81-108.
- Sambatti JBM, Rice KJ (2006) Local adaptation, patterns of selection, and gene flow in the Californian serpentine sunflowers (*Helianthus exilis*). *Evolution* **60**, 696-710.
- Sauquet H, Doyle JA, Scharaschkin T, *et al.* (2003) Phylogenetic analysis of Magnoliales and Myristicaceae based on multiple data sets: implications for character evolution. *Botanical Journal of the Linnean Society* **142**, 125-186.
- Savolainen O, Bokma F, Garcia-Gil R, Komulainen P, Repo T (2004) Genetic variation in cessation of growth and frost hardiness and consequences for adaptation of *Pinus sylvestris* to climatic changes. *Forest Ecology and Management* **197**, 79-89.
- Savolainen O, Pyhajarvi T, Knurr T (2007) Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics* **38**, 595-619.
- Savolainen V, Anstett M-C, Lexer C, *et al.* (2006) Sympatric speciation in palms on an oceanic island. *Nature* **441**, 210-213.
- Schliewen U, Rassmann K, Markmann M, *et al.* (2001) Genetic and ecological divergence of a monophyletic cichlid species pair under fully sympatric conditions in Lake Ejagham, Cameroon. *Molecular Ecology* **10**, 1471-1488.
- Schluter D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution* **16**, 372-380.
- Schluter D (2009) Evidence for ecological speciation and its alternative. *Science* **323**, 737-741.
- Scotti-Saintagne C, Mariette S, Porth I, *et al.* (2004) Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics* **168**, 1615-1626.
- Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution* **19**, 198-207.
- Shaw J, Lickey E, Beck J, *et al.* (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* **92**, 142-166.
- Simpson EH (1949) Measurement of diversity. *Nature* **163**, 688-688.
- Slatkin M (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics* **16**, 393-430.
- Smouse P, Long J, Sokal R (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology* **35**, 627-632.
- Smouse P, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* **82**, 561-573.
- Solomon S, Plattner GK, Knutti R, Friedlingstein P (2009) Irreversible climate change due to carbon dioxide emissions. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 1704-1709.
- Stace C (1975) *Hybridization and the flora of the British Isles* Academic Press, London.
- Steeves RAD (2011) *An Intrageneric and Intraspecific Study of Morphological and Genetic Variation in the Neotropical Compsoeura and Virola (Myristicaceae)* University of Guelph.
- Stockwell CA, Hendry AP, Kinnison MT (2003) Contemporary evolution meets conservation biology. *Trends in Ecology & Evolution* **18**, 94-101.
- Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* **14**, 671-688.

- Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437-460.
- Takahata N, Satta Y, Klein J (1992) Polymorphism and balancing selection at major histocompatibility complex Loci. *Genetics* **130**, 925-938.
- Templeton AR (1980) The theory of speciation via the founder principle. *Genetics* **94**, 1011-1038.
- Tice KA, Carlon DB (2011) Can AFLP genome scans detect small islands of differentiation? The case of shell sculpture variation in the periwinkle *Echinolittorina hawaiiensis*. *Journal of Evolutionary Biology* **24**, 1814-1825.
- Tieleman BI, Williams JB, Bloomer P (2003) Adaptation of metabolism and evaporative water loss along an aridity gradient. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**, 207-214.
- Tsitrona A, Kirkpatrick M, Levin DA (2003) A model for chloroplast capture. *Evolution* **57**, 1776-1782.
- Turelli M, Barton NH, Coyne JA (2001) Theory and speciation. *Trends in Ecology & Evolution* **16**, 330-343.
- Unwin GL, Kriedemann PE (1990) Drought tolerance and rainforest tree growth on a North Queensland rainfall gradient. *Forest Ecology and Management* **30**, 113-123.
- Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution* **16**, 381-390.
- Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences* **106**, 9939-9946.
- Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**, 1811 - 1823.
- Wang R-L, Stec A, Hey J, Lukens L, Doebley J (1999) The limits of selection during maize domestication. *Nature* **398**, 236-239.
- Watterson GA (1974) The sampling theory of selectively neutral alleles. *Advances in Applied Probability* **6**, 463-488.
- Weinberg W (1908) Über den Nachweis der Vererbung beim Menschen. *Jahreshefte des Vereins Väterländische Naturkundler in Württemberg* **64**, 369-382.
- Weir B, Cockerham C (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358 - 1370.
- White GM, Boshier DH, Powell W (1999) Genetic variation within a fragmented population of *Swietenia humilis* Zucc. *Molecular Ecology* **8**, 1899-1909.
- White GM, Boshier DH, Powell W (2002) Increased pollen flow counteracts fragmentation in a tropical dry forest: An example from *Swietenia humilis* Zuccarini. *Proceedings of the National Academy of Sciences of the USA* **99**, 2038-2042.
- Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{st} \neq 1/(4Nm+1)$. *Heredity* **82**, 117-125.
- Williams J, Auge H, Maron J (2008) Different gardens, different results: native and introduced populations exhibit contrasting phenotypes across common gardens. *Oecologia* **157**, 239-248.
- Williams L, Oleksiak M (2008) Signatures of selection in natural populations adapted to chronic pollution. *BMC Evolutionary Biology* **8**, 282.
- Wilson TK (2004) *Myristicaceae*. In *Flowering Plants of the Neotropics* N. Smith, S.A. Mori, A. Henderson, D.W. Stevenson & S.V. Heald edn. New York Botanical Garden & Princeton University Press, Princeton.
- Wolfe KH, Li W-H, Sharp PM (1987) Rates of Nucleotide Substitution Vary Greatly among Plant Mitochondrial, Chloroplast, and Nuclear DNAs. *Proceedings of the National Academy of Sciences* **84**, 9054-9058.

- Wright S (1929) Fisher's theory of dominance. *American Naturalist* **63**, 274-279.
- Wright S (1931) Evolution in mendelian populations. *Genetics* **16**, 97-159.
- Wright S (1943) Isolation by distance. *Genetics* **28**, 114-138.
- Wright S (1951) The genetical structure of populations. *Ann. Eugen. (Lond.)* **15**, 323-354.
- Wright S (1969) *Evolution and the genetics of populations: a treatise in four volumes* Chicago Press, USA.
- Wu C-I (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**, 851-865.
- Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution* **17**, 32-43.
- Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology & Evolution* **11**, 413-418.