



**HAL**  
open science

# Dynamique d'hybridation dans le complexe d'espèces des chênes blancs européens (Chênes pédonculés - *Quercus robur* L., sessiles - *Q. petraea* (Matt.) Liebl., pubescents - *Q. pubescens* Willd. et tauzins - *Q. pyrenaica* Willd.)

Olivier Lepais

## ► To cite this version:

Olivier Lepais. Dynamique d'hybridation dans le complexe d'espèces des chênes blancs européens (Chênes pédonculés - *Quercus robur* L., sessiles - *Q. petraea* (Matt.) Liebl., pubescents - *Q. pubescens* Willd. et tauzins - *Q. pyrenaica* Willd.). Sciences du Vivant [q-bio]. Université des Sciences et Technologies (Bordeaux 1), 2008. Français. NNT: . tel-02821939

**HAL Id: tel-02821939**

**<https://hal.inrae.fr/tel-02821939>**

Submitted on 6 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N° d'ordre :

# THÈSE

présentée à

**L'UNIVERSITÉ BORDEAUX 1**

ÉCOLE DOCTORALE SCIENCES ET ENVIRONNEMENTS (ED304)

par **Olivier LEPAIS**

POUR OBTENIR LE GRADE DE

**DOCTEUR**

SPÉCIALITÉ : ÉCOLOGIE ÉVOLUTIVE, FONCTIONNELLE ET DES COMMUNAUTÉS

Sujet :

## **Dynamique d'hybridation dans le complexe d'espèces des chênes blancs européens**

Chênes pédonculés - *Quercus robur* L., sessiles - *Q. petraea* (Matt.) Liebl.,  
pubescents - *Q. pubescens* Willd. et tautzins - *Q. pyrenaica* Willd.

Soutenue le 26 septembre 2008

Après avis de :

M. Joël CUGUEN	Professeur, Université de Lille 1	Rapporteur
M. Christophe THÉBAUD	Professeur, Université Paul Sabatier Toulouse	Rapporteur

Devant la commission d'examen formée de :

M. Joël CUGUEN	Professeur, Université de Lille 1	Rapporteur
M. Christophe THÉBAUD	Professeur, Université Paul Sabatier Toulouse	Rapporteur
Mme Marie-Laure DESPREZ-LOUSTAU	Directrice de recherche, INRA Bordeaux	Examinatrice
M. Jean-Pierre RENAUDIN	Professeur, Université Bordeaux 1	Examineur
Mme Réjane STREIFF	Chargée de recherche, INRA Montpellier	Examinatrice
Mme Sophie GERBER	Chargée de recherche, INRA Bordeaux (Directrice de thèse)	Examinatrice



**UMR INRA 1202 BIOGECO**  
**Equipe de génétique**  
69 route d'Arcachon  
F-33612 CESTAS Cedex





# Table des matières

<b>Remerciements</b> .....	<b>7</b>
<b>Prologue</b> .....	<b>9</b>
<b>Chapitre 1- Introduction</b> .....	<b>11</b>
1.1. Synthèse bibliographique .....	13
1.1.1. Espèce et spéciation .....	13
1.1.2. Hybridation .....	20
1.1.3. Conclusion .....	25
1.2. Modèle d'étude : le complexe d'espèces des chênes blancs européens .....	26
1.2.1. Présentation du genre <i>Quercus</i> .....	26
1.2.2. Le complexe d'espèces des chênes blancs .....	27
1.3. Objectif de la thèse .....	38
<b>PARTIE A - MÉTHODOLOGIE</b> .....	<b>41</b>
<b>Chapitre 2- Marqueurs génétiques discriminant les espèces de chênes</b> .....	<b>43</b>
2.1. Introduction .....	45
2.2. Mise au point d'une méthode de génotypage rapide .....	46
2.2.1. Sélection de microsatellites différenciant les espèces de chênes .....	46
2.2.2. Sélection de microsatellites sur critères techniques .....	48
2.2.3. Mise au point du protocole .....	50
2.3. Normalisation des données entre projets .....	53
2.4. Calcul du taux d'erreur de génotypage .....	54
2.4.1. De la prise de conscience de l'erreur en génétique des populations .....	54
2.4.2. Estimation du taux d'erreur de génotypage .....	55
2.4.3. Conséquences des erreurs dans les analyses génétiques .....	57
2.5. Conclusion .....	58
<b>Chapitre 3- Tests d'assignation génétique des espèces et des hybrides</b> .....	<b>61</b>
3.1. Introduction .....	63
3.2. Revue des principales méthodes d'assignation .....	64
3.2.1. Introduction .....	64
3.2.2. Classification – assignation supervisée .....	65
3.2.3. Clustering – assignation non supervisée .....	66
3.2.4. Délimitation spatiale des populations .....	66
3.2.5. Détection des hybrides .....	67
3.2.6. Conclusion .....	67
3.3. Comparaison de méthodes d'assignation .....	69
3.3.1. Introduction .....	69
3.3.2. Matériels et méthodes .....	69
3.3.3. Résultats et discussions .....	71
3.3.4. Conclusion .....	73
3.4. Tests de sensibilité pour l'identification de deux espèces .....	76
3.4.1. Introduction .....	76
3.4.2. Matériels et méthodes .....	76
3.4.3. Résultats et discussions .....	81
3.4.4. Conclusion .....	83
3.5. Identification des hybrides au sein d'un complexe d'espèces .....	84
3.5.1. Introduction .....	84
3.5.2. Matériels et méthodes .....	84
3.5.3. Données génétiques .....	84
3.5.4. Résultats et discussions .....	86
3.5.5. Conclusion .....	91
3.6. Conclusion générale .....	92



<b>PARTIE B – DÉLIMITATION DES ESPÈCES, ASSIGNATION GÉNÉTIQUE EN POPULATIONS</b> .....	<b>95</b>
<b>Chapitre 4- Hybridation en populations naturelles</b> .....	<b>97</b>
4.1. Introduction.....	99
4.2. Matériels et méthodes.....	102
4.2.1. Echantillonnages.....	102
4.2.2. Génotypage .....	102
4.2.3. Assignation des individus .....	103
4.2.4. Analyse phylogénétique.....	103
4.2.5. Caractérisation de l'hybridation dans les parcelles .....	103
4.3. Résultats .....	106
4.3.1. Assignation génétique des individus.....	106
4.3.2. Relations génétiques entre espèces, hybrides et populations .....	111
4.3.3. Caractéristiques génétiques des hybrides dans les parcelles.....	112
4.4. Discussion.....	120
4.4.1. Taux d'hybridation comparé entre populations.....	120
4.4.2. Caractéristiques des hybrides .....	121
4.4.3. Hybridation hors zone de sympatrie .....	121
4.4.4. Hybridation et introgression fréquence-dépendante.....	122
<b>Chapitre 5- Conséquences phénotypiques et génétiques de l'hybridation dans une parcelle</b> .....	<b>125</b>
5.1. Introduction - Caractéristiques générales du bois de Briouant.....	127
5.2. Morphologie foliaire .....	137
5.3. Structure génétique.....	142
5.3.1. Méthodes .....	142
5.3.2. Résultats et discussions .....	142
5.4. Structure génétique spatiale fine .....	146
5.4.1. Méthodes .....	146
5.4.2. Résultats et discussions .....	147
5.5. Conclusion.....	151
<b>PARTIE C – SYSTÈME DE REPRODUCTION</b> .....	<b>153</b>
<b>Chapitre 6- Flux de gènes interspécifiques en conditions contrôlées</b> .....	<b>155</b>
6.1. Introduction.....	157
6.2. Matériels et méthodes .....	158
6.2.1. Croisements contrôlés chez les chênes .....	158
6.2.2. Chênes utilisés dans les croisements .....	158
6.2.3. Plan de croisement .....	160
6.2.4. Analyses génétiques.....	161
6.2.5. Analyses statistiques .....	162
6.3. Résultats .....	164
6.3.1. Identification des pères.....	164
6.3.2. Compatibilité génétique .....	165
6.3.3. Compétition pollinique .....	166
6.3.4. Taux de germination et de survie .....	169
6.3.5. Isolement reproducteur.....	171
6.4. Conclusion.....	172
<b>Chapitre 7- Flux de gènes interspécifiques en population naturelle</b> .....	<b>175</b>
7.1. Introduction.....	177
7.2. Matériels et méthodes .....	177
7.2.1. Parcelle étudiée .....	177
7.2.2. Observations de la phénologie du débourrement.....	178
7.2.3. Echantillonnages.....	179
7.2.4. Suivi de germination des glands.....	181
7.2.5. Génotypage des glands et des plantules .....	182
7.2.6. Assignation génétique des glands et des plantules.....	183
7.2.7. Analyse de paternité .....	184
7.2.8. Analyse de parenté .....	184
7.2.9. Identification directe des événements de reproduction .....	185
7.2.10. Corrélation de paternité intra- et inter-familles .....	186
7.3. Résultats .....	186

7.3.1. Phénologie des espèces et des hybrides .....	186
7.3.2. Caractéristiques des descendance étudiées .....	190
7.3.3. Assignment génétique des glands .....	192
7.3.4. Analyse de paternité .....	195
7.3.5. Analyse de parenté .....	202
7.3.6. Synthèse des événements de reproduction détectés .....	205
7.3.7. Corrélations de paternité intra-famille .....	208
7.3.8. Différenciations des nuages polliniques .....	209
7.4. Discussion .....	211
<b>Chapitre 8- Conclusion générale .....</b>	<b>217</b>
8.1. Synthèse des résultats obtenus .....	219
8.1.1. Délimitation des espèces et assignation .....	219
8.1.2. Hybridation – introgression .....	220
8.1.3. Isolement reproducteur .....	221
8.2. Conséquences de l'hybridation sur l'évolution du complexe d'espèces .....	223
<b>Références bibliographiques .....</b>	<b>225</b>
<b>ANNEXES .....</b>	<b>235</b>
ANNEXE 1 - Résultats complets des croisements contrôlés .....	236
ANNEXE 2 - Identification des deux générations d'hybrides .....	238
ANNEXE A – Lepais et <i>al.</i> , <i>Silvae Genetica</i> , 55, 4-5, 2006 .....	240
ANNEXE B - Lepais et <i>al.</i> , <i>Les Actes du BRG</i> , 6, 543-557, 2006. ....	243
ANNEXE C - Viscosi et <i>al.</i> , <i>Plant Biosystems</i> (in press).....	259
ANNEXE D – Rekik et <i>al.</i> , <i>Hortscience</i> (2008) .....	275
ANNEXE E – Terrab et <i>al.</i> , <i>American Journal of Botany</i> (2008).....	277
<b>Résumé .....</b>	<b>279</b>



# Remerciements

Je tiens en premier lieu à remercier Sophie Gerber qui a dirigé cette thèse. Ton soutien au quotidien et ta disponibilité pendant ces quatre années ont été décisifs à l'aboutissement du travail.

Merci à Antoine Kremer, directeur du laboratoire, de m'avoir accueilli au sein de l'unité et de m'avoir permis de concrétiser ce projet de recherche. Tes recommandations et tes remarques lors de la définition du sujet et à plusieurs étapes du projet étaient encourageantes et motivantes.

Un grand merci à Rémy Petit, directeur de l'équipe génétique, pour les nombreuses discussions scientifiques et son intérêt pour ce travail de recherche.

Mes remerciements vont aux membres de mon comité de thèse Christophe Andalo, Stuart Baird, Marie Etienne, Vincent Badeau, qui lors de deux réunions m'ont permis de faire le point et de continuer avec un œil neuf.

Merci aux rapporteurs Joël Cuguen et Christophe Thébaud ainsi qu'aux examinateurs Marie-Laure Desprez-Loustau, Réjane Streiff et Jean-Pierre Renaudin d'avoir accepté d'évaluer ce travail.

J'ai pu préparer ce doctorat grâce à une bourse du ministère (bourse MNRT) de l'école doctorale Sciences et Environnement. Un contrat d'ATER à l'UFR des Sciences Biologies de l'Université Bordeaux 1 m'a permis d'approfondir mon travail une année supplémentaire et de goûter au stimulant métier d'enseignant.

Les diverses étapes des expériences de cette thèse ont été financées par divers projets : Projet BRG, 2004-2006 ("Impact de l'écologie, de l'histoire et du mode de gestion des populations sur la diversité en forêts fragmentées.") ; Projet européen OAKFLOW, 2001-2004 ("Intra & interspecific gene flow in oaks as mechanisms promoting genetic diversity and adaptive potential", coordination A. Kremer.) ; Projet Inter-régional Aquitaine / Midi-Pyrénées, 2004-2005 ("Evolution de la diversité des forêts sous l'effet des changements globaux, changements d'usage et changements climatiques") ; projet ANR QDiv, 2006-2008 ("Quantification of the effects of global change on plant diversity"). Les analyses génétiques ont été effectuées à la plateforme de Génotypage et de Séquençage de Bordeaux (financement du Conseil Régional d'Aquitaine n°20030304002FA et 20040305003FA et de l'Union Européenne FEDER n°2003227).

Cette thèse est le fruit d'un travail collectif à plusieurs étapes du projet.

Je remercie en premier lieu Guy Roussel qui a réalisé l'ensemble des croisements contrôlés et pour toute l'aide en pépinière, au labo graine ou sur le terrain. Ton ingéniosité pare à tous les problèmes et ta vision intime des chênes m'ont ouvert d'autres perspectives. Le dur labeur fut bien agréable.

La connaissance du terrain de l'UMR DYNAFOR (Dynamique forestière dans l'espace rurale) a été décisive dans la réalisation de cette thèse. Merci à Jérôme Willm pour sa participation aux divers échantillonnages à Briouant ainsi que son investissement important dans la cartographie et la collecte de données des parcelles de la Paguères et Briouant. Je tiens également à remercier Maya Gonzalez, Julie Saurel, Laurent Brunel, Alain Cabanette et Gérard Balent. Merci également aux propriétaires de nous avoir permis d'accéder à leur parcelle et aux étudiants de la formation des ingénieurs forestiers (FIF ENGREF).

Un grand merci à Jean Marc Louvet pour sa participation au travail de terrain et son intérêt pour la comparaison des méthodes de cartographie.

Merci aux personnels de l'unité Expérimentale de l'Hermitage pour leur participation aux croisements contrôlés, à travers leur soutien technique et humain (irrigation, fertilisation, préparation du matériel, nacelle...).

Merci à toutes les « petites mains » qui m'ont accompagnée lors de certaines étapes longues et fastidieuses. François Hubert et Delphine Vincent pour l'échantillonnage des jamais trop nombreux

glands. Mimi pour l'aide à l'embranchement de centaines de branches à 20 mètres du sol. Florian Alberto pour les 14 mesures sur les 3300 feuilles. François Hubert pour la mise au point de l'extraction sur glands et les centaines d'extractions d'ADN. Merci à Patrick Léger, Pierre Yves Dumoulin et Franck Salin pour leur participation au génotypage. Merci à Valérie Léger pour l'aide lors du développement technique. Merci à Erwan Guichoux, Florian Alberto et Jessica Lavabre pour le partage des données qui ont permis l'émergence de nouvelles perspectives. Le travail présenté dans cette thèse est enrichi grâce aux expériences de plusieurs stagiaires avec qui j'ai eu le plaisir de travailler : Jessica Lavabre, Phillippe Matter, Séverine Mary.

Merci également à Alexis Ducouso, Pauline Garnier-Géré, Henri Caron, Arndt Hampe, Carlos Magni, Manuela Casasoli, Catherine Bodénès, Sabine Dumas, Marie Pierre Reviron, Jean Pierre Rossi et Alain Franc pour leur conseils et remarques qui m'ont beaucoup aidé. Merci également à l'équipe de choc au secrétariat : Corinne, Véronique, Annie, Danièle et Florence. Merci plus généralement à tous les collègues de Pierroton pour la bonne ambiance et le bouillonnement scientifique.

Je remercie Martin Lascoux pour son invitation et son accueil au laboratoire à Uppsala ainsi que Susanne Gustafsson pour son soutien au labo sur place. Merci Martin de m'avoir initié aux coutumes suédoises et au mode de vie expatrié du chercheur. Ce séjour m'a beaucoup conforté dans ma motivation pour la recherche.

Merci à Erwan, Mimi, Jérôme pour le partage de nombreuses heures au bureau à Jérémie pour ces nombreux conseils.

Merci à Emmanuelle Eveno pour m'avoir entraîné dans l'organisation du 11<sup>ème</sup> *European Meeting of PhD Student of Evolutionary Biology* à Bordeaux (2005). Ce fût une expérience intense et inoubliable tant au niveau organisationnel que scientifique. Merci à Eric Bazin, Emmanuelle Billard, Stéphane De Mita, Julien Dutheil, Baptiste Faure, Matthieu Faure et Delphine Muths pour la longue organisation à distance et la sympathique semaine passée dans le feu de l'action. Un grand merci à Florence pour son implication dans le casse-tête de la gestion financière de ce colloque.

Merci aux collègues de l'université pour leur aide et leur accompagnement pour la préparation des cours et des divers TP et TD : Emmanuelle Revardelle, Sophie Javerzat, Patrick Auguste, Sophie North et Annabelle Redonnet. Merci au personnel du Département Universitaire des Sciences d'Agen pour leur accueil et leur soutien : Carine Ferrand, Viviane Domec, Yves Gillis, Daniel Gastou et Françoise Marc.

Mes remerciements les plus profonds vont à M. Masson et M. Laporte-Cru pour m'avoir transmis leur passion de la biologie et de l'écologie.

Merci à toutes aux personnes avec qui j'ai partagé de bons moments aussi bien au labo qu'à l'extérieur : Manu, Laurent, Antonio, Igor, Pablo, Jean François, Iria, Luis, Marcello, Tristan, Delphine, Le Jeune, Thibault, MH, Fabiano, Ana Luisa, Juan Manuel et Naoki.

Merci aux colloc' Muriel et Hakim pour leur présence agréable bien que relative.

Merci Marie & Jeff, Brice & Séverine, Jérôme & Céline pour m'avoir permis de revenir un peu sur terre de temps en temps.

Merci à toi ma Cécile.

Enfin, merci à mes parents pour leur ouverture d'esprit et leur soutien de toujours.

# Prologue

Ce manuscrit présente les résultats de ma thèse sur l'hybridation au sein du complexe des chênes blancs européens. Certaines parties ont fait l'objet de publications scientifiques qui sont présentées en annexe (Annexes A, B et C), d'autres articles sont également prévus. Par ailleurs, le travail de recherche conduit souvent à des collaborations dont certaines sont valorisées par une publication. Ma contribution à l'analyse de données génétiques chez des variétés d'olivier tunisiennes ainsi que chez le cèdre atlantique a produit deux articles dont les résumés sont joints en annexe (Annexes D et E).

La thèse est organisée en trois grandes parties composées chacune de deux chapitres. Chaque chapitre est structuré comme un article autonome et peut donc être lu indépendamment. La première partie (chapitre 1) introduit la notion d'espèce, fait une synthèse sur les processus de spéciation et d'hybridation avant d'aborder la place des chênes dans ces problématiques et de présenter la démarche mise en œuvre pour étudier l'hybridation contemporaine entre espèces de chênes. La méthodologie employée est expliquée et testée (partie A) avant d'être appliquée sur des croisements contrôlés et en populations naturelles pour quantifier l'hybridation et ses conséquences (partie B) et étudier le système de reproduction des espèces de chênes (partie C). Le manuscrit se termine par une conclusion générale (chapitre 8) qui synthétise les résultats obtenus et propose quelques perspectives de recherches.



# Chapitre 1- Introduction





Page précédente : Le vieux chêne de Baggesen au Danemark (A. Schovelin, XIX<sup>ième</sup> siècle).

Les populations naturelles évoluent sous l'effet des processus de mutation, de sélection, de dérive génétique et de migration. La combinaison de ces forces évolutives peut conduire des populations à évoluer différemment les unes des autres, ce qui peut aboutir à la spéciation lorsque la divergence est importante. Cependant ce processus est graduel et la reproduction entre des espèces qui ne sont pas totalement isolées reproductivement conduit parfois au phénomène d'hybridation. Ces flux de gènes entre espèces peuvent avoir des conséquences évolutives diverses et font l'objet de débats scientifiques récurrents, dont les travaux autour des chênes blancs européens illustrent la vivacité de ces débats. Cette introduction dresse un état des lieux de la question de la définition de l'espèce et du processus de spéciation, préalable indispensable pour aborder le phénomène d'hybridation. Le cas des chênes blancs européens est replacé dans ce contexte puis la démarche générale choisie dans ce travail pour étudier de la dynamique d'hybridation chez ces espèces est présentée.

## 1.1. Synthèse bibliographique

### 1.1.1. Espèce et spéciation

#### 1.1.1.1. La notion d'espèce

Historiquement, une espèce est définie comme un groupe d'individus qui se différencient d'autres groupes par des différences morphologiques héréditaires. Ce critère a permis d'identifier et de classer les espèces décrites selon la nomenclature Linéenne (genre – espèce) qui est toujours aujourd'hui le système de référence pour la classification taxonomique. Cependant, confronté à la diversité et à la complexité des systèmes biologiques, la reconnaissance des espèces sur la base de différences morphologiques est parfois difficile (Mayr, 1992). Les variations morphologiques entre individus, liées aux variations génétiques, à l'âge, à l'environnement ou au sexe, peuvent créer davantage de différences entre individus d'une même population qu'entre individus d'espèces différentes. A l'inverse, les différences morphologiques peuvent être tellement faibles que des espèces ne sont pas reconnues. Les espèces cachées (*cryptic species*) ne sont pas distinguables morphologiquement alors qu'elles coexistent sans se reproduire.

La prise en compte de l'espèce comme unité reproductrice a permis le développement du concept biologique de l'espèce (BSC : *Biological Species Concept* ; Dobzhansky, 1937 ; Mayr, 1942). « Une espèce est une communauté de populations d'individus reproducteurs qui sont isolées reproductivement des autres communautés similaires. » (Mayr, 1992). Cette définition regroupe deux notions. Elle introduit la notion de continuité de reproduction au sein des populations constituant l'espèce (définition multidimensionnelle ; Beurton, 2002). Parmi les individus appartenant à une même espèce, seuls ceux d'une même population se reproduisent et donc concrétisent le brassage génétique assurant les processus d'anagenèse (évolution graduelle et lente d'une lignée). Une seconde notion importante est l'isolement reproducteur entre espèces (définition non-dimensionnelle permettant de délimiter les espèces à un moment donné en un lieu donné ; Beurton, 2002) qui empêche

le brassage de gènes entre espèces déclenchant un processus de cladogenèse (spéciation par divergence). Cette définition n'est applicable que chez les organismes à reproduction sexuée car en cas de reproduction asexuée stricte chaque individu (ensemble de clones) constituerait une espèce à lui seul. De plus il s'appuie sur le critère d'isolement reproducteur sans préciser les mécanismes qui y conduisent. Ainsi ce concept se focalise surtout sur le critère permettant d'identifier une espèce (définition opérationnelle). De tels concepts se sont multipliés dans les années 60 et 70, on en recense maintenant plus d'une vingtaine (phénétique, reproductive, phylogénétique, cluster génotypique...). La pluralité des champs d'investigation scientifiques ainsi que la diversité des systèmes biologiques nécessitent de multiples définitions opérationnelles pour identifier les espèces.

D'autres concepts se focalisent sur ce que représente une espèce et pourquoi elle existe et se maintient. Le concept évolutif de l'espèce définit l'espèce comme une lignée évoluant séparément des autres avec ses tendances et ses rôles évolutifs propres (Simpson, 1951). Ce concept se veut plus général mais il reste cependant peu opérationnel. Ainsi, bien que ce concept théorique soit largement reconnu parmi la communauté scientifique, le débat sur les critères permettant de délimiter les espèces reste intense.

### 1.1.1.2. Spéciation

Si le débat sur la définition de l'espèce est vif, divers mécanismes responsables de la formation de nouvelles espèces ont été proposés et étudiés.

- Modalités biogéographiques

On distingue plusieurs types de spéciation en fonction de la localisation géographique des populations qui divergent lors du processus de spéciation (Figure 1).

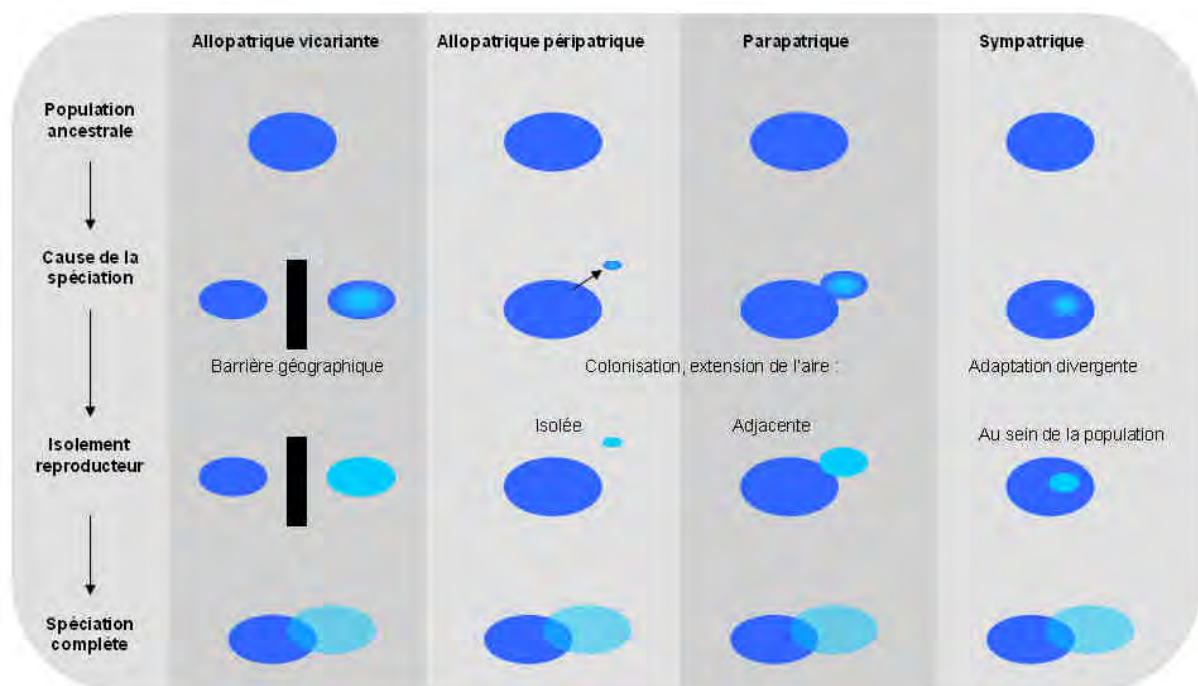


Figure 1 : Différents modes de spéciation classés selon le critère géographique.

#### - Spéciation allopatrique

La spéciation allopatrique a lieu lorsqu'une population est séparée en deux par une barrière géographique, rompant les flux de gènes. C'est historiquement le premier modèle proposé pour expliquer la spéciation qui est liée à la mise en place de barrières reproductives. Les populations qui ne sont plus en mesure de se reproduire lorsqu'elles sont remises en contact sont alors reconnues comme des espèces (Figure 1). Lors de la spéciation vicariante, une barrière majeure sépare l'aire en deux. Les populations divergent alors indépendamment soit par dérive ou par adaptation à des habitats différents de part et d'autre de la barrière. Lors de la spéciation péripatrique une petite population est isolée en dehors de l'aire de répartition. La spéciation s'amorce alors soit par dérive (facilitée par la petite taille de la population) ou par adaptation au nouveau milieu. Cependant les effets de la dérive face à la sélection dans la divergence et la spéciation font débat (divergence liée à la sélection ou à la dérive).

#### - Spéciation sympatrique

La possibilité de spéciation sympatrique n'a été découverte que plus tard. La plus facile à mettre en évidence est le cas de spéciations soudaines. Une spéciation sympatrique peut aussi se développer comme résultante de la sélection divergente entraînant une spécialisation de niche (au sens écologique ou plus généralement au sens adaptatif) qui soit directement ou par entraînement peut permettre le développement de barrières reproductives (pré- et postzygotiques). Dans ce modèle, l'isolement reproducteur n'est qu'un sous-produit de la sélection divergente sur des traits fonctionnels (Rieseberg & Burke, 2001 ; Wu, 2001). Lorsque la sélection est le moteur de la divergence, l'effet homogénéisant des flux de gènes en sympatrie ( $M$  : proportion d'échange génétique) peut être contrebalancé par la sélection ( $S$ ) si  $S > M$  (Figure 2). En condition allopatrique ( $M=0$ ), un niveau très faible de sélection peut conduire à la spéciation. En condition de sympatrie, une divergence peut se développer même en présence de flux de gènes si la sélection est forte, et les cas de spéciation malgré la présence de flux de gènes ont été identifiés (Niemiller *et al.*, 2008 ; Nosil, 2008).

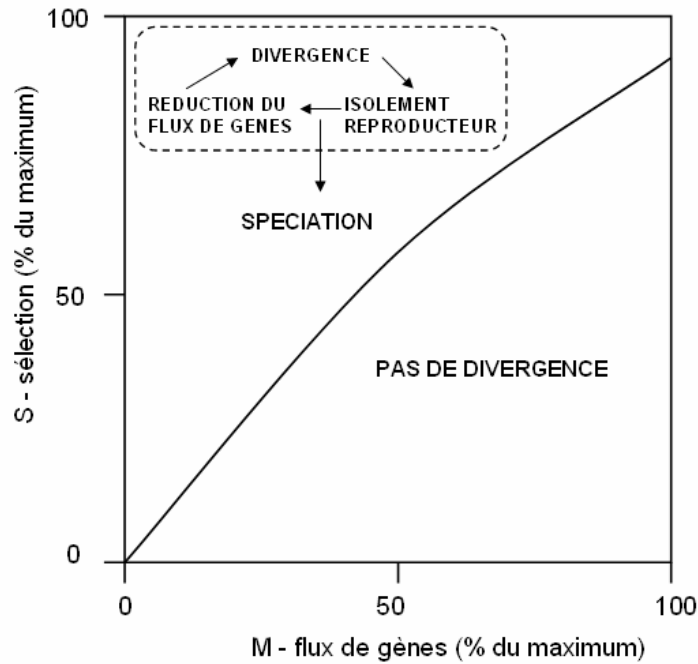


Figure 2 : Spéciation en présence de flux de gènes (adapté de Rice & Hostert, 1993)

#### - Spéciation parapatrique

Deux populations sont isolées géographiquement pour la majeure partie de leur aire de répartition mais elles présentent une sympatrie limitée à quelques aires restreintes. Il existe de nombreuses configurations de parapatrie car tout dépend de la structure des aires de répartition et du degré de sympatrie effectives au niveau local. Aussi, la configuration géographique des aires peut varier dans le temps, des populations peuvent donc expérimenter différentes phases de sympatrie et d'allopatrie. La spéciation parapatrique est ainsi le modèle le plus général de spéciation. Le degré de flux de gènes entre populations est intermédiaire entre la sympatrie ( $M=100\%$ ) et l'allopatrie ( $m=0$ ).

- Isolement reproducteur

L'isolement reproducteur est l'élément qui maintient l'intégrité des espèces. Les barrières reproductives qui se mettent en place lors de la spéciation sont plus ou moins fortes et conduisent à l'isolement reproducteur total entre espèce. Les barrières sont classées selon le moment où elles interviennent. L'isolement pré-reproducteur se réfère à l'absence de reproduction interspécifique lié à des facteurs d'ordre écologique ou comportemental. L'isolement post-reproducteur inclut les mécanismes qui, en cas de reproduction interspécifique, vont empêcher la fécondation ou rendre les hybrides non viables, fertiles ou non adaptés. L'isolement post-reproducteur est constitué de mécanismes prézygotiques (avant la fécondation) et postzygotique (après la fécondation).

#### - Isolement préreproducteur

Les causes d'isolement préreproducteur peuvent être variées. Lors de remises en contact de deux espèces ayant divergé, elles peuvent être interfertiles mais leur descendance sera moins viable. Dans ce cas, les mécanismes évitant les croisements interspécifiques (préreproducteurs et prézygotiques) seront sélectionnés afin d'éviter la production de descendants non viables, un phénomène nommé renforcement (Dobzhansky, 1937).

Chez les animaux, l'isolement comportemental (éthologique) semble particulièrement important dans l'isolement entre espèces. La reconnaissance entre mâle et femelle est essentielle à la reproduction. Ainsi des chants différents isolent deux espèces d'oiseaux par ailleurs totalement identiques morphologiquement (Grant & Grant, 1997). Il en va de même pour l'ensemble des signaux sensoriels qui interviennent dans la communication lors des parades. Dans ces cas, la sélection pour la reproduction surpasse la sélection de survie, la sélection sexuelle peut être une cause importante d'isolement reproducteur.

Chez les plantes à fleur dont la pollinisation est assurée par des animaux (insectes, oiseaux, chauve-souris...) la spécialisation pour une espèce de pollinisateur particulier (*pollinator syndrome*) isole reproductivement les espèces (exemples chez les iris : Martin *et al.*, 2008 ; chez des arbres tropicaux : Kay, 2006). La spécialisation de pollinisateur est la cause fréquente de radiations adaptatives chez plusieurs genres de plantes à fleurs menant à la grande diversité des types floraux. Plus généralement, des modifications d'interactions biotiques étroites (par exemple un changement d'hôte chez des champignons pathogènes de plantes ; Zaffarano *et al.*, 2008) peuvent conduire à une spéciation rapide. L'isolement temporel est lié au décalage des événements de reproduction dans le temps. Ce type d'isolement peut être la conséquence d'un changement de comportement chez les animaux mais peut concerner tous les organismes qui se reproduisent par voie sexuée. Il peut concerner les saisons de reproduction aussi bien que la synchronisation de la libération des gamètes à l'échelle plus fine dans le temps. Chez certaines espèces de plantes, le décalage de la phénologie est une composante importante de l'isolement reproducteur (iris : Martin *et al.*, 2007 ; frênes : Gerard *et al.*, 2006). Les espèces ayant des périodes de floraison identique ou chevauchantes auront davantage l'opportunité d'interagir que des espèces aux phénologies distinctes.

Les espèces ayant des aires de répartition sympatriques peuvent occuper localement des niches écologiques différentes, l'isolement écologique peut alors conduire à l'isolement reproducteur. Ce type d'isolement peut être fort si la spécialisation des espèces à leur habitat est forte et si la sélection contre les migrants l'est aussi. L'isolement écologique est un des moteurs principaux de la spéciation et comporte aussi une composante postzygotique comme nous le verrons plus loin.

- Isolement post-reproducteur

- . Isolement prézygotique

L'isolement génétique est la cause interne (intrinsèque) à l'absence de formation de gamètes. Des recombinaisons chromosomiques, un nombre variable de chromosomes entre espèces sont des freins majeurs qui vont empêcher la méiose. La reconnaissance entre gamètes est nécessaire à la réussite de la fécondation autant chez les animaux que chez les plantes (Howard, 1999 ; Swanson & Vacquier, 2002). Cette reconnaissance a parfois un déterminisme simple, impliquant un nombre réduit de gènes décrits en détail chez certaines espèces (Swanson *et al.*, 2004). L'isolement génétique peut conduire à une divergence de traits liés à l'isolement préreproducteur, directement par sélection sexuelle ou indirectement par liaison génétique. Ceci ayant pour effet d'augmenter encore l'incompatibilité génétique en retour.

- . Isolement postzygotique

Les mécanismes postzygotiques concernent la viabilité et la fertilité des hybrides formés. L'isolement postzygotique extrinsèque est lié au fait que les hybrides ont des phénotypes intermédiaires non adaptés à l'environnement des espèces parentales alors que l'isolement postzygotique intrinsèque est la conséquence d'épistasies négatives entre allèles provenant d'espèces différentes. L'isolement postzygotique intrinsèque peut se manifester à trois niveaux : la non viabilité des hybrides de première génération (F1), la stérilité des hybrides F1 et la dépression hybride qui peut produire une baisse de viabilité et fertilité au-delà de la première génération hybride.

Les hybrides peuvent se révéler non viables à toutes les étapes précédant la reproduction : du développement embryonnaire au stade jeune. Les zygotes F1 produits peuvent avorter à des stades précoces du développement embryonnaire ou plus tard. Les hybrides qui parviennent à l'âge adulte présentent chez un certain nombre d'espèces une faible fertilité, une stérilité partielle (par exemple, perte de fertilité mâles chez certaines plantes) ou totale. Les cas de stérilité totale représentent une barrière infranchissable pour les flux de gènes. Les hybrides ne pouvant se reproduire, le flux de gènes entre espèces est coupé et l'hybridation aura pour conséquence d'occuper le milieu par des individus stériles, mais parfois vigoureux. Chez les organismes qui peuvent se reproduire par voie asexuée (plantes, champignons, bactéries...), des hybrides vigoureux mais stériles peuvent vite devenir envahissants. Dans la plupart des organismes monoïques à déterminisme sexuel génétique (chromosomique), c'est le sexe hétérogamétique qui est affecté par la stérilité (*Haldane rule*).

Le modèle d'incompatibilité Dobzhansky-Muller (Dobzhansky, 1937) propose qu'une épistasie négative entre un gène de l'espèce 1 et un gène de l'espèce 2 va créer un manque de viabilité ou de fertilité. L'interaction négative serait assimilée à une mutation délétère dont l'effet est non exprimé (ou supprimé) dans le fond génétique de son espèce d'origine mais exprimé une fois au contact d'une partie du génome de l'autre espèce (Burke & Arnold, 2001). Ce modèle se confirme chez certaines espèces où les gènes responsables de la létalité ou de la stérilité des hybrides sont identifiables, par

exemple chez la drosophile (Presgraves *et al.*, 2003; Orr *et al.*, 2004, Orr, 2005; Masly *et al.*, 2006 ; Presgraves, 2007). Une mauvaise régulation de l'expression de gènes impliquant des ARNi (gene silencing) dans le génome hybride peut être la cause de leur létalité (Bomblies, 2006 ; Bomblies & Weigel, 2007). Ces résultats récents indiquent que les gènes responsables sont des gènes ordinaires de fonctions variées mais les facteurs de transcription semblent particulièrement nombreux (Orr, 2005). De plus, ces gènes semblent évoluer rapidement en réponse à la sélection. L'isolement postzygotique serait donc un sous-produit de l'évolution divergente des espèces.

- Evolution de l'isolement au cours de la divergence

- Isolement reproducteur soudain

L'isolement reproducteur peut être dans certains cas acquis très rapidement et mener à une spéciation sans grand changements dans la composition génétique au niveau allélique (Avisé, 2004). L'allopolypléidisation générée par hybridation entre espèces qui diffèrent par leur composition chromosomique conduit à un isolement reproducteur immédiat, la reproduction entre, par exemple, un individu tétraploïde et un autre diploïde produisant des descendants triploïdes stériles. Cependant, l'allopolypléidisation n'entraîne pas toujours un isolement reproducteur complet (Slotte *et al.*, 2008).

Les réarrangements chromosomiques tels que des inversions ou des translocations peuvent contribuer à un isolement reproducteur rapide. Les différences de structure chromosomiques entre espèces peuvent poser des problèmes d'appariement chromosomique ou de disjonctions à la méiose chez les hybrides, entraînant leur stérilité. La viabilité des hybrides peut être affectée par une modification de l'expression de gène par effet de position ou réduire la recombinaison dans les régions génomiques réarrangées causant une réduction des flux de gènes (Rieseberg, 2001).

Une dernière cause d'isolement reproducteur rapide peut résulter d'un changement brutal de système de reproduction. Chez les plantes, une mutation du système d'auto-incompatibilité ou une évolution de l'hétérostylie vers l'homostylie peut apparaître rapidement et favoriser les réarrangements chromosomiques chez les individus autoféconds (Avisé, 2004). La période de reproduction ou l'interaction avec des pollinisateurs spécialisés peut aussi conduire à un isolement reproducteur immédiat, par exemple chez une espèce du genre *Mimulus*, la mutation d'un seul gène affectant l'aspect de la fleur a produit un changement de pollinisateur (Bradshaw & Schemske, 2003).

- Apparition graduelle de l'isolement

Outre les cas particuliers présentés précédemment, on considère généralement que les barrières reproductives apparaissent de façon graduelle au fur et à mesure de la divergence quel que soit le mécanisme de spéciation (allo- para- ou sympatrique). Ainsi lorsque deux populations en voie de divergence sont remises en contact, l'isolement reproducteur peut ne pas être complètement achevé : les flux de gènes peuvent être rétablis et conduire à la convergence génétique des deux populations ou au contraire à la spéciation (par renforcement), suivant le degré de divergence accumulé. Il semble en fait que la spéciation en présence de flux de gènes soit un phénomène assez courant (Rice & Hostert,



1993 ; Nosil, 2008 ; Strasburg & Rieseberg, 2008), soit en sympatrie ou bien lors d'alternances de phases allopatriques et parapatriques par exemple.

Les mécanismes menant à la formation de nouvelles espèces sont variés et parfois difficiles à mettre en évidence. Toutes les définitions d'espèce et les méthodes permettant de les délimiter sont confrontées à la difficulté de décrire et classer le résultat d'évolutions graduelles. Des étapes de spéciation incomplètes sont fréquemment observées à différents moments et ainsi toutes les populations ne peuvent être classées dans des espèces biologiques discrètes (Mayr, 1992 ; Rieseberg *et al.*, 2006).

### 1.1.2. Hybridation

L'hybridation a participé pour beaucoup au débat sur la définition des espèces. Il est vite apparu que chez certaines espèces, les flux de gènes existaient entre ce qui apparaissait être de "bonnes" espèces. Malgré ces flux de gènes, les formes parentales semblaient maintenues.

#### 1.1.2.1. Historique de la notion d'hybridation

Bien que non intentionnelle au départ, l'hybridation a accompagné l'histoire de la domestication des plantes et des animaux par l'homme depuis le Néolithique (Rieseberg & Carney, 1998). L'hybridation permet d'une part de créer de nouvelles variétés comestibles et est d'autre part utilisée pour obtenir un rendement élevé grâce à l'effet d'hétérosis des hybrides de première génération entre lignées pures (homozygotes).

Trois théories sur l'importance évolutive de l'hybridation existent. La première, la plus répandue parmi les zoologistes, considère que l'hybridation n'a aucune conséquence évolutive sur le long terme (Wagner, 1970). Une deuxième hypothèse est que l'hybridation est un phénomène important dans le processus de spéciation, conduisant soit au renforcement de l'isolement reproducteur, soit à une convergence des espèces par disparition des barrières reproductives (Harrison, 1993). La troisième hypothèse considère l'hybridation comme source de nouveauté évolutive (Anderson & Stebbins, 1954, Rattenbury, 1962).

Dans une synthèse sur l'hybridation chez les plantes, Rieseberg et Carney (Rieseberg & Carney, 1998) font un bref historique de la notion d'hybridation et soulignent trois découvertes importantes qui ont fondé l'étude moderne de l'hybridation. L'hybridation a d'abord été vue comme un phénomène permettant la création de nouvelles espèces (spéciation hybride). Wingé (Winge, 1917) découvre qu'il est possible de créer une nouvelle espèce hybride stable instantanément par doublement de ses chromosomes (allopolyploïdie). La population hybride se retrouve isolée reproductivement des espèces parentales et peut constituer une nouvelle espèce. La spéciation hybride semble être particulièrement répandue chez les plantes. La spéciation par allopolyploïdisation est l'un des phénomènes qui aurait pu intervenir dans la diversification de la vie sur terre. Une seconde découverte importante concerne la spéciation hybride sans changement du nombre de génome (spéciation hybride

homoploïde). Müntzing (Müntzing, 1930) émet l'hypothèse que le tri chromosomique au cours des générations hybrides peut conduire à une nouvelle espèce stable, fertile et partiellement isolée reproductivement des espèces parentales du fait du réarrangement chromosomique. Des études plus récentes indiquent que ce type de spéciation hybride peut être dû à des facteurs de stérilité géniques et non seulement chromosomiques.

Enfin, les travaux expérimentaux d'hybridation en population naturelle d'Anderson et Stebbins (Anderson & Hubricht, 1938 ; Anderson & Stebbins, 1954) montrent que l'hybridation et l'introgession pourraient être favorisées par la sélection et contribuer à l'évolution adaptative des populations.

Comme nous l'avons vu précédemment, l'isolement reproducteur est une notion clé de la définition biologique de l'espèce. Néanmoins, les cas d'hybridation décrits dans la littérature sont nombreux. Les questions liées au processus de spéciation et au maintien d'espèces différentes malgré l'hybridation restent encore d'actualité, cependant plusieurs modèles ont été proposés pour rendre compte des processus intervenant dans les zones hybrides.

#### 1.1.2.2. Modèles de zones hybrides

Trois types de modèles de zone hybride existent pour expliquer la dynamique d'hybridation et en particulier le maintien des espèces parentales malgré des flux de gènes entre espèces. Le débat concerne en premier lieu la fitness (viabilité et fertilité) des hybrides. Certains modèles supposent une faible fitness des hybrides et d'autres considèrent que la fitness des hybrides est au moins égale à celle des espèces parentales. Selon les modèles la fitness des hybrides peut être affectée par des facteurs intrinsèques, d'autres modèles donnent plus d'importance aux facteurs extrinsèques dépendant de l'environnement. Ainsi, la viabilité et la fertilité des hybrides peuvent varier et l'hybridation aura des conséquences évolutives différentes en fonction des modèles.

- Modèles à hybrides ayant une faible fitness

- *Dynamic equilibrium model* (Bigelow, 1965) ; *Tension zone model* (Barton & Hewitt, 1985) :

La fitness des hybrides est indépendante de l'environnement et la zone hybride est maintenue par un équilibre entre dispersion des espèces parentales et sélection contre les hybrides. Les barrières postzygotiques intrinsèques sont responsables de la faible fitness générale des hybrides. Dans une méta-analyse recensant les résultats sur la fitness des hybrides, Arnold et Hogdes (Arnold & Hodges, 1995) enregistrent 13 cas d'hybrides montrant une fitness réduite sur 44 étudiés, ainsi que 24 cas pour lesquels les hybrides ont une fitness plus élevée que l'un des parents ou les deux. Dans 7 cas seulement la fitness des hybrides est équivalente à celle des parents. Ces résultats, bien que discutés (Day & Schluter, 1995), démontrent que les modèles classiques qui supposent une faible fitness des hybrides ne suffisent pas à expliquer tous les cas d'hybridations.

- Modèles où la fitness des hybrides dépend en partie de l'environnement

- *Bounded hybrid superiority model* (Moore, 1977) : Dans ce modèle, les hybrides, formés lors du contact secondaire entre deux taxons, peuvent avoir une fitness supérieure aux parents dans leur environnement. Si tel est le cas, les reproductions hétérosécifiques seront favorisées car les individus y participant auront également une fitness plus élevée. Depuis 25 ans, Good et collaborateurs (Good *et al.*, 2000) comparent le succès reproducteur des hybrides et des espèces parentales des oiseaux du genre *Larus* (*Larus occidentalis* et *L. glaucescens*). Ils montrent que les reproductions entre femelle d'espèce pure et mâle hybride donnent plus de descendants. Les mâles hybrides choisissent un territoire moins exposé à la prédation ce qui explique une partie des résultats. De plus, ils offrent un régime alimentaire qui donne un meilleur taux de survie aux jeunes. Les auteurs concluent que les modèles de zones hybrides classiques (impliquant la sélection contre les hybrides, et le renforcement des barrières prézygotiques par l'isolement postzygotique) ne correspondent pas aux données observées. Chez les oiseaux, les barrières comportementales sont les premières à évoluer et les barrières postzygotiques n'évoluent que plus tard, ce qui augmente les opportunités d'hybridation. L'intégrité des espèces est maintenue en dehors de la zone hybride car l'avantage hybride y est restreint géographiquement et la dynamique d'hybridation peut varier au cours du temps.

- Modèles à hybrides pouvant avoir une meilleur fitness

- *Evolutionary novelty model* : Reprenant les idées d'Anderson et de Stebbins (Anderson & Stebbins, 1954), Arnold fait remarquer que les hybrides ne représentent pas une seule unité mais au contraire une multitude de génotypes très variés (Arnold, 1997). Certains génotypes hybrides peuvent avoir une fitness inférieure aux génotypes parentaux, mais certaines combinaisons d'allèles peuvent produire des individus aussi viables et fertiles, voire plus, que les espèces parentales. Ce modèle considère que l'hybridation peut mener à la formation de nouveaux génotypes qui exploiteraient d'autres ressources ou s'adapteraient à un nouvel habitat. Ainsi selon ce modèle, l'hybridation est souvent associée à des milieux perturbés (Anderson, 1948). Enfin, l'introgession, qui est le transfert de gènes entre espèces suite à plusieurs générations d'hybridation, permettrait aux espèces d'abriter plus de variation et de s'adapter plus vite à des changements environnementaux. L'hybridation pourrait accélérer l'évolution en générant de la variation adaptative, conduire à la formation de nouveaux complexes de gènes, voire de nouvelles espèces. A l'instar de la mutation, de la recombinaison, de la dérive et de la migration, l'hybridation et l'introgession sont des processus qui génèrent de la nouvelle diversité.

Le rôle évolutif de l'hybridation varie en fonction des systèmes biologiques considérés. Les modèles descriptifs et prédictifs sont donc nombreux. Si l'importance évolutive de l'hybridation reste un sujet très discuté, ses conséquences sont observées chez de nombreuses espèces.

### 1.1.2.3. Conséquences de l'hybridation

- Spéciation par hybridation
- Spéciation inverse

La levée soudaine de certaines barrières prézygotiques peut conduire à une forte augmentation du taux d'hybridation pouvant aller jusqu'à la fusion de plusieurs espèces en une seule (phénomène appelé spéciation inversée). Des études sur le long terme portant sur des espèces d'épinoches benthique et limnique du Lake Enos ont montré que le nombre d'hybride entre ces deux espèces augmente depuis une vingtaine d'année. De 1 % détecté en 1984 et 1992, un taux d'hybride de 17% est trouvé en 1999 (Kraak *et al.*, 2001). Des études morphologiques et génétiques basées sur un échantillonnage datant de 1977 à 2002 montrent deux groupes morphologiques et génétiques jusqu'en 1994, correspondant aux deux espèces mais à partir de 1997, les individus correspondent à un seul groupe morphologique très variable et à un seul groupe génétique (Taylor *et al.*, 2006). La concordance entre données morphologiques et génétiques tend à prouver que les deux espèces ont fusionnées en une seule par rupture de la barrière prézygotique. Ce phénomène brusque est probablement lié à l'introduction d'une espèce d'écrevisse exotique (Seehausen, 2006). Celles-ci ont augmenté la turbidité de l'eau et entraîné la disparition de la flore aquatique. Ces modifications de l'environnement ont perturbé la reconnaissance conspécifique et augmenté le nombre de croisements hétérosécifiques. La disparition de la flore aquatique a réduit les lieux de ponte et fait disparaître la différence d'habitat pour la ponte entre les deux espèces. Ce phénomène se retrouve également pour les espèces de cichlidés du lac Victoria en Afrique de l'est. La turbidité du lac liée à l'anthropisation et l'introduction de la perche du Nil a abouti à une érosion de la diversité spécifique. La sélection sexuelle, diminuée par le manque de visibilité, a conduit à un phénomène d'homogénéisation des espèces par hybridation (Seehausen *et al.*, 1997). La reconnaissance sexuelle chez les poissons repose souvent sur la couleur ce qui implique une dépendance du milieu, en particulier à la clarté de l'eau (Endler & Houde, 1995). Une levée de cet isolement prézygotique comportemental peut conduire au mélange généralisé des espèces.

#### - Spéciation hybride

Outre la spéciation allopolyploïde présentée précédemment, l'hybridation peut créer de nouvelles espèces sans changement de ploïdie (spéciation hybride homoploïde). Les espèces isolées par des barrières de stérilité chromosomique ou génique peuvent donner naissance, par recombinaison de ces facteurs de stérilité, à une nouvelle espèce isolée reproductivement des deux espèces parentales (Avisé, 2004). Ce modèle de spéciation par recombinaison nécessite une forte sélection en faveur des hybrides les plus viables, une évolution chromosomique rapide et la disponibilité d'un habitat adapté pour l'établissement de la nouvelle espèce (Rieseberg & Carney, 1998 ; pour un exemple chez le tournesol, voir Rieseberg *et al.*, 1990 ; Rieseberg, 1991).

#### • Introgression

L'introgression est le transfert de gènes d'une espèce à l'autre par hybridation (Anderson & Hubricht, 1938). L'introgression de caractères agronomiques d'intérêts d'une espèce sauvage à une espèce cultivée est utilisée en agronomie pour l'amélioration variétale. La facilité avec laquelle un gène va pouvoir passer d'une espèce à l'autre dépend de l'architecture génétique des barrières reproductives

(Rieseberg & Carney, 1998). Lorsque de nombreux gènes sont impliqués dans la létalité des hybrides, la majeure partie du génome sera résistante à l'introgression à cause du déséquilibre de liaison entre le gène d'intérêt et ces gènes. Si la recombinaison est faible le déséquilibre de liaison sera d'autant plus fort et l'introgression difficile. De plus, si les gènes impliqués dans la létalité des hybrides interagissent ou sont coadaptés, l'introgression sera d'autant plus difficile. Si les barrières reproductives sont sous contrôle génétique simple (quelques gènes), une grande partie du génome devrait être perméable à l'introgression et seuls les gènes localisés à proximité des facteurs de la létalité des hybrides seront difficile à introgresser. La différenciation génétique entre certaines espèces proches apparaît très hétérogène le long du génome. Certaines régions génomiques, probablement liées à la divergence (adaptative ou reproductive) entre espèces, montrent un fort niveau de différenciation alors que d'autres régions montrent un très faible niveau de différenciation et représentant probablement des zones génomique perméables aux flux de gènes interspécifiques. Cette théorie renvoie directement à des résultats obtenus chez les plantes (Rieseberg & Burke, 2001) ainsi que des études récentes chez la drosophile (Wu, 2001).

Les résultats de rétrocroisements (un hybride de première génération croisé avec une des espèces parentales) montrent souvent des distorsions de ségrégation dans la descendance, c'est-à-dire une ségrégation anormale par rapport aux proportions mendéliennes normales. Ces anomalies sont liées au fait que le génome d'un des deux parent est éliminé plus rapidement que ce qui est normalement attendu dans les descendances successives de rétrocroisements. Ainsi, quelques rétrocroisements récurrents suffisent à restaurer une descendance, qui bien qu'issue d'un hybride, ressemble phénotypiquement à l'espèce parentale récurrente.

- Asymétrie et déplacement des zones hybrides

La structure des zones hybrides varie en fonction des différences génétiques et écologiques entre les espèces qui sont en contact, de leur taux de migration ainsi que de la viabilité des hybrides (Buggs, 2007). Selon ces caractéristiques, une zone hybride peut être stable ou bien se déplacer. Si les hybrides ont un avantage sélectif dans un environnement intermédiaire par rapport aux espèces parentales (*Bounded Hybrid Superiority Model*), la zone hybride sera stable. Si en revanche les hybrides présentent une meilleure adaptation que les espèces parentales dans leur environnement, la zone hybride peut s'étendre au détriment des espèces parentales. L'expansion peut être unidirectionnelle si l'une des deux espèces parentale est moins adaptée que l'autre, et aboutir au remplacement de l'espèce la moins compétitive par la zone hybride.

Une extension directionnelle de la zone hybride est aussi possible si l'hybridation est asymétrique et que des rétrocroisements successifs ont lieu entre les hybrides et une seule des deux espèces (Rieseberg & Carney, 1998). Un tel phénomène peut être mis en évidence en comparant la structuration géographique de marqueurs génétiques à hérédité uniparentale (cytoplasmiques ou liés à un chromosome sexuel) et biparentale (nucléaire autosomal). Le génome mitochondrial, à hérédité maternelle chez les mammifères et paternelle chez la plupart des angiospermes, ainsi que le génome

chloroplastique, à hérédité maternelle chez les angiospermes, sont des bons outils pour étudier la direction des flux de gènes causée par une incompatibilité sexuelle asymétrique. L'étude de l'association entre génotypes nucléaires et cytoplasmiques (déséquilibre cytonucléaire, Arnold, 1993) permet de préciser la dynamique d'hybridation entre espèces, et notamment de détecter une hybridation directionnelle. Un exemple classique concerne deux espèces de souris, *Mus domesticus* et *Mus musculus*, qui possèdent normalement des allèles spécifiques dans leur génome mitochondrial. En Suède, des souris présentaient une morphologie et un génome nucléaire caractéristique de *Mus musculus* alors que leur génome mitochondrial était typique de l'espèce *Mus domesticus* (Gyllensten & Wilson, 1987). Une petite population de femelles *M. domesticus* aurait migré en Scandinavie à l'occasion de l'arrivée de l'agriculture il y a 4000 ans. Des rétrocroisements répétés avec des souris *Mus musculus* mâles auraient permis l'introduction de l'ADN mitochondrial *domesticus* dans les populations ayant un fond génétique nucléaire majoritairement *musculus*. Ce phénomène de capture cytoplasmique a été montré chez de nombreuses espèces animales et végétales (Awise, 2004). Cependant, de tels résultats peuvent s'expliquer par le partage d'un polymorphisme ancestral et des analyses complémentaires et indépendantes sont nécessaires pour trancher entre les deux hypothèses. Néanmoins dans la plupart des cas, l'introgression est la cause la plus probable de ces observations et démontre que les flux de gènes entre espèces sont courants dans l'histoire évolutive récente.

Chez les plantes, les mouvements de gènes sont liés à deux vecteurs distincts : le pollen et les graines. Lorsque le flux de gènes est lié uniquement au pollen, les marqueurs chloroplastiques (à transmission maternelle) ne s'introgressent pas. Chez des espèces d'iris de Louisiane, dans certaines populations le flux de gènes entre *Iris fulva* et *Iris hexagona* est asymétrique (Arnold *et al.*, 1992), de nombreux hybrides possèdent le génome chloroplastique de *I. hexagona* ce qui montre qu'ils sont issus du transfert de pollen de *I. fulva* vers *I. hexagona*. De tels cas d'introgression de marqueurs nucléaires sans transfert de marqueurs chloroplastiques ne sont pas rares chez les plantes. L'origine de cette asymétrie peut être liée à divers phénomènes comme la préférence des pollinisateurs, l'incompatibilité interspécifique unidirectionnelle ou un système de stérilité cytoplasmique conduisant à la stérilité partielle (mâle ou femelle) chez les hybrides. Quelle qu'en soit la cause, de telles asymétries permettent à certaines espèces de migrer via leur pollen par hybridation et rétrocroisement récurrent (eucalyptus : Potts & Reid, 1988 ; chênes : Petit *et al.*, 2003). Ce modèle sera détaillé plus tard car il constitue le centre de la problématique de cette thèse.

### 1.1.3. Conclusion

Cette revue bibliographique met en évidence la multiplicité des concepts nécessaires à décrire des phénomènes aussi complexes que la spéciation ou l'hybridation. Ainsi l'hybridation peut dans certains cas avoir des conséquences évolutives sur les espèces qui y participent. Il paraît en outre évident vu la pluralité des concepts que l'étude de l'hybridation nécessite une définition précise du sens du mot "espèce". La délimitation des espèces n'est pas une chose facile dans bien des cas et reste pourtant

primordiale pour étudier l'hybridation. Les chênes blancs européens constituent un modèle illustrant différents points de cette problématique comme nous allons le voir maintenant.

## 1.2. Modèle d'étude : le complexe d'espèces des chênes blancs européens

Le statut taxonomique incertain des chênes européens à feuille caduques fait l'objet de débat, ainsi que le degré d'isolement reproducteur des taxa, et la présence d'hybridation dans les peuplements composés de plusieurs espèces sympatriques.

### 1.2.1. Présentation du genre *Quercus*

Le genre *Quercus* appartient à la famille des Fagacées (qui comprend également le hêtre, *Fagus sylvatica*, et le châtaignier, *Castanea sativa*). Ce genre compte à son dernier recensement 531 espèces (Govaerts *et al.*, 2006) réparties dans 5 sections phylogénétiques (Figure 3). La section *Lobatae* comprend les chênes rouges et noirs d'Amérique du Nord et Centrale, la section *Cerris* inclue les espèces de chênes méditerranéennes à feuilles persistantes (chêne vert, *Quercus ilex*, et le chêne liège, *Quercus suber*, entre autre) et la section *Quercus* correspond aux chênes blancs qui font l'objet de la présente étude (Manos *et al.*, 1999 ; Manos, 2001; Manos *et al.*, 1999).

L'hybridation est un phénomène étudié de longue date chez les chênes. Des croisements contrôlés entre espèces de chênes nord-américains ont été pratiqués longtemps à l'Université de l'Utah (Cottam *et al.*, 1982) et illustrent l'interfertilité de nombreuses espèces, parfois même entre espèces de sections différentes. Au départ, des études de populations de chênes nord-américains utilisaient des critères morphologiques pour analyser le degré d'hybridation entre espèces (Muller, 1952). L'hybridation entre 8 paires d'espèces tenant compte de l'environnement édaphique et climatique a ainsi été décrite. Muller montre que le taux d'hybridation dépend des facteurs affectant la survie des hybrides et notamment des caractéristiques écologiques des espèces parentales. Si deux espèces sont adaptées à des sols différents, leurs hybrides survivront sur les sols intermédiaires. Les préférences climatiques des espèces vont jouer un rôle sur leur aire de répartition et donc leur degré de sympatrie, ce qui peut limiter les possibilités d'hybridations. Prenons l'exemple de *Q. grisea*, une espèce adaptée à un sol granitique mais qui peut aussi se contenter de calcaire dolomitique et *Q. Mohriana* une espèce qui se retrouve sur sol calcaire. Lorsque ces espèces se retrouvent en contact alors qu'elles occupent des sols différents, l'hybridation observée est très faible. Par contre, dans d'autres zones de contact caractérisées par un sol calcaire également adapté aux deux espèces, l'hybridation est très fréquente.

Plus récemment, l'utilisation de marqueurs moléculaires a permis de confirmer l'importance de l'hybridation chez les chênes appartenant à plusieurs sections (Nason *et al.*, 1992a ; Craft *et al.*, 2002 ; Ishida *et al.*, 2003 ; Dodd & Afzal-Rafii, 2004 ; Gonzalez-Rodriguez *et al.*, 2004 ; Tovar-Sanchez & Oyama, 2004 ; Kanno *et al.*, 2004 ; Scareli-Santos *et al.*, 2007). Chez les chênes blancs (section *Quercus*), *Q. grisea* et *Q. gambelii* sont deux espèces qui forment des hybrides mais l'introggression est limitée (Howard *et al.*, 1997). Dans une étude comparant le succès reproducteur de ces deux espèces

dans le cadre de croisements contrôlés répétés dans deux environnements, Williams *et al.* proposent que le stress environnemental augmente la probabilité d'hybridation (Williams *et al.*, 2001). En effet, en environnement stressant, le pollen de *Q. gambelii* a tendance à être moins fertile, ce qui induit une plus grande proportion de croisements hétérospécifiques. Les auteurs concluent que l'hybridation peut être accentuée par un stress qui diminue la compétitivité reproductive du pollen d'une espèce subissant un stress environnemental, phénomène qu'ils nomment émascation environnementale.

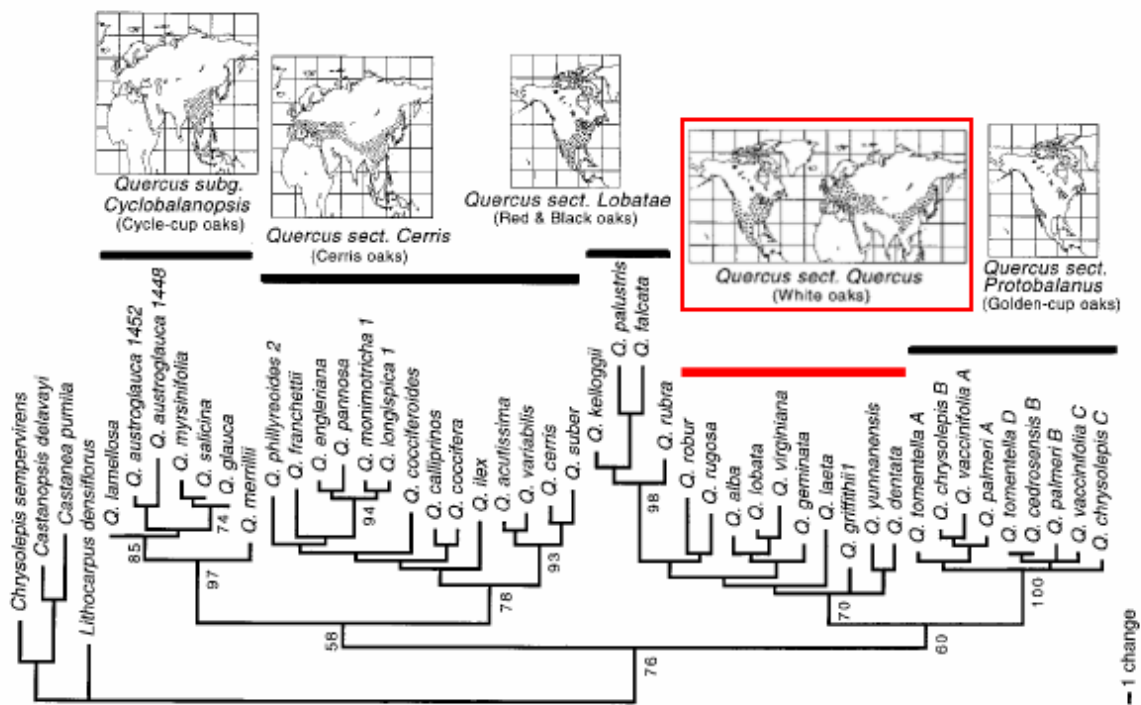


Figure 3 : Arbre phylogénétique du genre *Quercus* obtenu par analyse de parcimonie des séquences ITS (adapté de Manos, 2001)

### 1.2.2. Le complexe d'espèces des chênes blancs

Nous allons nous intéresser aux espèces de chênes blancs européens, arbres dominants de nombreuses forêts tempérées et subméditerranéenne en Europe. Ces espèces ont fait l'objet de nombreuses études de biologie évolutive. Avant de détailler la problématique de cette thèse, les principales caractéristiques des espèces et leur histoire sont présentées ce qui va permettre de préciser le sens de la notion d'espèce chez les chênes avant d'aborder les différents modèles d'hybridation chez ces espèces.

#### 1.2.2.1. Caractéristiques des quatre espèces

- Aire de répartition

Les chênes pédonculé et sessile (*Quercus robur* L. et *Q. petraea* (Matt.) Liebl.) ont une large aire de répartition en Europe qui s'étend du nord de la péninsule ibérique jusqu'à la Russie, l'aire du chêne sessile est limitée au nord et à l'est par rapport au chêne pédonculé à cause de sa sensibilité au froid (Figure 4 A et B). Le chêne pubescent (*Quercus pubescens* Willd.) a une distribution plus méridionale,



largement confiné à la façade méditerranéenne (Figure 4 C). Le chêne tauzin (*Quercus pyrenaica* Willd.) a une aire de distribution restreinte à la façade atlantique du sud-ouest de l'Europe (Figure 4 D).

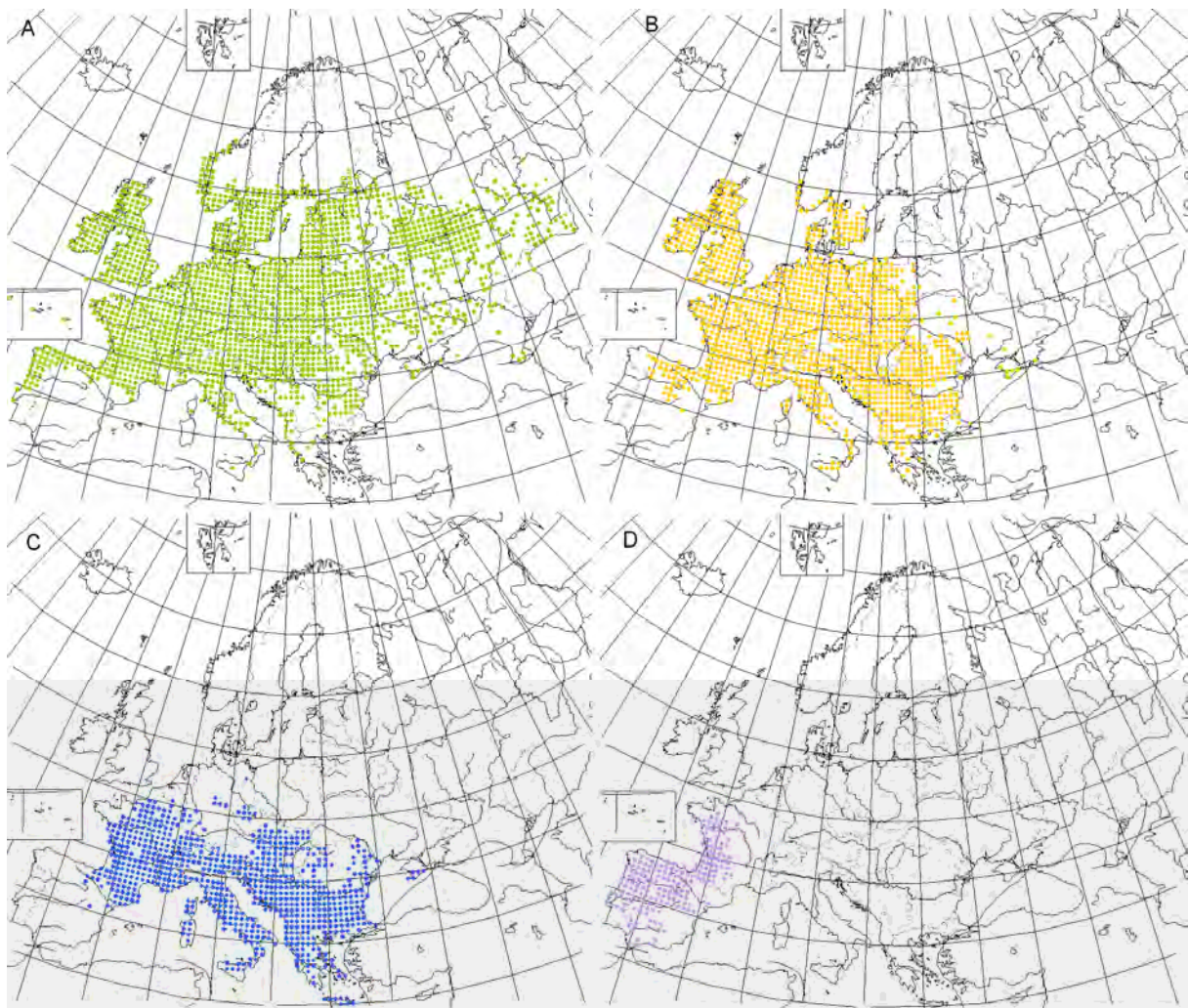


Figure 4 : Aires de répartition des espèces de chênes blancs en Europe (A : *Quercus robur*, chêne pédonculé ; B : *Quercus petraea*, chêne sessile ; C : *Quercus pubescens*, chêne pubescent ; D : *Quercus pyrenaica*, chêne tauzin ; adapté de Atlas Florae Europaeae 1999).

- Morphologie

Plusieurs caractères morphologiques des feuilles et des fruits permettent de reconnaître les espèces en forêt. Elles présentent aussi de légères différences au niveau de l'aspect de l'écorce et de la forme du houppier mais ces caractères ne sont pas détaillés ici. Le **chêne pédonculé** doit son nom à la présence d'un long pédoncule qui attache les akènes (glands) au rameau (Figure 5 A). Ce long pédoncule contraste avec le très court pétiole qui rattache les feuilles au rameau, donnant l'impression que les feuilles sont regroupées en paquets, alors que les glands apparaissent bien visibles. La combinaison de caractères propres à cette espèce (présence d'un long pédoncule et forme allongée des glands) pourrait rendre les glands plus visibles et attractifs pour le geai des chênes, vecteur important pour leur dispersion (Jones, 1959 ; Bossema, 1979 ; den Ouden *et al.*, 2005). Les feuilles présentent aussi

quelques particularités telles que la présence d'oreillettes à la base ou l'absence totale de pilosité. Les trois autres espèces ont un pédoncule très réduit ou absent : leurs glands sont directement fixés en paquets sur le rameau. En outre, leurs feuilles présentent de longs pétioles (Figure 5). Elles se distinguent le plus évidemment par leur degré de pubescence. Le **chêne sessile** ne possède pas de pubescence (ou de façon très réduite au niveau de la nervure centrale), le **chêne pubescent** a des poils sur la face inférieure des feuilles alors que le **chêne tauzin** présente de la pilosité sur toute la feuille. La forme de leurs feuilles est assez différente (Figure 5 B, C et D). Le chêne tauzin présente des sinus profonds pouvant pratiquement atteindre la nervure centrale. Le chêne pubescent a des lobes plus nombreux et irréguliers que le chêne sessile.

Comme le fait remarquer Darwin à Alphonse de Candolle (Darwin, 1863), ces espèces ont une très grande variabilité intraspécifique :

*"I have been fairly astonished at the amount of individual variability in the oaks."*

Cette grande variabilité rend parfois l'identification des espèces difficile dans les peuplements où plusieurs espèces cohabitent. Des analyses morphologiques à l'aide de plusieurs mesures permettent de séparer les espèces globalement (Dupouey & Badaeu, 1993 ; Kremer *et al.*, 2002a) mais il n'y a pas de limite franches et certains individus ont une morphologie intermédiaire.

- **Ecologie**

Les quatre espèces de chênes présentent des caractéristiques écologiques propres (Figure 6) qui vont expliquer leur présence dans les peuplements. Le chêne pédonculé supporte l'ennoyage racinaire temporaire mais craint les fortes sécheresses estivales. On le trouve donc sur des sols toujours frais, il apprécie particulièrement les sols riches et profonds des zones alluviales. Le chêne sessile supporte la sécheresse estivale et les sols relativement pauvres. C'est une espèce franchement forestière qui apparaît plus tardivement que les autres dans la succession forestière. Les chênes pubescents et tauzins supportent bien les sols secs et les stations thermophiles. Ces deux espèces ont en revanche des préférences marquées pour la nature du sol : le chêne tauzin est présent sur sol acide et sableux alors que le chêne pubescent apprécie les sols calcaires.

Cependant, ces préférences ne sont que des tendances et il est fréquent de trouver des forêts dans lesquels cohabitent plusieurs espèces. Par exemple, les chênes pédonculés et sessiles sont souvent associés dans les grandes forêts des plaines alluviales, les chênes pédonculés et tauzins se rencontrent sur les sables du plateau landais et les chênes pédonculés et pubescents dans les bois des coteaux calcaires du sud-ouest de la France. Au sein de chacune de ces zones, les espèces se répartissent suivant les conditions pédo-écologiques locales : elles sont rarement mélangées pied à pied sauf dans les zones de transitions écologiques marquant la limite de répartition plus ou moins nette des espèces.



**Figure 5** : Dessins botaniques de chênes pédonculés (A), sessiles (B), pubescents (C) et taurins (D).  
(Dessins A, B et D : Duhamel du Monceau, 1755 et dessin C : Hempel & Wilhelm, 1889).



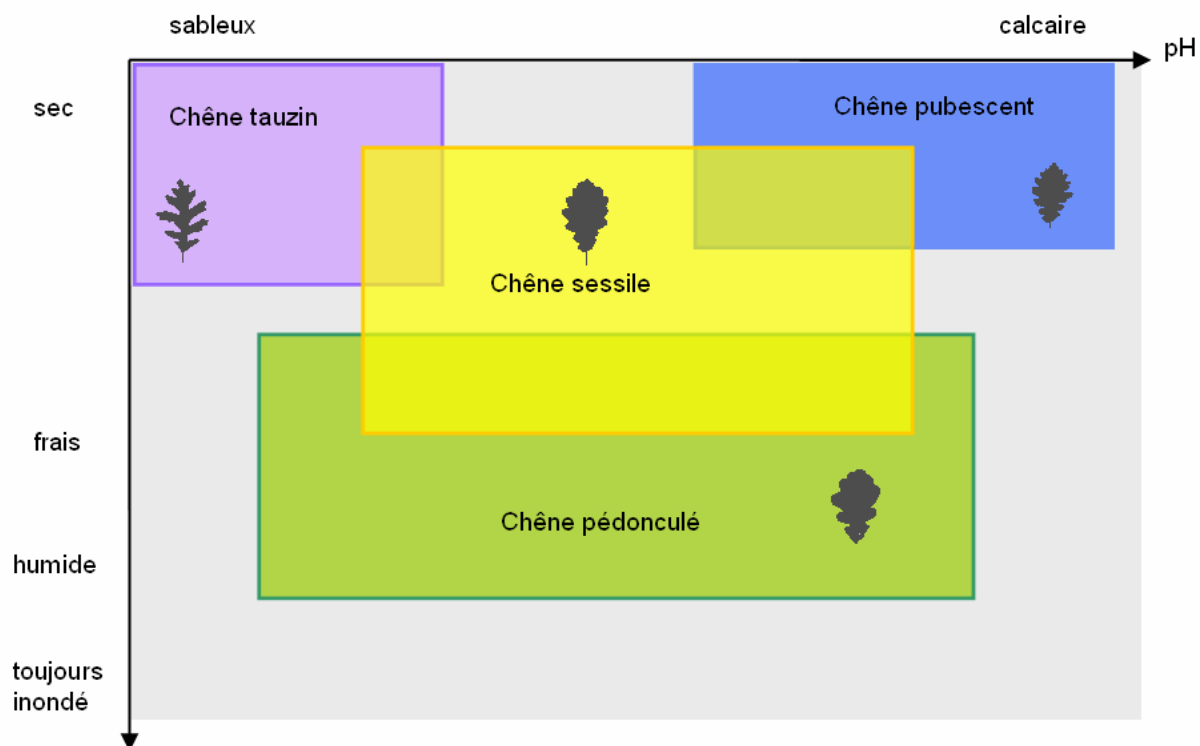


Figure 6 : Préférences écologiques des espèces (d'après Rameau *et al.*, 1994) pour la composition du sol (abscisse) et son niveau d'hygrométrie (ordonnée).

#### 1.2.2.2. La notion d'espèce chez les chênes

La grande variabilité morphologique des chênes a dans un premier temps conduit à définir de multiples espèces. Cependant, au 19<sup>ème</sup> siècle, la prise en compte de cette grande variabilité ainsi que la relative similarité écologique de certaines de ces espèces a conduit à faire des regroupements (de Candolle, 1862a ; de Candolle, 1862b), ce qui a abouti au classement que l'on connaît aujourd'hui (*Quercus robur*, *Q. petraea*, *Q. pubescens*, *Q. toza* ancienne dénomination de *Q. pyrenaica*). La reconnaissance des formes particulières à l'intérieur de ces espèces persiste aujourd'hui avec l'existence de plusieurs sous-espèces qui se distinguent morphologiquement et géographiquement (par exemple, *Quercus robur ssp. pedunculiflora* en Europe de l'est).

L'hybridation pose un problème par rapport à la définition des taxons qui y participent. Selon le concept biologique de l'espèce, les espèces sont isolées reproductivement les unes des autres. Une application stricte de cette définition revient à classer les taxons qui s'hybrident comme une seule espèce, éventuellement composée de sous-espèces (Mayr, 1992). Les espèces de chênes n'étant pas isolées reproductivement (des hybrides peuvent être obtenus en croisements contrôlés et sont suspectés en population) elles peuvent être traitées comme des sous-espèces, par exemple *Quercus robur ssp. robur* et *Quercus robur ssp. petraea* (Kleinschmit & Kleinschmit, 2000). Cependant au regard du concept biologique de l'espèce, les chênes représentent un système de populations très complexes qui sont difficile à délimiter (Burger, 1975). La distinction des quatre formes en tant qu'espèce a donc

persisté notamment parce qu'elle permet de définir quatre unités différenciées du point de vue morphologique et écologique qui correspondent à des utilisations sylvicoles spécifiques. Cette difficulté d'application du concept biologique de l'espèce chez les chênes à amené une nouvelle définition de l'espèce, le concept écologique de l'espèce (Van Valen, 1976), qui insiste sur l'importance de l'écologie comme facteur de spéciation, l'isolement reproducteur n'étant qu'un sous-produit de la divergence écologique. Selon ce concept, nos quatre taxons forment une multi-espèce composée de quatre espèces écologiques qui peuvent échanger des gènes.

La notion de complexe d'espèces (Pernès, 1984) est souvent utilisée pour définir un groupe d'espèces connectées par des flux de gènes. Cependant, cette définition va plus loin puisqu'elle suppose que les flux de gènes interspécifiques font parti du processus évolutif, les barrières reproductives ne sont pas passives (elles ne sont pas seulement un sous-produit de la sélection) mais au contraire, elles sont activement ajustées par la sélection en fonction du contexte évolutif.

Tout au long de cette thèse, nous allons utiliser le terme "espèce" pour désigner les quatre différentes espèces morphologiques ou écologiques qui représentent des entités concrètes au niveau évolutif et adaptatif. J'insiste sur le fait que, d'après la définition standard (biologique) de l'espèce, ce que l'on appellera espèce est en fait une sous-espèce. Finalement, pour marquer cette différence de définition, nous parlerons de complexe d'espèces afin de rappeler que ces espèces ne sont pas forcément isolées reproductivement.

### 1.2.2.3. *Limites du complexe*

En Europe, il existe environ une vingtaine d'espèces regroupées dans deux sections : *Quercus* et *Cerris*. Les espèces de la section *Cerris* comprennent des espèces à aire de répartition méditerranéenne. *Quercus ilex*, *Q. suber* et *Q. coccifera* sont largement répandus en climat méditerranéen, du nord de l'Afrique au sud de l'Europe, *Q. cerris* est une espèce présente de l'Italie à la Turquie. Les espèces des sections *Cerris* et *Quercus* ont une divergence ancienne et sont des unités évolutives bien distinctes.

Dans les régions tempérées on trouve des espèces de la section *Quercus* formant un complexe d'espèces. Deux espèces, *Quercus robur* et *Q. petraea*, ont une large aire de répartition sympatrique en Europe. D'autres espèces à aires de répartitions plus restreintes composent les forêts plus méridionales, parfois en mélange avec *Q. robur* et *Q. petraea*. *Q. pubescens* occupe une large façade méditerranéenne, *Q. frainetto* l'Europe de l'est, *Q. pyrenaica* la façade atlantique ouest européenne et *Q. faginea* la façade atlantique ibérique. Ces espèces forment un complexe d'espèces car elles peuvent s'hybrider naturellement ou artificiellement.

#### 1.2.2.4. *Biogéographie et migrations quaternaire*

La faible diversité en espèces sur le territoire européen comparé au continent américain est attribuée aux successions des cycles glaciaires / interglaciaires au cours du quaternaire qui ont entraîné l'extinction de certaines espèces (Hewitt, 2000). Les espèces qui ont pu migrer lors de ces brusques changements climatiques ont survécu, il y a donc eu une forte pression de sélection pour les caractères intervenant dans la migration. Le dernier épisode de migration (entre 12000 BP et 6000 BP) consécutif au dernier épisode glaciaire a laissé des traces visibles dans la structuration génétique des populations de chênes (Dumolin-Lapègue *et al.*, 1997 ; Petit *et al.*, 2002b ; Petit *et al.*, 2002a). L'analyse de marqueurs chloroplastiques, marqueur à hérédité maternelle chez la plupart des angiospermes et donc transmis par les graines, a permis, grâce à l'analyse conjointe de pollen fossile, de retracer les migrations en réponse à ce changement climatique rapide. L'étude intensive des espèces de chênes (génétique et analyse de pollen fossils) a permis de montrer qu'elles étaient isolées dans des refuges glaciaires localisés dans trois régions en Europe. Ces régions refuges (péninsules ibérique, italienne et balkanique) ont pour caractéristiques communes une localisation méridionale et une topographie montagneuse permettant aux espèces de trouver des conditions favorables à leur survie lors des épisodes glaciaires. Chaque refuge est caractérisé par une lignée chloroplastique regroupant des haplotypes génétiquement proches (Figure 7). La répartition actuelle de ces haplotypes (Figure 7) à l'échelle de l'Europe, montre que chaque refuge a recolonisé une partie du territoire, poursuivant des voies de migration globalement sud-nord. Différentes lignées se sont parfois rencontrées, formant des zones de contact nettement délimitées. Les variants chloroplastiques (haplotypes) ne permettent pas de distinguer les espèces car la plupart des haplotypes sont localement partagés entre espèces. Néanmoins, certaines lignées semblent associées à certaines espèces. Ainsi, le chêne pubescent est associé à la lignée italienne et balkanique, le chêne tauzin possède uniquement des haplotypes de la lignée ibérique. Le chêne sessile est lié à la lignée balkanique alors que le chêne pédonculé est caractérisé par une origine ibérique (Petit *et al.*, 2002a). Cependant, cette association n'est que partielle et les différentes zones refuges aient abrité plusieurs espèces de chênes qui semblent avoir suivi des voies de colonisation parallèles. Les analyses de pollen fossiles ont permis de dater les étapes de cette migration qui s'est réalisée entre 12000 et 6000 BP (Brewer *et al.*, 2002), soit une vitesse de recolonisation moyenne de 500 mètres par an, ce qui est considérable pour des espèces qui ont des temps de génération d'environ 25 ans.

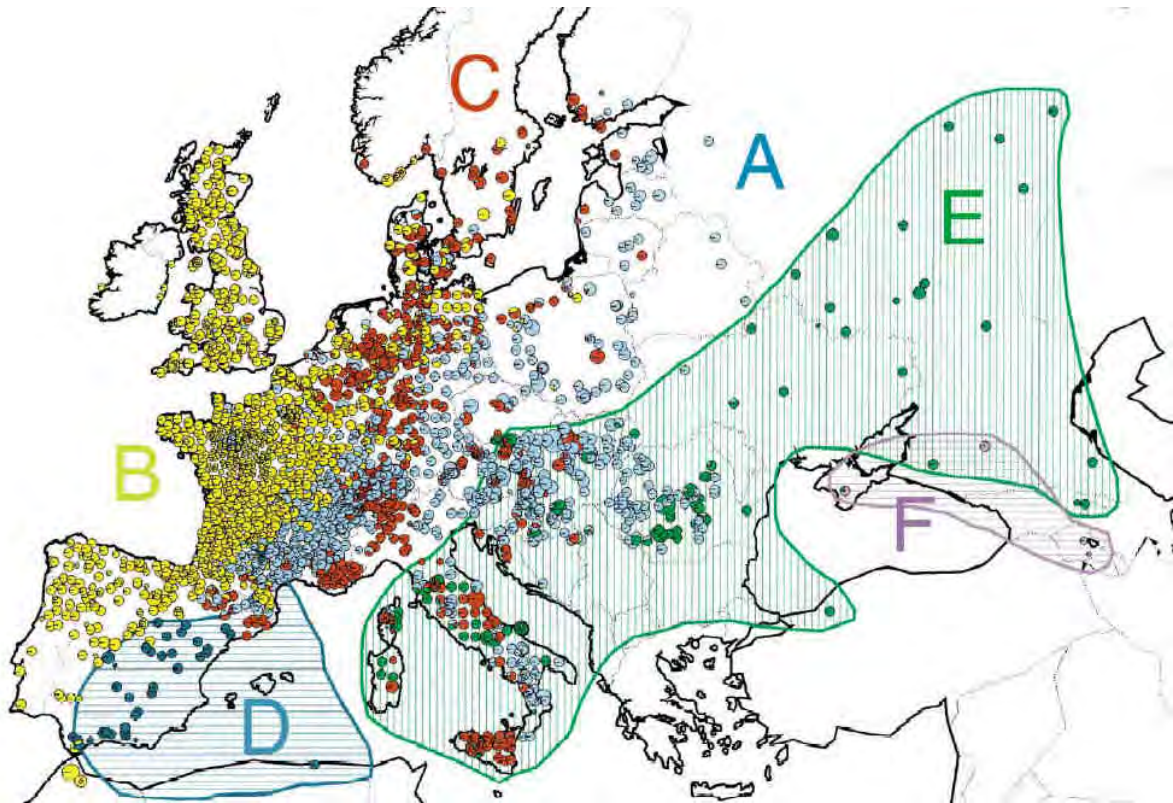


Figure 7 : Répartition des 6 lignées maternelles (lignées chloroplastiques) chez les chênes blancs européens (d'après Petit *et al.*, 2002b). La répartition des lignées A, B et C témoigne des routes de recolonisation depuis les zones refuges méridionale (A : refuges balkaniques, B : refuges ibériques et C : refuges italiens) jusqu'au nord de l'Europe (Petit *et al.*, 2002a). Chaque lignée comprend plusieurs types d'ADN chloroplastique (haplotypes) génétiquement proches les uns des autres.

A l'échelle plus fine (régionale), il a été montré que les haplotypes étaient structurés géographiquement en taches à l'intérieur desquelles un seul haplotype prédomine (Figure 8). Cette structuration est expliquée par des dispersions à longue distance d'un nombre limité de glands (originaire d'un arbre ou d'un petit nombre d'arbres appartenant à une lignée maternelle commune). Ces quelques individus fondent une nouvelle population en avant du front de colonisation avec un seul haplotype. Cette population s'étend jusqu'à être rejointe par le front de colonisation qui avance par dispersion des glands à courte distance. Ainsi, chaque processus de migration à longue distance qui arrive à fonder une population va apparaître sous forme de zone possédant un seul haplotype (Figure 8). La vitesse de migration obtenue par ce modèle de migration composite, comprenant un processus de diffusion par dispersion de proche en proche (locale) des glands ainsi qu'un processus très rare de dispersion à (très) longue distance, a été testé par des simulations (LeCorre *et al.*, 1997). Elles montrent que la prise en compte d'événements très rares de migration à longue distance est nécessaire pour obtenir la vitesse de recolonisation observée. Le geai des chênes (*Garrulus glandarius*) se nourrit principalement de glands qu'il récolte et enterre pour faire des réserves (Bossema, 1979 ; Ducouso & Petit, 1994). Cette espèce d'oiseau est capable de disperser un grand nombre de glands (plus de 4000 glands par an et par individu) parfois sur de longues distances. C'est donc un des vecteurs possibles des dispersions rares à longue distance. Le rôle de l'homme dans cette colonisation a parfois été suggéré (Kremer, 2007), il a

lui aussi subi les conditions climatiques postglaciaires et a migré vers les nouvelles terres déglacées. De plus, des écrits anciens indiquent que les chênes servaient à divers usages : source de nourriture (glands), utilisation du bois pour la fabrication d'objets.

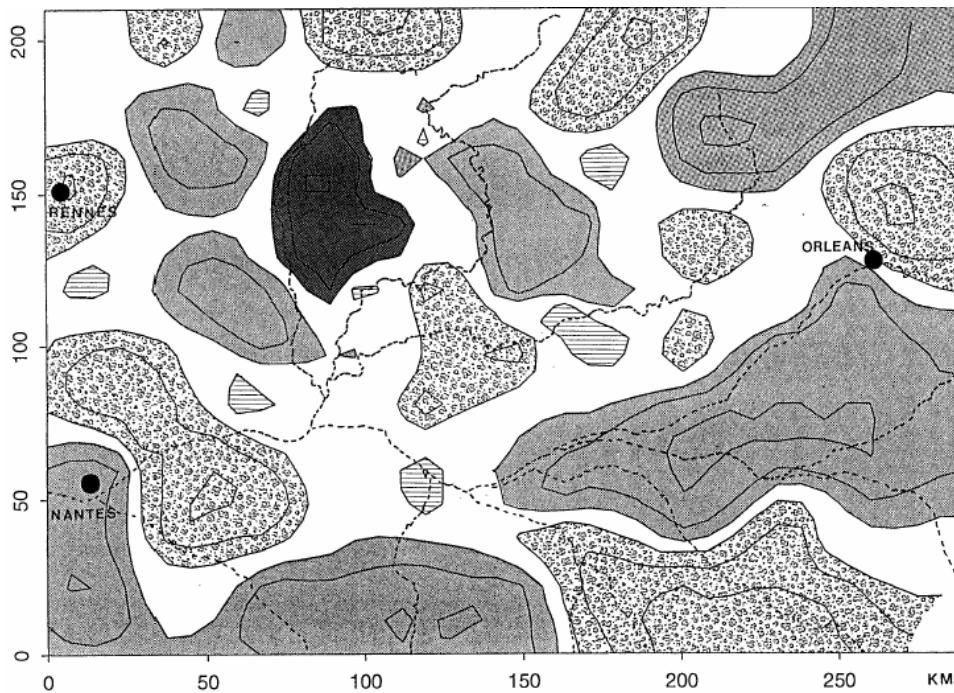


Figure 8 : Carte des fréquences des haplotypes (d'après Petit *et al.*, 1997) à l'échelle régionale (200 x 300 km) estimées à l'aide de méthodes géostatistiques (isolignes de fréquences : 0.6, 0.8 et 1.0). Les quatre haplotypes identifiés sont représentés par différents niveaux de gris ou symboles. Les données des chênes pédonculés et sessiles sont combinées (voir texte).

#### 1.2.2.5. L'hybridation comme moyen de migration : le modèle pédonculé / sessile

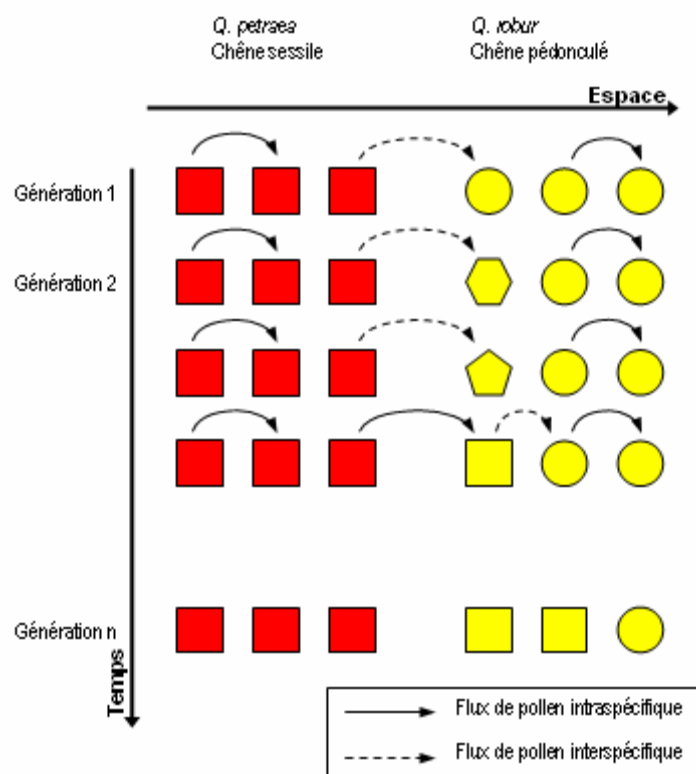
A une échelle régionale, les espèces des chênes présentes en sympatrie (forêt mixte) partagent le même haplotype (Figure 8; Petit *et al.*, 1997). Ce partage systématique d'un même haplotype local entre espèces a également été montré à une échelle géographique plus large (Petit *et al.*, 1993 ; Dumolin-Lapègue *et al.*, 1997 ; Dumolin-Lapegue *et al.*, 1999 ; Petit *et al.*, 2002c).

Parallèlement, des analyses de croisements contrôlés entre chênes pédonculés et sessiles montrent que ces espèces sont interfertiles et que l'hybridation est directionnelle : le pollen de chêne sessile donne une descendance nombreuse lorsqu'il est appliqué sur des fleurs femelles de chêne pédonculé alors que l'inverse produit un nombre d'hybride faible et des glands peu viables (Steinhoff, 1993 ; Steinhoff, 1998 ; Kleinschmit & Kleinschmit, 2000). Cette asymétrie dans les croisements interspécifiques pourrait créer une hybridation directionnelle en forêt. Enfin, des marqueurs génétiques nucléaires de différentes natures (isozymes (Gomory *et al.*, 2001), RAPD et SCAR (Bodenes *et al.*, 1997), AFLP (Coart *et al.*, 2002 ; Mariette *et al.*, 2002), microsattellites (Muir *et al.*, 2000)) montrent que les chênes pédonculés et sessiles sont peu différenciés génétiquement. Aucun marqueur diagnostique n'a été identifié et les espèces diffèrent seulement par leurs fréquences alléliques. Cette différenciation faible,



malgré de relativement fortes différences phénotypiques (morphologiques et écologiques), pourrait être lié à une hybridation actuelle ou ancienne entre ces espèces.

L'ensemble de ces résultats associés aux caractéristiques écologiques des deux espèces a conduit à proposer un modèle d'hybridation unidirectionnel dans lequel le chêne sessile est capable de migrer par son seul pollen (Figure 9). Si les croisements interspécifiques sont unidirectionnels et récurrents, les hybrides formés vont être fécondés par du pollen sessile ce qui va permettre au phénotype sessile de réapparaître au fur et à mesure des générations successives de rétrocroisements (Figure 9). Cette dynamique d'hybridation pourrait ainsi renforcer la succession forestière, le chêne sessile étant une espèce arrivant après le chêne pédonculé, espèce plus pionnière, lors de la succession forestière (Petit *et al.*, 2003).



**Figure 9** : Modèle d'hybridation interspécifique et de migration du chêne sessile par le pollen (d'après Kremer *et al.*, 2002b). Les ellipses représentent les chênes pédonculés, les carrés les chênes sessiles et les formes intermédiaires sont des chênes hybrides. Les couleurs représentent l'haplotype des arbres. Conditions initiales : deux populations mono-spécifiques séparées géographiquement et ayant des génomes chloroplastiques différents (rouge et jaune). La migration du chêne sessile par le pollen se caractérise par une conservation au fil des générations de l'haplotype du chêne pédonculé maternel.

Des analyses fines du génome, utilisant un grand nombre de marqueurs génétiques couvrant l'ensemble des chromosomes (Scotti-Saintagne *et al.*, 2004b), montrent que la différenciation interspécifique est particulièrement élevée dans plusieurs zones (12% des marqueurs étudiés) localisés dans 9 des 12 chromosomes du chêne alors qu'elle est très faible voire nulle ailleurs (Bodenes *et al.*, 1997 ; Saintagne *et al.*, 2004). De plus, des QTLs (Quantitative Trait Loci : zones du génome impliquées dans la variation quantitative d'un caractère, Orr, 1998) liés à des caractères

morphologiques différenciant les espèces ont été localisés sur 6 chromosomes et co-localisent avec certains des marqueurs fortement différenciés. Ces locus particuliers (*outliers loci*) qui s'écartent de l'attendu neutre montrent que la sélection pourrait être la cause de cette forte différenciation. Les zones fortement différenciées ont une taille moyenne de 2 centiMorgans (cM) avec des blocs pouvant atteindre 5 à 10 cM (Scotti-Saintagne *et al.*, 2004a). Ces blocs correspondent à des zones d'influence de gènes potentiellement soumis à une sélection divergente entre espèces (effet d'entraînement, *hitchhiking effect*, lié à la faible distance et donc à la faible recombinaison entre marqueurs neutres et gènes sélectionnés). Ces résultats sont en accord avec la théorie génique de la divergence des espèces (Rieseberg & Burke, 2001 ; Wu, 2001) selon laquelle un petit nombre de locus présente un haut niveau de différenciation lié à la sélection divergente alors que le reste du génome est homogénéisé par les flux de gènes.

#### 1.2.2.6. Controverse : hybridation ou partage de polymorphisme ancien

Récemment, l'étude de locus microsatellites dans plusieurs populations de chênes pédonculés et sessiles en Europe (Muir & Schlotterer, 2005) a remis en cause l'importance du phénomène d'hybridation entre ces deux espèces. Les auteurs proposent trois hypothèses pour expliquer la faible différenciation génétique entre espèces :

- Faible flux de gènes interspécifique : le taux de flux de gènes serait trop faible pour compenser la dérive à l'intérieur de chaque espèce (notons que selon cette vision, les espèces divergent par dérive et non par sélection divergente, ce qui est une vision quelque peu archaïque du phénomène de spéciation).
- Sélection sur un nombre limité de locus qui maintiennent l'intégrité des espèces alors que le flux de gènes interspécifique peut être important et homogénéiser le reste du génome.
- Partage du polymorphisme ancien (l'hypothèse la plus probable selon les auteurs) : la différenciation génétique interspécifique concerne l'ensemble du génome (et non certaines zones seulement) et la faible différenciation générale observée serait liée au partage d'un polymorphisme ancestral entre les deux espèces dont la divergence serait récente.

Cependant, une bonne partie de la bibliographie sur l'hybridation et la différenciation génétique interspécifique chez les chênes n'est pas considérée dans cette étude et l'hypothèse du partage du polymorphisme ancien ne peut expliquer la présence d'hybridation et l'hétérogénéité de la différenciation à l'échelle du génome. Une réponse (Lexer *et al.*, 2006) à cette article immédiatement suivie d'une contre-réponse (Muir & Schlotterer, 2006) laissent les auteurs sur leurs point de vue respectifs, alors que des analyses en populations montrent que l'hybridation est bien une réalité au sein de ce complexe d'espèces (Valbuena-Carabana *et al.*, 2005 ; Curtu *et al.*, 2007 ; Valbuena-Carabana *et al.*, 2007).

### 1.3. Objectif de la thèse

Dans ce contexte scientifique l'objectif de cette thèse est de quantifier l'hybridation entre quatre espèces de chênes afin d'estimer ces conséquences éventuelles sur l'évolution des espèces.

#### Méthode générale :

Nous allons utiliser des méthodes permettant d'étudier l'hybridation récente (quelques générations) et comparer les résultats à différents stades de vie (graine, jeune plant et adulte) afin de comprendre la dynamique des flux de gènes au sein de ce complexe d'espèces.

Les démarches mises en œuvre ainsi que les différentes expériences réalisées lors de cette thèse sont décrites dans trois parties composées chacune de deux chapitres. Cependant, chaque chapitre est autonome (structuré comme un article) et peut être lu indépendamment.

#### Démarche :

La **PARTIE A** décrit les développements méthodologiques nécessaires pour identifier les espèces de chênes sur des bases génétiques.

- Pendant cette thèse, nous avons analysé l'hybridation en conditions artificielles et dans des populations naturelles. Ces différentes expériences sont basées sur l'utilisation de marqueurs génétiques afin d'identifier les relations de parentés et de délimiter les espèces. Un protocole expérimental a été développé pour analyser rapidement l'information fournie par plusieurs marqueurs génétique (microsatellites). Cette technique a fait l'objet d'une publication (Lepais *et al.*, 2006b ; Annexe A) et des informations complémentaires sont présentées dans le **chapitre 2**. L'accent est mis en particulier sur l'estimation du taux d'erreur de ces expériences ainsi que sur leur prise en compte dans les analyses ultérieures.
- La délimitation et l'identification des espèces grâce aux outils de la génétique a connu un regain d'intérêt ces dernières années, principalement grâce à deux facteurs : la disponibilité de marqueurs génétiques très polymorphes dont l'information peut être acquise rapidement ainsi que le développement récent de nouvelles méthodes d'analyses. La génétique des populations fournit en particulier plusieurs outils permettant d'assigner les individus à une espèce à partir des seules données génétiques. Ces méthodes sont particulièrement bienvenues pour l'étude des chênes puisqu'elles permettent de délimiter les espèces et surtout d'identifier les hybrides, sans a priori sur la morphologie. Dans le **chapitre 3**, après une revue rapide des différentes méthodes disponibles et de leurs applications, plusieurs d'entre elles sont comparées avec des données réelles. Afin d'estimer la fiabilité de ces analyses, des ré-échantillonnages des données sont testés pour estimer l'impact du nombre de locus et d'individus sur les performances d'assignation des individus aux espèces. Enfin, des simulations sont utilisées

afin de tester la performance de la méthode pour assigner les individus aux espèces et identifier les hybrides.

La **PARTIE B** présente l'application de l'assignation génétique en population naturelles qui nous a permis de quantifier le taux d'hybridation et ses conséquences sur les peuplements de chênes.

- Dans le **chapitre 4**, la méthode d'assignation est appliquée à plusieurs populations à travers la France afin de quantifier et de caractériser l'hybridation entre espèces par paires. Des analyses détaillées sont menées dans plusieurs parcelles afin de comparer les dynamiques d'hybridation. La prise en compte de paramètres locaux permet d'expliquer la dynamique d'introgession. Cette étude va faire l'objet d'un article pour le journal *Molecular Ecology*.
- L'hybridation récente détectée dans les populations peut avoir des conséquences sur le phénotype des individus ainsi que sur la diversité génétique des espèces et sa structuration spatiale. Pour tester ces hypothèses, une série d'analyses concentrées sur une parcelle comprenant les quatre espèces (Briouant) est menée dans le **chapitre 5**. L'analyse de la morphologie foliaire permet de comparer les assignations génétiques et morphologiques et de caractériser la forme des hybrides. Les données morphologiques acquises lors de cette thèse ont également servi à l'étude comparative de plusieurs méthodes d'analyses des formes de feuilles de chênes grâce à une collaboration internationale valorisé par un article (Viscosi *et al.*, 2008b ; Annexe C). Enfin, l'effet de l'hybridation sur l'organisation de la diversité génétique entre espèces et sur la structuration génétique spatiale est étudié à l'échelle de ce peuplement.

Dans la **PARTIE C** nous nous intéressons au système de reproduction des chênes, un paramètre important pour la compréhension de l'hybridation entre ces espèces.

- Le **chapitre 6** présente les résultats d'expériences de croisements contrôlés permettant l'étude de certains paramètres du système de reproduction des espèces. L'isolement reproducteur entre espèces peut avoir des causes génétiques dont un certain nombre peut être mis en évidence grâce à des expériences d'hybridation artificielle. Nous avons soumis nos quatre espèces à une expérience de croisements contrôlés dans le but premier d'identifier les compatibilités et incompatibilités entre espèces. Des analyses plus fines ont permis de quantifier le rôle relatif de plusieurs barrières reproductives. De plus, le suivi des descendants en pépinière nous a permis de calculer les taux de germination et de survie des différents croisements et donc d'avoir une estimation de la viabilité des hybrides aux stades précoces de vie et d'identifier des barrières postzygotiques. Ces résultats feront l'objet d'un futur article.
- Dans le **chapitre 7** nous nous intéressons au système reproducteur des espèces et des hybrides en peuplement naturelle. La phénologie du débourrement foliaire, corrélée à la phénologie de floraison et donc à une barrière pré-reproductive importante, a été suivie pendant deux années

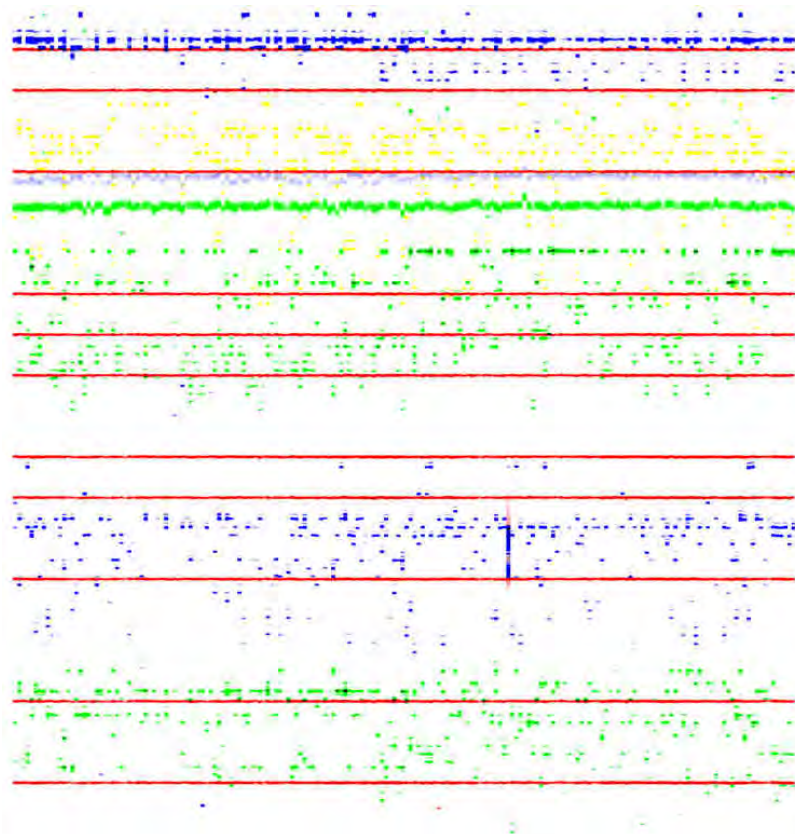
consécutives. Cette analyse est motivée par la présence du chêne tauzin, espèce particulièrement tardive. Ces données nous ont permis d'étudier les différences entre espèces, le phénotype des hybrides pour ce caractère et leurs conséquences éventuelles sur la dynamique d'hybridation. Les conséquences de l'hybridation dans ce complexe d'espèces vont dépendre en grande partie du comportement reproducteur des hybrides. Une image des flux de gènes réalisés lors d'une saison de reproduction a été obtenue suite à l'analyse de descendances maternelles (glands récoltés sur plusieurs arbres) et de plantules à Briouant. Des analyses de paternité et de parenté sont menées, complétées par des assignations génétiques, afin d'identifier directement les différents types d'événements reproducteurs (hybridations, croisements intraspécifiques,...). Cette approche directe est complétée par une approche indirecte permettant de caractériser plus finement la reproduction dans ce complexe d'espèces. Ces résultats feront l'objet d'un futur article.

Une conclusion générale (**chapitre 8**) rappelle les principaux résultats et les intègre dans la problématique générale à laquelle cette thèse tente de répondre.

# **PARTIE A - MÉTHODOLOGIE**



## Chapitre 2- Marqueurs génétiques discriminant les espèces de chênes





Page précédente : Résultats bruts d'une analyse génétique vus sous forme de gel contenant les allèles de 5 microsatellites pour 192 individus. Chaque colonne contient un individu, les lignes rouges représentent des tailles connues grâce à un marqueur de taille.

## 2.1. Introduction

La génétique des populations étudie les variations des fréquences alléliques (variation des différentes formes d'un gène) au cours du temps sous l'influence de la dérive génétique, de la sélection naturelle, de la mutation et de la migration afin de comprendre l'adaptation et l'évolution des espèces et des populations. Ce domaine scientifique s'est développé depuis les années 30 sous l'influence de Haldane, Fisher et Wright qui ont développé les concepts permettant d'appliquer la génétique mendélienne aux populations naturelles. Les marqueurs utilisés pour estimer les variations d'origines génétiques ont rapidement évolués. Les premiers marqueurs moléculaires, les iso-enzymes (différentes formes d'une même enzyme correspondant à différents allèles d'un gène) ont permis d'estimer la structure de la variabilité génétique au sein et entre populations et d'étudier les processus de reproduction et de dispersion. Cependant, le développement plus récent de nombreux types de marqueurs basés sur l'ADN a donné accès à davantage de variation et donc un plus grand pouvoir résolutif. Parmi ces nouveaux marqueurs ADN, les microsatellites sont les plus souvent utilisés car leur taux de mutation élevé leur confère un grand nombre d'allèles et donc un polymorphisme important permettant des applications variées. Les microsatellites sont des répétitions de séquences simples constituées de 1 à 6 nucléotides (on les nomme aussi SSR pour *Simple Sequence Repeats* ou STR pour *Short Tandem Repeats*). Cette structure répétée confère une certaine instabilité lors de la réplication de l'ADN ce qui explique leur taux de mutation élevé. Ces marqueurs sont présents chez la plupart des espèces et sont répartis dans l'ensemble du génome, aussi bien dans les parties codantes (dans les gènes) que dans les parties non codantes du génome. Dans notre cas, nous allons utiliser des microsatellites génomiques (non localisés dans des gènes) considérés comme neutres pour les applications en génétique des populations. Le grand polymorphisme des microsatellites nous permet de distinguer chaque individu et donc d'estimer les flux de gènes contemporains.

Une cinquantaine de microsatellites ont été développés chez les chênes pédonculés (Kampfer *et al.*, 1998) et sessiles (Steinkellner *et al.*, 1997). Ces marqueurs ont été utilisés pour étudier la structure génétique des populations (Degen *et al.*, 1999; Jensen *et al.*, 2003; Mariette *et al.*, 2002; Muir *et al.*, 2004; Streiff *et al.*, 1999; Streiff *et al.*, 1998) ainsi que pour construire des cartes génétiques notamment celle du chêne pédonculé (Barreneche *et al.*, 1998). Certains de ces marqueurs ont été amplifiés avec succès chez le châtaignier (Barreneche *et al.*, 2004; Boccacci *et al.*, 2004) permettant d'obtenir des points relais afin de comparer les cartes génétiques des chênes et des châtaigniers. Cette analyse a permis de montrer que l'ordre des marqueurs génétiques le long de la carte génétique était remarquablement conservé entre ces espèces (Barreneche *et al.*, 2004; Casasoli *et al.*, 2006). Cette application montre que bien que le nombre de répétitions du motif de base est variable entre individus et entre espèces, les séquences flanquant les microsatellites restent relativement constantes ce qui permet d'amplifier les mêmes marqueurs dans différentes espèces. Les 6 microsatellites utilisées par Réjane Streiff lors de sa thèse (Streiff, 1998) sont devenus les marqueurs de référence pour l'analyse

des flux de gènes chez les chênes pédonculés et sessiles à travers l'Europe (projet européen OAKFLOW). Une partie de ces locus a été récemment intégrée dans un protocole d'amplification simultanée (PCR multiplexe ; Dzialuk *et al.*, 2005) facilitant leur analyse en masse. L'homoplasie, le fait que deux allèles détectés comme identiques aient une origine différente, concerne certains de ces locus ( par exemple Qp9, Curtu *et al.*, 2007) ce qui les rend difficilement utilisables pour des approches phylogénétiques (le critère d'identité par descendance n'étant plus respecté). Cependant, pour étudier des phénomènes micro-évolutifs, c'est-à-dire intervenant depuis quelques générations, le biais créé par l'homoplasie n'aura pas beaucoup de conséquences (Selkoe & Toonen, 2006). Un problème plus conséquent lié à l'utilisation des marqueurs génétiques, en particulier des microsatellites, concerne les erreurs liées à l'acquisition des données génotypiques. Un taux d'erreur élevé peut sérieusement biaiser les interprétations, cependant la quantification et la prise en compte de ces erreurs a fait l'objet de développements récents.

Dans ce chapitre, nous présentons comment nous avons utilisé les dernières technologies disponibles pour mettre au point une technique permettant l'analyse rapide de plusieurs microsatellites chez les chênes. Le protocole élaboré est présenté en annexe (ANNEXE A), et l'accent est mis ici sur des informations complémentaires comme le choix des microsatellites, le fonctionnement du séquenceur capillaire pour une application en génotypage, la normalisation des données entre projets ainsi que l'estimation du taux d'erreur de la méthode.

## **2.2. Mise au point d'une méthode de génotypage rapide**

### **2.2.1. Sélection de microsatellites différenciant les espèces de chênes**

Une cinquantaine de microsatellites étant disponibles chez les chênes pédonculés et sessiles, une sélection a été nécessaire. La première propriété essentielle d'un bon marqueur génétique, ce qui va conditionner son contenu en information et donc son utilité, est qu'il soit polymorphe c'est-à-dire qu'il présente plusieurs allèles différents. Les microsatellites sont des marqueurs qui sont en général très polymorphes car ils ont un taux de mutation rapide lié à leur nature répétée. De plus, les chênes sont des espèces à population de grande taille et à fort flux de gènes ce qui se traduit au niveau génétique par une très forte diversité génétique et donc un polymorphisme très élevé pour les marqueurs neutres. Le niveau de polymorphisme des microsatellites n'a donc pas été dans notre cas le premier critère de sélection des marqueurs à utiliser puisque un polymorphisme élevé en population est attendu pour tous les marqueurs développés.

Notre but principal a été d'utiliser des microsatellites afin de reconnaître les espèces de chênes grâce à des analyses d'assignation génétique. Pour atteindre cet objectif, il nous a fallu sélectionner les microsatellites qui présentaient le maximum de différences entre espèces afin d'augmenter le pouvoir résolutif de l'assignation génétique. Nous nous sommes concentrés sur l'indice de différenciation entre espèces ( $F_{ST}$ ) comme critère permettant de sélectionner les microsatellites les plus informatifs pour notre étude (Tableau 1). Certains locus ont des taux de différenciation entre chêne sessile et pédonculé

particulièrement élevés (Qr96, Qr112 et Qr11 dans une moindre mesure). Ces trois locus ont pu être inclus dans le protocole PCR. Nous avons ensuite cherché des locus qui présentaient un bon taux d'amplification et donnait des résultats clairs et exploitables. Plusieurs locus ont été testés sans succès (Tableau 1). Les locus inclus dans le protocole final comprennent quelques microsatellites avec de fort niveau de différenciation interspécifique alors que les autres locus ont des niveaux de différenciation moyen voir faible (Tableau 1). Il faut noter que nous avons effectué notre sélection sur le niveau de différenciation génétique ( $F_{ST}$ ), mais d'autres indices prenant en plus en compte le nombre d'allèles et leur distribution entre espèces peuvent être de meilleurs critères de sélection que la simple différenciation (Banks *et al.*, 2003 ; Rosenberg, 2005). Ces nouvelles méthodes, incluses dans le logiciel WhichRun (Banks & Eichert, 2000), seront particulièrement utiles lorsque le nombre de locus disponible sera plus important (développement de marqueurs grâce aux ressources génomiques notamment).

Tableau 1 : Liste des microsatellites testés pour leur différenciation génétique entre chêne sessile et pédonculé (Pablo Goikoetxea / projet européen OAKFLOW, données non publiées).

Locus	LG	Fst	Multiplexe
<b>QrZAG96</b>	<b>10</b>	<b>0.218</b>	<b>inclus</b>
QrZAG30	12	0.129	
<b>QrZAG112</b>	<b>12</b>	<b>0.126</b>	<b>inclus</b>
<b>QpZAG110</b>	<b>8</b>	<b>0.085</b>	<b>inclus</b>
QrZAG73	5	0.083	
<b>QrZAG11</b>	<b>10</b>	<b>0.082</b>	<b>inclus</b>
QpZAG46	2	0.057	testé
QpZAG119	2	0.055	testé
<b>QrZAG87</b>	<b>2</b>	<b>0.049</b>	<b>inclus</b>
<b>QrZAG39</b>	<b>5</b>	<b>0.042</b>	<b>inclus</b>
QrZAG75	nd	0.04	
QrZAG101	1	0.037	
QpZAG3/64	2	0.036	testé
<b>QrZAG7</b>	<b>2</b>	<b>0.031</b>	<b>inclus</b>
QrZAG44	7	0.029	
QpZAG15	9	0.027	
QpZAG7	nd	0.021	testé
QpZAG1/2	nd	0.019	
QrZAG8	12	0.0184	
<b>QrZAG65</b>	<b>10</b>	<b>0.015</b>	<b>inclus</b>
QrZAG31	9	0.014	
QrZAG15	9	0.013	
QrZAG108	7	0.013	testé
QpZAG102	9	0.013	testé
QrZAG74	1	0.011	
QpZAG16	6	0.011	testé
<b>QrZAG20</b>	<b>1</b>	<b>0.01</b>	<b>inclus</b>
QpZAG58	5	0.009	
<b>QrZAG5b</b>	<b>6</b>	<b>0.007</b>	<b>inclus</b>
<b>QrZAG20</b>	<b>1</b>	<b>nd</b>	<b>inclus</b>
QrZAG90	9	nd	testé
QpZAG108	2	nd	testé
QrZAG58_new (*)	2	nd	testé
QrZAG58	2	nd	testé

LG : groupe de liaison (*linkage group*) sur la carte génétique du chêne pédonculé (Barreneche *et al.*, 1998) ; *Fst* : différenciation génétique interspécifique ; Multiplexe : sélection de certains locus pour test en amplification multiplexe (testé : non conservé car problèmes d'amplification ; inclus : locus, indiqué en gras, conservé dans le protocole final) ; nd : non déterminé. (\*) nouvelles amorces (Muir *et al.*, 2004).

### 2.2.2. Sélection de microsatellites sur critères techniques

Le choix des locus à intégrer est dicté par l'application (les plus différenciés entre espèce dans notre cas) mais aussi par des contraintes techniques comme la taille des allèles attendus ou la température d'hybridation des amorces par exemple. En effet, plusieurs microsatellites vont être amplifiés simultanément (technique de PCR multiplexe : amplification de plusieurs fragments en une seule réaction d'amplification PCR (Kimpton *et al.*, 1993)) ce qui impose que les amorces aient des températures d'hybridation relativement proches et des tailles d'allèles suffisamment éloignées (voir Figure 3). L'information disponible sur ces marqueurs était regroupée dans un fichier de tableur rendant le tri d'information fastidieux. Nous avons donc organisé ces informations dans une base de données hébergée sur un serveur local (<http://mururoa/SSR/chenes/bdd-ssr.php3>) et interfacée par une

page de type internet programmée en langage PHP permettant d'interroger la base de données grâce à des requêtes MySQL (Figure 10). Cet outil permet de trouver rapidement les locus dont les caractéristiques sont compatibles avec les contraintes techniques imposées (recherche multicritère avancée). Les bases de données devraient devenir un outil efficace pour regrouper et connecter différents types d'informations (marqueurs génétiques : la base présentée ici ; informations sur les échantillons : outils de traçabilité existant ; mais aussi des données telles que génotypes, informations autres (géographiques, mesures diverses...), données sur la cartographie génétique, etc...). Outre l'aspect traçabilité et stockage rendu indispensable vu la quantité de données générée par les techniques d'analyses modernes, le développement de bases de données devrait de plus permettre de croiser les données issues de plusieurs types d'analyses et ainsi faciliter l'échange et le partage d'informations entre différents domaines de recherche.

Requête par défaut listant l'ensemble des 196 microsattellites potentiellement amplifiables chez les chênes, dont 50 ont été développés chez les chênes pédonculé et sessile. Le résultat des requêtes apparaît sous forme d'un tableau dont le nombre de colonne dépend des informations souhaitées par l'utilisateur (ici : nom du locus, espèce d'origine, localisation sur la carte génétique du chêne pédonculé, motif répété, température d'hybridation des amorces, taille minimale et maximale attendue des allèles, nombre d'allèles détectés/nombre d'individus testés et référence).

Requête SQL par défaut :

```
SELECT f_name, species, Qc_lg, motif, Tm, min_size, max_size, range_size, num_allèles, Reference FROM sat
```

Envoyer

196 résultats

f_name	species	motif	Tm	min_size	max_size	range_size	num_allèles	Reference
QpZAB10	Quercus petraea	not mapped	50	97	117	20	7/27	Steinkollnes, et al 1997
QpZAG15	Quercus petraea	7-(GT)16(GA)9	55	160	190	30	7/37	Steinkollnes, et al 1997
QpZAG162	Quercus petraea	10-(A)54	58	111	145	32	6/19	Steinkollnes, et al 1997
QpZAG164	Quercus petraea	2-(AG)21	59	164	199	35	9/34	Steinkollnes, et al 1997
QpZAG7	Quercus petraea	not mapped	52	140	160	20	8/20	Steinkollnes, et al 1997
QpZAG8	Quercus petraea	7-(AG)12	50	182	210	28	11/28	Steinkollnes, et al 1997
QpZAG15	Quercus petraea	9-(AG)23	56	106	152	44	11/43	Steinkollnes, et al 1997
QpZAG16	Quercus petraea	6-(AG)19	50	147	164	17	12/36	Steinkollnes, et al 1997

Exemple de requête-type permettant d'identifier des microsattellites répondant à certaines contraintes techniques (tailles des fragments amplifiés et température d'hybridation des amorces) et d'afficher la séquence des amorces permettant d'amplifier les locus :

Requête SQL :

```
SELECT f_name, species, motif, Tm, min_size, max_size, Rsequence, Fsequence, Reference FROM sat WHERE min_size>120 AND max_size<220 AND Tm>50 AND Qc_carré='yes' ORDER BY min_size ASC
```

Envoyer

4 résultats

f_name	species	motif	Tm	min_size	max_size	Rsequence	Fsequence	Reference
QpZAG96	Quercus robur	37(G)20	53	133	194	CCC AGT CAC ATC CAC TAC TOT CC	GGT TGG GAA AAG GAG ATC AGA	Caubère, et al 1998
QpZAG15	Quercus petraea	7-(GT)16(GA)9	55	160	190	GCT TGA GAG TTA AGA TTT GT	GCA ACA CCC TTT AAG TAC CA	Steinkollnes, et al 1997
EMC113	Castanea sativa	6(CAA)8	59	182	188	TAATCCAAATACGGGACAG	TGATATGACATTTCACCTTGATT	Burca et al 2003
EMC11	Castanea sativa	10(CCA)6	53	170	190	TTC GTA TGC ACC TGC GAA TGC TC	GAC GGA AGA AGA CCC TAC CTC AG	Kern Buisson

"Trouver la séquence des amorces de microsattellites ayant une taille de produit PCR comprise entre 120 et 220 pb, une température d'hybridation des amorces de 55°C et dont la position sur la carte génétique est connue" : 4 résultats.

Figure 10 : Principe de fonctionnement de la base de données microsattellites chêne.

### 2.2.3. Mise au point du protocole

Le polymorphisme élevé des microsatellites rend cet outil particulièrement attractif pour les études en génétique des populations. Bien que relativement facile à génotyper comparé à d'autres types de marqueurs génétiques, certaines étapes de l'expérience, telles que l'extraction d'ADN ou la migration et la détermination de la taille des fragments amplifiés (allèles), sont fastidieuses et ne permettent de traiter qu'un nombre limité d'individus en même temps (l'ADN de 36 voire 48 individus était extrait par jour et le génotypage d'un microsatellite pour 96 individus nécessitait plusieurs heures de travail au laboratoire et d'analyse méticuleuse afin de déterminer la taille de chaque allèle). Afin de répondre aux objectifs de cette thèse, la première étape a consisté à optimiser le débit d'extraction d'ADN en adaptant le protocole d'extraction en tube Eppendorf 2mL couramment utilisé chez les chênes (Dumolin *et al.*, 1995) à l'utilisation de microtubes 1.2mL sur portoir au format plaque 96 puits. La principale difficulté a été de trouver des microtubes suffisamment solides pour résister au broyage et aux différents solvants. Il a ensuite suffi de réduire les volumes des différents produits utilisés afin d'obtenir un protocole permettant d'obtenir de l'ADN de bonne qualité pour 192 individus en une journée de travail au laboratoire (Figure 11A).

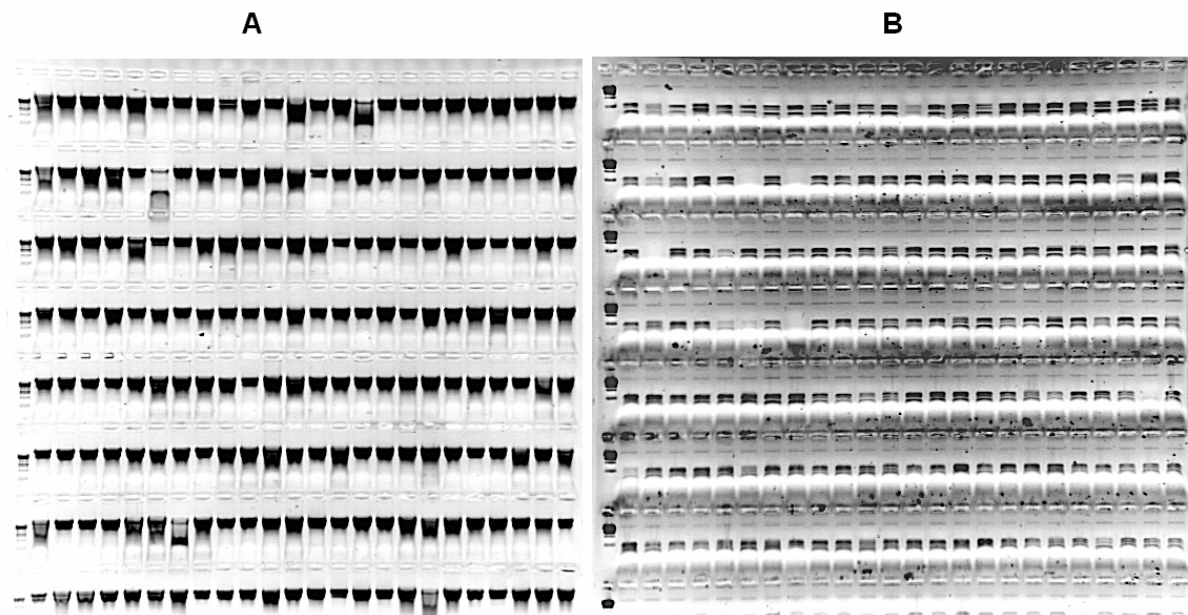


Figure 11 : (A) Migration sur gel d'agarose 1.2% de l'ADN extrait de 192 glands de chênes issus des croisements contrôlés et (B) migration sur gel d'agarose 3% des produits PCR obtenus par amplification simultanée des 5 microsatellites du Kit1 pour les même 192 individus (Plaques 7 et 8 hybcc2006). La résolution de la matrice d'agarose n'est pas suffisante pour distinguer les différents locus qui apparaissent ici sous forme de deux bandes à cause de leurs tailles très proches. La migration sur gel d'agarose est nécessaire pour vérifier qu'une amplification homogène a été obtenue pour tous les individus avant d'effectuer la migration des fragments au séquenceur capillaire.

L'acquisition par le laboratoire d'un séquenceur 96 capillaires Amersham MegaBACE 1000 a permis d'envisager d'augmenter considérablement le débit des analyses génétiques (Koumi *et al.*, 2004). Cette génération de séquenceur utilise deux types de laser permettant de distinguer des fragments d'ADN marqués par quatre fluorochromes différents (contre un seul pour les "anciens" séquenceur à plaque de

type Licor), ce qui permet de distinguer les allèles de même taille de trois locus différents (voir Figure 12). Après migration des fragments dans la matrice de polyacrylamide, un logiciel spécifique (Fragment Profiler) permet de déterminer précisément la taille de chaque allèle de façon semi-automatique, ce qui rend cette étape nettement plus fiable et rapide (Figure 13). Cependant, cette étape, nommée *allele scoring* ou *binning*, qui consiste à transformer une taille d'allèle, variable continue (en paires de bases) et relative (puisque mesurée par un appareil spécifique), en nom d'allèle (données discrète) n'est pas une tâche aisée et demande une rigueur et une attention particulière (Ghosh *et al.*, 1997 ; Presson *et al.*, 2006). Dans notre cas, nous avons défini des filtres logiciels spécifiques afin d'analyser toutes les données de la même façon (concrètement, deux fichiers *peakfilter* ont été paramétrés une fois pour toute et mis à disposition sur un serveur permettant une bonne concordance entre projets). Cette étapes pourrait bénéficier d'un traitement plus rigoureux en utilisant les tailles d'allèles relatives (Amos *et al.*, 2007) afin de tenir compte de la spécificité de chaque locus ou en effectuant des modifications du jeu de données à l'issu de cette étape pour limiter l'impact des allèles non détectés ou des allèles rares (Bacles & Ennos, 2008).

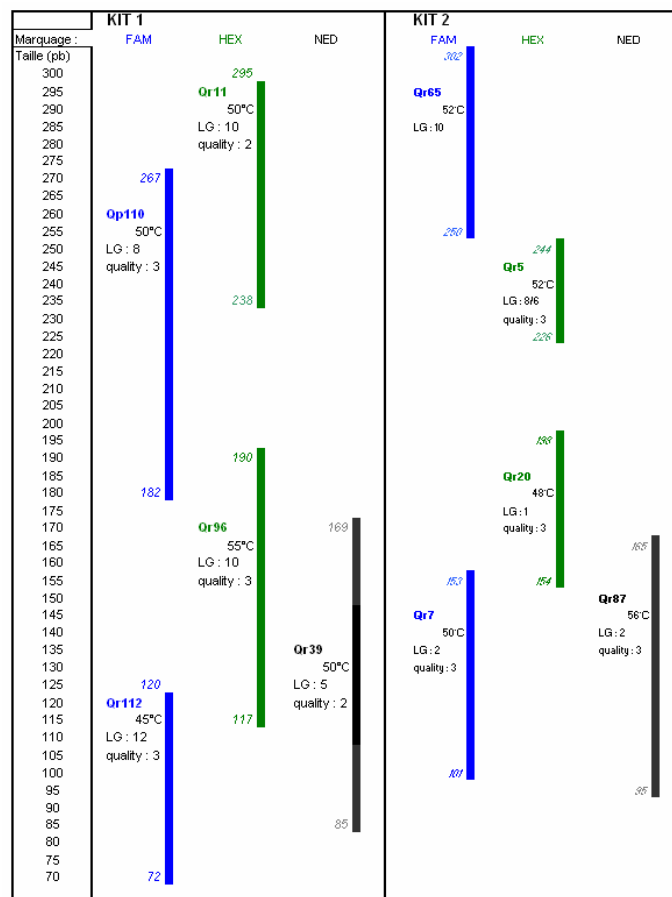


Figure 12 : Schéma des deux ensembles de microsatellites amplifiés en multiplexes (Kit1 à gauche et Kit2 à droite). Chaque locus est représenté par un rectangle dont la longueur correspond à l'étendue des tailles d'allèles possibles et la couleur correspond au fluorochrome utilisé pour marquer l'amorce forward.



Le séquenceur capillaire permet de faire co-migrer plusieurs locus à la fois, il ne restait plus qu'à mettre au point un protocole d'amplification PCR permettant de co-amplifier plusieurs locus à la fois (PCR multiplexe) pour permettre un débit d'analyse fluide. Un premier protocole incluant 10 microsatellites amplifiés en deux PCR mis au point début 2005, a été utilisé par Jessica Lavabre (Lavabre, 2005) pour génotyper 220 chênes du bois de la Paguères (voir chapitre 4). Après ces premiers résultats encourageants, nous avons rajouté 5 autres locus (15 microsatellites en deux PCR) pour génotyper 1054 chênes (adultes et plantules) du bois Briouant lors d'un séjour de 10 jours (13-24 juin 2005) au laboratoire de génomique fonctionnelle et évolutive de l'Université d'Uppsala (Suède) sur invitation de Martin Lascoux (expériences financées par le projet OAKFLOW). Cependant, confronté à des difficultés techniques, l'amplification des microsatellites n'a été que partielle causant un taux de données manquantes important (environ 25% de données manquantes pour 10 microsatellites) et l'amplification de 5 microsatellites a échoué pour tous les individus. Malgré cet échec relatif, ce séjour a permis de montrer que la technique de multiplexe permet de génotyper très rapidement un grand nombre d'individus pour un nombre de locus élevé. Un génotypage équivalent effectué par les techniques classiques (amplification d'un microsatellite à la fois et migration sur séquenceur à plaque de type Licor) nécessiterait plusieurs mois de travail.

A la suite de cette expérience, nous avons sélectionné les 10 locus qui présentaient un taux d'amplification correct (figure 3) et nous avons simplifié et optimisé le protocole d'amplification PCR au début de l'année 2006. La technique de génotypage rapide a fait l'objet d'une publication détaillant la méthode d'extraction d'ADN et l'amplification des microsatellites par PCR multiplexe (voir ANNEXE B, Lepais *et al.*, 2006b). Cette technique a permis de génotyper des individus à partir de différents tissus (feuilles, bourgeons et morceaux de glands) et s'est révélée également efficace pour génotyper des espèces mexicaines appartenant à la même section des chênes blancs (*Lepidobalanus*). Cependant, chez des espèces phylogénétiquement plus éloignées (par exemple les chênes rouges *Erythoballanus* ou les chênes méditerranéens tels que la section *Ilex* ou *Suber*), seulement une partie des microsatellites sont amplifiées.

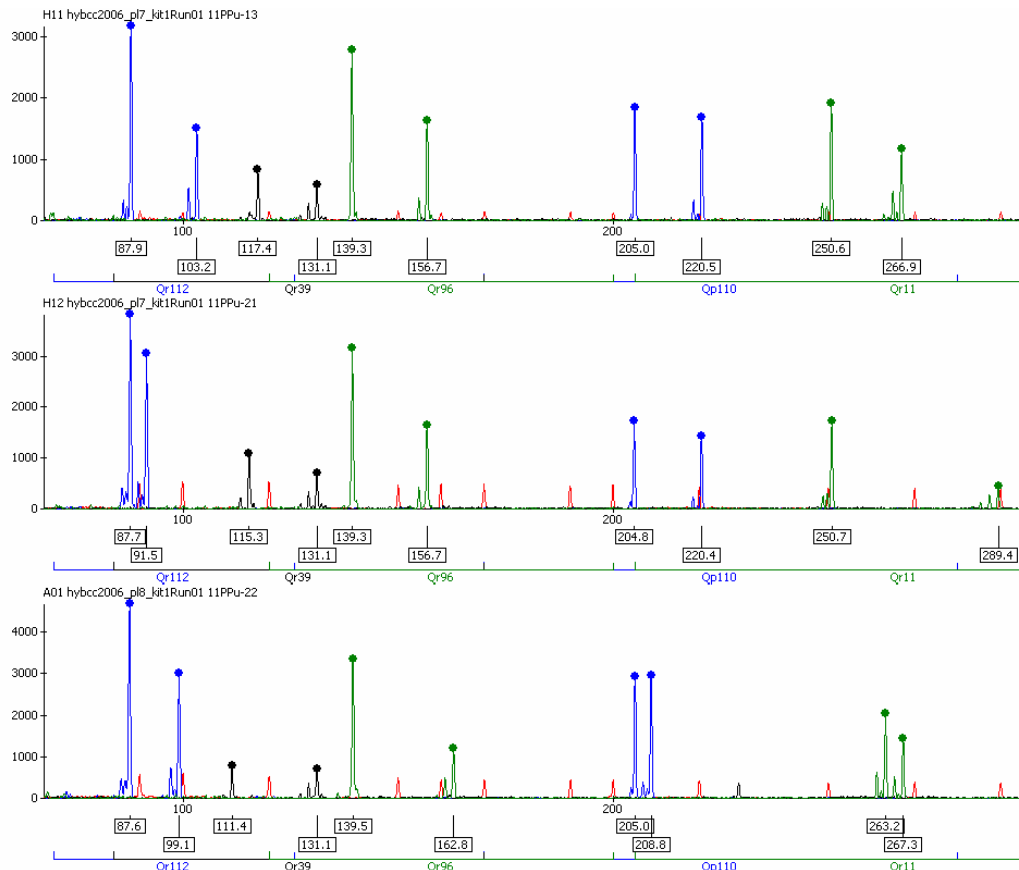


Figure 13: Exemple de profils obtenus dans le logiciel *Fragment Profiler* à l'issu de l'analyse par le séquenceur capillaire Amersham MegaBACE 1000. La taille des allèles (points de couleur) des 5 microsatellites du Kit1 est montrée pour trois individus. La taille de chaque allèle est obtenue grâce à la co-migration d'un marqueur de taille (pics rouges). La couleur des différents pics correspond à différents fluorochromes utilisés pour distinguer des locus ayant des tailles d'allèles proches (voir Figure 12).

### 2.3. Normalisation des données entre projets

Un avantage non négligeable des séquenceurs capillaires est que les données brutes issues de plusieurs projets peuvent être analysées ensemble avec les mêmes paramètres de lecture dans le logiciel *Fragment Profiler*, permettant une bonne concordance dans la façon dont sont effectuées l'identification des allèles, ce qui n'était pas possible précédemment. Par exemple, bien qu'analysés avec le même modèle de séquenceur à plaque Licor, les laboratoires impliqués dans le projet européen OAKFLOW n'ont pas réussi à faire correspondre la détermination des tailles d'allèles, rendant une analyse globale impossible. L'utilisation de modèles de séquenceur différents (Delmotte *et al.*, 2001, Baric *et al.*, 2008), les conditions de migrations (par exemple une différence de 5°C pour la température ambiante du laboratoire peut changer l'estimation de la taille des allèles jusqu'à près de 1pb (Davison & Chiba, 2003)) ou la méthode de lecture des données brutes peuvent rendre particulièrement difficile la correspondance de la dénomination des allèles entre laboratoires, entre projets au sein du même laboratoire voire même au cours du temps au sein du même projet. L'inclusion d'individus témoins dans chaque expérience (chaque plaque d'ADN contenant quelques

individus témoins) peut faciliter la résolution des décalages de tailles liées à la grande sensibilité des séquenceurs capillaires et permettre une étude globale des génotypes obtenus par différents laboratoires ou projets. Dans notre cas, le protocole de génotypage appliqué en Suède et à Pierroton a montré qu'une lecture globale (par le même utilisateur) permettait d'obtenir des annotations d'allèles comparables entre les deux laboratoires. Le génotypage grâce au protocole multiplexe Kit1 et Kit2 par plusieurs expérimentateurs dans plusieurs projets (différentes analyses indépendantes de données brutes) a permis d'obtenir des données génotypique présentant parfois quelques différences dans la notation des allèles (des décalages de plus ou moins 1 pb) mais ces différences se sont montrées rares, flagrantes et faciles à corriger une fois les différents jeux de données mis en commun (Figure 14). Une première étape de vérification et de normalisation entre jeux de données est absolument nécessaire au préalable de toute analyse en commun de génotypes obtenus indépendamment sinon de gros biais (en général flagrants, voir Figure 14) sont introduits dans les analyses ultérieures (Delmotte et al., 2001 ; Pasqualotto et al., 2007).

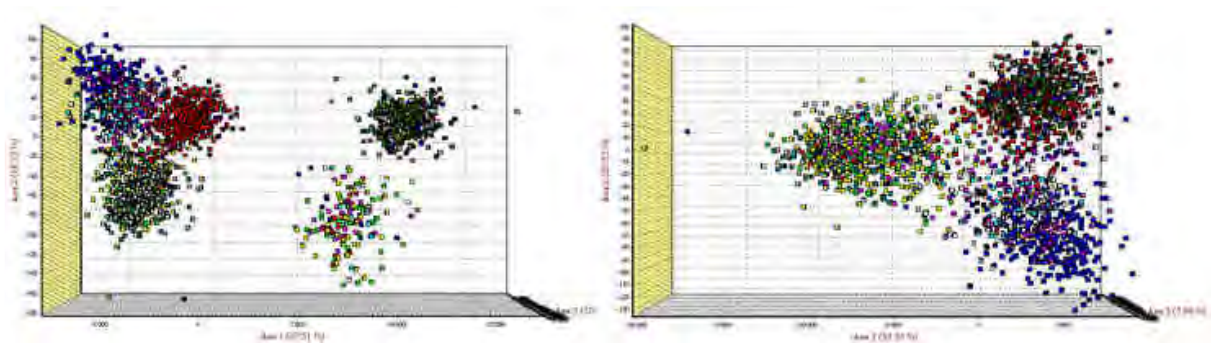


Figure 14 : Résultat d'une Analyse Factorielle des Correspondances (option "AFC 3D sur pop" dans le logiciel Genetix (Belkir *et al.*, 1996-2004)) sur les données génotypiques avant normalisation (A). L'ajout de nouveaux génotypes (populations ONF analysées par Erwan Guichoux, voir chapitre 4) provenant d'un autre projet montre que ces populations se distinguent particulièrement des autres dans l'espace factoriel (les deux nuages de points à droite). Une analyse des fréquences alléliques du jeu de données complet montre que trois locus sont annotés avec une différence systématique de -1 paire de base, créant de nouveaux allèles fictifs pour ces trois locus et ce qui différencie artificiellement ces populations des autres. La normalisation a consisté à ajouter 1 à la dénomination des allèles de ces nouveaux génotypes. Une nouvelle AFC à l'issue de cette normalisation (B) montre que le biais disparaît : les nouvelles populations se regroupent avec les autres et chaque nuage de point correspond à une espèce de chêne.

## 2.4. Calcul du taux d'erreur de génotypage

### 2.4.1. De la prise de conscience de l'erreur en génétique des populations

Une erreur de génotypage intervient lorsque le génotype observé pour un individu ne correspond pas à son vrai génotype. Ce problème est longtemps resté un sujet tabou malgré l'existence systématique des erreurs de génotypage, quel que soit le type de marqueur génétique analysé. Le développement récent des analyses génétiques à partir de matériel biologique dégradé (échantillonnage dit non-invasif de fèces ou poils d'animaux (Broquet & Petit, 2004 ; Buchan *et al.*, 2005), échantillonnage d'ADN ancien ou issu de collections dans les muséums (Cooper & Poinar, 2000 ; Paabo *et al.*, 2004 ; Wandeler *et al.*,

2007) nécessite l'utilisation de nombreuses répétitions des analyses afin d'estimer le taux d'erreur associé à l'utilisation d'ADN de mauvaise qualité. De plus, plusieurs répétitions du génotypage du même individu sont parfois nécessaires afin de déterminer un génotype consensus. La même nécessité de contrôler le taux d'erreur de génotypage est apparue également essentiel à l'application des outils de la génétique en police scientifique (Lin *et al.*, 2006 ; Gilder *et al.*, 2007). En fait, même dans les analyses plus classiques réalisées avec de l'ADN de bonne qualité, les erreurs de génotypage existent inévitablement (Soulsbury *et al.*, 2007). Rares sont les études en génétique des populations qui reportent le taux d'erreur associé au génotypage, alors que cette estimation permettrait d'estimer la fiabilité des résultats. En effet, les erreurs de génotypage sont susceptibles de biaiser sérieusement certains types d'analyses et de fausser leurs interprétations. Il est donc important d'estimer le taux d'erreur de génotypage et de prendre en compte ces erreurs dans les analyses effectuées à partir des données génotypiques (Broquet & Petit, 2004 ; Bonin *et al.*, 2004 ; Hoffman & Amos, 2005).

#### 2.4.2. Estimation du taux d'erreur de génotypage

Plusieurs méthodes, conduisant à des estimations comparables (Hoffman & Amos, 2005), existent afin d'estimer le taux d'erreur de génotypage (Bonin *et al.*, 2004). La méthode la plus précise mais aussi la plus longue et coûteuse consiste à répéter à l'aveugle le génotypage de 5 à 10 % des individus. Une autre technique consiste à vérifier l'absence de déficit en hétérozygote (test de Hardy-Weinberg) pour chaque locus, un excès d'homozygote à certains locus pouvant être le signe de présence d'allèles nuls ou de non amplification d'un des deux allèles (Van Oosterhout *et al.*, 2004 ; Chapuis & Estoup, 2007 ; Johnson & Haydon, 2007). Une méthode alternative consiste à comparer le génotype d'une mère avec ceux de sa descendance, cette approche est celle qui donne l'estimation du taux d'erreur la plus faible (Hoffman & Amos, 2005). C'est la méthode que nous avons utilisé afin d'estimer le taux d'erreur associé au génotypage de nos 10 microsatellites en multiplexe chez les chênes. Nous disposions du génotype de 626 glands récoltés sur 16 chênes. Ainsi le taux d'erreur de génotypage a pu être estimé en comptant le nombre de locus incompatibles chez les descendants avec pour référence celui de leur mère ; un taux d'erreur est calculé par descendance et par locus.

Le taux d'erreur moyen pour l'ensemble des locus et des descendance est de 2.1% (Tableau 2). Ce taux d'erreur varie entre familles de 0.2% à 5.0% et entre locus de 0.2% à 7.2%. Ces variations correspondent à des erreurs concentrées sur certains locus chez certaines familles comme le montre le pourcentage de discordance élevé de certains locus, principalement Qr39, Qr5 et Qr65. Lorsque les trois locus présentant plus de 5% d'erreur sont retirés, l'erreur globale passe de 2.1% à 0.85%. En d'autres termes, pour les 10 microsatellites (2.1% d'erreur) nous estimons une erreur à un locus environ tous les 50 génotypes mono-locus soit une erreur à un locus chez un individu sur cinq (20% des individus, ou des génotypes multilocus, contiennent une erreur). Pour 7 microsatellites (0.85% d'erreur), nous obtenons une erreur à un locus environ tous les 120 génotypes mono-locus soit une erreur à un locus chez un individu sur dix-sept (6% des individus, ou des génotypes multilocus,

contiennent une erreur). Les taux d'erreur estimés ici sont équivalents à ce que l'on peut trouver dans la littérature (Ewen *et al.*, 2000 ; Bonin *et al.*, 2004, en général moins de 2 à 3% d'erreur est jugé assez faibles (DeWoody *et al.*, 2006) pour ne pas biaiser les analyses (mais voir plus loin pour plus de détails).

La nature des erreurs liées aux microsatellites peut être multiple : erreur d'identification des allèles (lié à des profils d'électrophorèse difficile à interpréter), artefact pris pour un allèle, allèle nul (lié à la mutation de la séquence flanquante ce qui empêche une amorce de s'hybrider et l'amplification de l'allèle correspondant), *allele dropout* (amplification préférentielle d'un des deux allèles à cause de la compétition lors de la réaction d'amplification PCR) et parfois allèle incohérent (sans raisons identifiables). Dans notre cas, nous avons identifié plusieurs de ces types d'erreurs parmi les descendances (Tableau 2). Les erreurs ne semblent pas réparties de façon homogène entre locus et entre mères mais elles sont au contraire fortement élevées dans certaines descendances à un locus particulier. Ce résultat pourrait être lié à une cause biologique comme la présence de mutations dans la séquence flanquante des amorces qui pourrait perturber l'amplification correcte de certains allèles.

Tableau 2 : Taux d'erreur par locus et descendance pour l'ensemble des 10ssr ou des 7ssr présentant un faible taux d'erreur (exclus : Qr39, Qr5 et Qr65).

desc	n	Qp110	Qr11	Qr112	Qr39	Qr96	Qr20	Qr5	Qr65	Qr7	Qr87	10ssr	7ssr
316	31		4.80%		5.90%	14.3%(F)	11.10%				4.50%	3.9%	<b>4.5%</b>
210	23								7.10%	5.90%		1.3%	0.9%
367	33		3.70%				5.30%				3.70%	1.2%	1.6%
192	76							3.10%		1.50%	1.40%	0.6%	0.4%
756	57				<b>42.8%(F)</b>				11.9%(F)			5.0%	0.0%
698	56		2.00%						16%(D)		1.80%	0.3%	0.0%
706	48							<b>30.4%(F)</b>				3.0%	0.0%
85	48									2.50%		0.2%	0.4%
741	48				2.20%			12.5%(N)		2.30%		2.0%	0.3%
101	37			3.60%	<b>30.8%(F)</b>				nd			3.5%	0.5%
834	59		2.00%		2.10%			<b>34.8%(F)</b>	4.50%			4.1%	0.3%
39	62										1.90%	0.2%	0.3%
634	38	2.90%	3.00%			8.80%		2.90%	9.10%	9.10%		3.7%	<b>3.7%</b>
594	53		3.40%					11.1%(F)	nd			1.6%	0.3%
461	30	3.80%			8.00%				nd			1.2%	0.5%
377	28								<b>33.3%(AD)</b>			1.9%	0.0%
Moyenne	719	0.30%	1.00%	0.20%	<b>5.40%</b>	0.90%	0.40%	<b>7.20%</b>	<b>5.30%</b>	1.10%	0.60%	2.10%	0.86%

Le taux d'erreur est estimé par le pourcentage de discordance entre la mère et ses descendants (les taux d'erreurs particulièrement élevés sont indiqués en gras). Lorsque l'erreur est fréquente dans une descendance, sa cause est expliquée : AD = allele dropout, D = décalage de plus ou moins deux paires de bases, F = faux génotype incohérent avec la mère, N = amplification non spécifique (artefact).n : nombre de demi-frères (glands récoltés sur un chêne) par descendance, les cases vides correspondent à l'absence d'erreur détectée.

### 2.4.3. Conséquences des erreurs dans les analyses génétiques

Les erreurs de génotypage peuvent conduire à une mauvaise interprétation des données et à des conclusions biologiques erronées. Cependant la robustesse des analyses à l'erreur de génotypage varie en fonction de l'hypothèse testée et du type d'analyse pratiqué (DeWoody *et al.*, 2006). L'estimation de paramètres démographiques à partir de données génétiques (estimation de la taille ou de la dynamique démographique des populations), la cartographie génétique et les études d'association entre génotypes et phénotypes sont particulièrement sensibles aux erreurs de génotypage (Pompanon *et al.*, 2005). Les analyses basées sur les fréquences alléliques, comme par exemple l'estimation de la différenciation génétique en populations, sont sensibles aux erreurs (en particulier lorsqu'un des deux allèles n'est pas détecté) qui conduisent à une surestimation de la différenciation et à un déficit d'hétérozygote (DeWoody *et al.*, 2006) et des méthodes existent pour corriger les fréquences alléliques en tenant compte des erreurs (Chapuis & Estoup, 2007). Nous verrons au chapitre 5 quelques conséquences des erreurs de génotypage sur l'estimation de tels paramètres génétiques en population.

Pour ce qui concerne le type d'analyses qui va être utilisé dans cette thèse, i. e. l'analyse de parenté / paternité et l'assignation génétique, les conséquences des erreurs de génotypage sont différentes.

Pour les analyses de parenté ou de paternité, une erreur de génotypage à un locus chez un individu va entraîner une discordance entre cet individu et ses parents et donc conduire à exclure par erreur un ou deux vrais parents (Dakin & Avise, 2004). Les erreurs de génotypage vont donc réduire l'efficacité d'identification parent - enfant en excluant par erreur des relations de parenté, ce qui peut conduire à des conclusions erronées si les descendant sans parent assigné sont interprétés en terme d'immigrants (Hoffman & Amos, 2005 ; Pompanon *et al.*, 2005). Cependant, la prise en compte des erreurs, en autorisant quelques discordances entre allèles du descendant et du parent, permet de réduire le taux de fausse exclusion de paternité / parenté (Slavov *et al.*, 2005 ; Vandeputte *et al.*, 2006).

Dans notre cas, lorsque les trois locus présentant plus de 5% d'erreur sont retirés (Tableau 2), le taux d'erreur global pour les 7 microsatellites tombe à 0.7%, réduisant fortement les incompatibilités alléliques liées aux erreurs entre les parents potentiels (mère et père) et leurs descendants. Les trois locus présentant des taux d'erreurs supérieurs à 5% dans la descendance seront donc exclus pour les analyses de paternité / parenté et de modèles de reproduction corrélée (analyses détaillées dans le chapitre 7). Le fait d'exclure 3 microsatellites ne compromet pas la puissance du jeu de données puisque la probabilité d'exclusion avec 7 microsatellites est de 0.999 pour un test de paternité, ce qui reste une valeur élevée (voir chapitre 7).

Si de nombreuses études et simulations ont été effectuées pour estimer l'impact des erreurs de génotypage sur la recherche de parenté (Dakin & Avise, 2004), très peu d'analyses ont été menées pour quantifier leurs impacts sur les analyses d'assignation génétiques. La présence d'allèles nuls ou d'*allele dropout* n'aura pas autant d'effet pour l'analyse d'assignation génétique en comparaison à l'analyse de paternité. Ce type d'erreur va augmenter le nombre d'individus homozygotes (non

amplification de l'un des deux allèles) et donc rendre l'assignation génétique plus conservative c'est à dire réduire la détection d'hybride (Chapuis M. P., communication personnelle). D'autre part, l'effet de l'erreur sur un locus va être contrebalancé par l'effet des ces autres locus pour l'assignation génétique d'un individu. En effet c'est le génotype à tous les locus d'un individu qui va déterminer le résultat de son assignation génétique et une erreur n'aura que peu d'effet si elle concerne un locus ayant peu d'influence sur l'assignation (c. a. d. faible différenciation génétique, voir chapitre 3). Les erreurs sont particulièrement concentrées sur des locus ayant un faible taux de différenciation (voir Tableau 1 et Tableau 2) et donc vont avoir peu d'impact sur les résultats d'assignation génétique. De plus, les analyses développées au chapitre 3 montrent que le nombre de locus utilisé pour l'assignation génétique doit être maximal pour fournir le meilleur pouvoir résolutif. L'ensemble des 10 microsatellites sera donc conservé pour effectuer les analyses d'assignation génétique multi-population développées au chapitre 4. Toutes les erreurs détectées ont été codées comme données manquantes, réduisant le taux d'erreur global pour les analyses ultérieures.

## 2.5. Conclusion

Nous avons détaillé dans ce chapitre les différentes étapes qui ont permis de développer une technique de génotypage rapide tirant parti des avantages conférés par l'utilisation d'un séquenceur à capillaires. Malgré un nombre important de microsatellites testés, le jeu final de 10 microsatellites comprend trois locus affecté par des erreurs de génotypage que nous avons pu quantifier grâce à la comparaison des génotypes entre mères et descendants. Nous pouvons utiliser nos données génotypiques en fonction du type d'analyse appliqué, en particulier, en sélectionnant les 7 locus les plus robustes pour les analyses sensibles au biais lié à ces erreurs. D'une façon générale, il est rare de disposer de descendance pour vérifier la ségrégation des marqueurs et quantifier l'erreur lors des analyses en populations, il serait alors nécessaire de procéder à une ré-analyse à l'aveugle d'environ 10% des individus afin d'estimer le taux d'erreur de toute expérience de génotypage. Outre la rapidité de génotypage, l'utilisation de séquenceur capillaire permet d'automatiser certaine tâches, réduisant les sources d'erreurs (Vemireddy *et al.*, 2007) et facilitant l'analyse conjointe de données issues de plusieurs projets voire de plusieurs laboratoires. Cependant, les résultats de telles analyses semblent hétérogènes entre modèles de séquenceurs (Baric *et al.*, 2008). Le remplacement tout récent du séquenceur MegaBACE 1000 par un modèle de la marque ABI a permis de comparer les données issues des deux machines. Il est apparu que la taille des allèles détectés était assez différente d'une machine à l'autre, rendant la combinaison des données difficile (Séverine Mary, communication personnelle). La normalisation des données risque de plus d'être compliquée par le fait que la nature des fichiers générés par les deux machines ne sont pas compatibles et ne pourront pas être analysés avec le même logiciel. Etant donné que les nombreux génotypes générés par le séquenceur MegaBACE pourraient servir de référence, afin d'assigner un individu à une espèce par exemple, il serait utile de pouvoir inclure d'autres individus à ces génotypes de référence, ce qui nécessite de pouvoir effectuer au préalable une normalisation

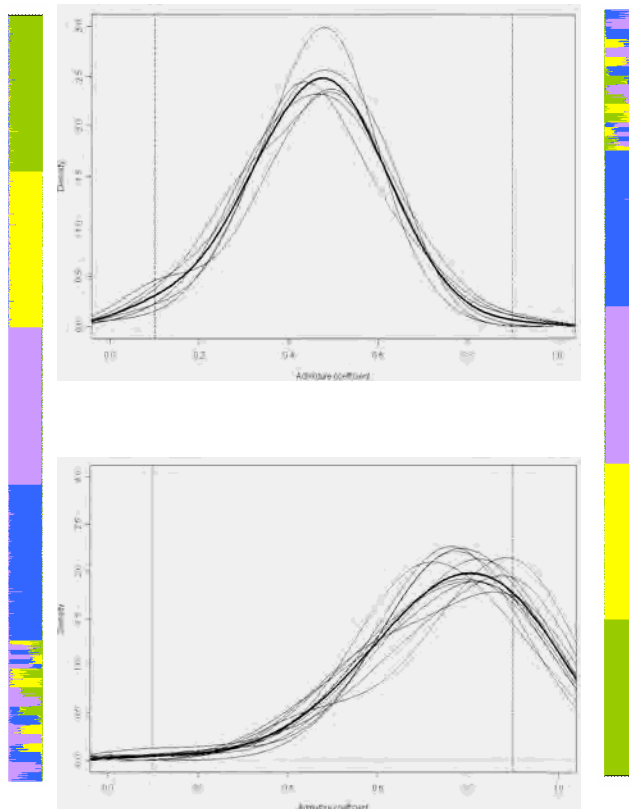
appropriée (par re-génotypage de quelques individus de référence avec le nouveau séquenceur par exemple). Nous attirons ici l'attention sur le fait de ne pas sous-estimer le biais important créé sans normalisation appropriée.

Plus généralement et en perspective de futures analyses de génotypage, l'utilisation du séquenceur capillaire permet d'accéder à la taille des fragments exprimée en paires de bases. Un gain important dans la qualité d'identification des allèles pourrait être fait à l'avenir en utilisation d'autres outils que les logiciels de lecture standards livrés avec les séquenceurs (par exemple le logiciel Flexibin, Amos *et al.*, 2007). D'autres méthodes récemment développées permettent de limiter l'effet des erreurs dans les analyses en réduisant le nombre d'allèles rares ou en tenant compte de l'éventuel existence d'un allèle nul en créant un "allèle zéro" (Amos *et al.*, 2007 ; Bacles & Ennos, 2008). Les erreurs sont ici prises en compte à la source, dans la façon dont on code les allèles. Le grand avantage est que l'on peut comparer les résultats obtenus avec les différentes codifications et ainsi comprendre l'origine éventuelle des erreurs pour chaque locus et trouver le format de donnée limitant l'effet des erreurs. Ces approches récentes montrent la prise de conscience de l'erreur en génétique des populations et fournissent des outils au généticien lui permettant d'apprécier, de tester et de confirmer l'adéquation entre la précision de ses données et l'interprétation de ses analyses.





# Chapitre 3- Tests d'assignation génétique des espèces et des hybrides



Page précédente : Les simulations permettent de tester les méthodes d'assignations génétiques. Ici, le résultat d'assignations génétiques d'individus hybrides F1 et rétrocroisements.

### 3.1. Introduction

Les espèces de chênes européens, bien que morphologiquement distinguables, peuvent être parfois difficile à identifier particulièrement lorsque plusieurs espèces sont en mélanges dans le même peuplement. Les études antérieures basées sur la morphologie foliaire ont montré que certains caractères différenciant les chênes sessiles et pédonculés sont particulièrement efficaces pour la distinction des espèces (Bacilieri *et al.*, 1996a) et que cette différence de morphologie est stable sur toute l'aire de répartition (Kremer *et al.*, 2002a). De plus, certaines espèces ont des caractéristiques morphologiques propres comme par exemple la présence de poils sous le limbe chez le chêne pubescent (Dupouey & Badeau, 1993 ; Bruschi *et al.*, 2000) ou la présence de trichome sur le limbe de chêne tauzin. Néanmoins, ces analyses indiquent que ces espèces ont une grande variabilité de morphologie et forment souvent un continuum pouvant rendre la distinction incertaine pour des individus à la morphologie intermédiaire ou combinée. Deux hypothèses ont été invoquées pour rendre compte de ce continuum morphologique : soit une variabilité morphologique intra-spécifique forte rend difficile la séparation des espèces avec les variables mesurées, et / ou la présence d'hybrides possédant des morphologies intermédiaires rend la distinction entre espèces difficile.

Ce constat démontre qu'il est impossible d'identifier des hybrides entre espèces de ce complexe par la seule morphologie. Afin de pouvoir étudier l'hybridation entre espèces appartenant au complexe des chênes blancs européen, il est indispensable d'utiliser d'autres méthodes, telles que les marqueurs génétiques, afin d'identifier précisément les espèces et les hybrides présents dans les populations.

Le chêne est une espèce d'importance majeure dans les forêts de France et d'Europe et représente un produit commercial non négligeable, rendant nécessaire une certification de l'espèce ou de l'origine du matériel, comme les semences, les plants ou le bois. Ces besoins de traçabilité nécessitent des analyses directes indépendantes de la nature du produit à tester. Bien que des réflexions soient lancées pour le développement de tels tests, les analyses d'ADN permettant la traçabilité en sont encore au stade de développement au laboratoire.

Le développement des marqueurs génétiques très polymorphes (microsatellites) ainsi que des méthodes d'assignations génétiques performantes permettent d'envisager l'identification des espèces grâce aux données génétiques, indépendamment des données morphologiques. Néanmoins, bien que d'une grande sensibilité, ces méthodes nécessitent une évaluation approfondie avant d'être appliquées en populations naturelles (Randi, 2008). L'objectif de ce chapitre, composé de quatre parties, est d'explicitier l'utilisation des marqueurs génétiques pour l'identification des espèces et des hybrides.

- La première partie comprend une courte synthèse des méthodes d'assignations génétiques existantes ainsi que leurs domaines d'application.

- Certaines de ces méthodes sont appliquées à deux espèces de chênes afin de comparer leurs résultats et d'examiner la concordance entre assignation génétique et détermination morphologique. Ceci fait l'objet de la deuxième partie du chapitre.

Des analyses par ré-échantillonnages et simulations permettent alors de tester la sensibilité de la méthode d'assignation génétique :

- Le nombre de marqueurs génétiques et d'individus nécessaires pour une assignation de bonne qualité est estimé par deux techniques de ré-échantillonnage dans la troisième partie.
- Enfin la dernière partie est consacrée à des simulations d'individus d'hybrides dans le but de tester la capacité de l'assignation génétique à identifier les différents types d'hybrides.

## 3.2. Revue des principales méthodes d'assignation

### 3.2.1. Introduction

Le terme d'assignation génétique est défini comme l'affectation d'un individu à un groupe à partir de l'analyse de données génétiques. Les premières techniques permettant une telle classification des individus utilisaient des analyses multivariées telles que l'analyse factorielle des correspondances afin de représenter les individus dans un espace factoriel. Des groupes d'individus proches dans l'espace peuvent être ainsi visualisés et si ces groupes correspondent à des espèces, les individus placés entre ces groupes peuvent être assimilés à des hybrides. Néanmoins, ces méthodes sont limitées car les groupes constitués ne correspondent pas toujours à une espèce et ils peuvent dépendre fortement du traitement statistique choisi (nature de l'analyse multivariée ou du type de distance). Il est alors souvent difficile d'obtenir une assignation précise et fiable.

Depuis le milieu des années 1990, de nombreuses méthodes d'assignation basées sur des modèles de génétique des populations ont été développées. L'incorporation d'un modèle permet de tirer d'avantage d'information des données génétiques. Ces modèles prennent en compte la structure des données génétiques afin d'identifier des groupes homogènes. La présence de plusieurs populations (ou espèces) crée un signal se traduisant par un écart à l'équilibre de Hardy-Weinberg (Encadré 1) et un déséquilibre de liaison (Encadré 2) entre marqueurs génétiques. Ces méthodes utilisent ce signal pour assigner les individus aux populations en minimisant ces deux déséquilibres autant que possible.

On distingue deux type d'approches : la classification (ou test d'assignation supervisé) qui nécessite des populations de référence afin d'assigner les individus et le clustering (ou assignation non supervisé) qui utilise les données génétique seules, sans population de référence ou de connaissance à priori sur l'origine des individus (Manel *et al.*, 2005). Ces deux approches sont présentées dans les deux paragraphes ci-dessous.

Le nombre important de techniques d'assignation rend parfois difficile le choix de la méthode appropriée à ses données et à la question biologique à résoudre. J'explique dans cette partie la particularité de chaque méthode ainsi que leurs applications possibles (un état des lieux est proposé dans le tableau).

### Encadré 1 – Le principe de Hardy-Weinberg

Selon le principe de Hardy-Weinberg, les fréquences alléliques d'une population diploïde idéale restent stables d'une génération à l'autre et ne dépendent que des fréquences alléliques de la génération initiale. Les hypothèses de départ sont nombreuses quant à la population idéale (fictive). En effet, elle est de grande taille (idéalement infinie), la reproduction est sexuée, la rencontre des gamètes se fait au hasard (panmixie), les gamètes sont équiprobables (pangamie). Enfin, il n'y a ni sélection, ni mutation, ni migration et les générations ne se chevauchent pas.

Pour un locus a deux allèles  $A$  et  $a$  de fréquence  $p$  et  $q$  respectivement. Les trois génotypes possibles et leurs fréquences sont :

$$\begin{aligned}f(AA) &= p^2 \\f(Aa) &= 2pq \\f(aa) &= q^2\end{aligned}$$

La somme des fréquences des trois génotypes étant de 1 :

$$p^2 + 2pq + q^2 = 1$$

A la génération suivante, la fréquence  $p'$  de l'allèle  $A$  est :

$$p' = f(AA) + \frac{1}{2} f(Aa) = p^2 + pq = p(p+q) = p$$

Dans cette population fictive, les fréquences alléliques restent constantes au cours des générations et correspondent aux fréquences alléliques initiales. Cette population n'évolue pas car les fréquences alléliques ne changent pas. Le processus évolutif a donc lieu si les propriétés de la population fictive ne se retrouvent pas comme c'est le cas très fréquemment dans les populations naturelles où la sélection, la migration et la mutation interviennent. Ainsi des déséquilibres entre fréquences génotypiques observées et attendues à partir des fréquences alléliques indiquent un écart à l'équilibre de Hardy-Weinberg.

### Encadré 2 – Le déséquilibre de liaison

La liaison génétique entre locus est définie comme l'association non aléatoire entre allèles à plusieurs locus. Cette association d'allèles peut être liée à la proximité physique des locus sur un même chromosome. Elle est d'autant plus forte que la recombinaison entre les locus est faible, c'est à dire que la distance entre locus est faible. Cette information est utilisée afin de positionner des locus sur les cartes génétiques.

Néanmoins, la liaison entre locus appartenant à des chromosomes différents est possible. Le déséquilibre de liaison apparaît lorsque quelques combinaisons d'allèles apparaissent plus ou moins fréquemment dans une population que ce qui est attendu sous association aléatoire des allèles basées sur leurs fréquences. Le déséquilibre de liaison va apparaître du fait des taux de recombinaison et de mutation, de la dérive, la reproduction non aléatoire et la structure de la population.

### 3.2.2. Classification – assignation supervisée

De nombreuses méthodes ont été développées dans le but de classer des individus parmi des populations de référence (Paetkau *et al.*, 1995 ; Rannala & Mountain, 1997 ; Paetkau *et al.*, 2004). La plupart de ces méthodes sont regroupées dans le logiciel GeneClass 2.0 (Piry *et al.*, 2004). Pour chaque population de référence, la probabilité d'observer chaque génotype est calculée à partir de leurs fréquences alléliques. Chaque individu d'origine inconnue est ensuite affecté à la population dans laquelle son génotype est le plus probable. Ce principe de base des tests d'assignation peut être complété par un test d'exclusion (Cornuet *et al.*, 1999) afin de vérifier que toutes les populations potentielles ont été échantillonnées parmi les populations de référence. En effet, le principal défaut des tests d'assignation est qu'un individu va être assigné à la population la plus probable parmi les populations de références ce qui nécessite (1) de pouvoir définir des populations de référence, (2) de

génotyper toutes les populations dont peuvent provenir les individus d'origine inconnue à assigner. Pour vérifier l'assignation des individus, chaque population est exclue comme origine de l'individu inconnu jusqu'à ce qu'il ne reste plus qu'une population d'origine possible pour l'individu (ce qui confirme la première assignation) ou, si toutes les populations sont exclues, alors l'individu appartient à une population non échantillonnée. Ces méthodes ont été très largement employées pour connaître la provenance d'un individu inconnu. La combinaison de test d'assignation et de test d'exclusion a par exemple permis de détecter un cas de fraude lors d'un concours de pêche organisé dans un lac en Finlande (Primmer *et al.*, 2000). L'analyse de sept marqueurs microsatellites du saumon "gagnant" a permis d'exclure le lac comme population d'origine de ce saumon. A la vue de ces résultats, le fraudeur a avoué avoir acheté cet individu chez le poissonnier.

### 3.2.3. Clustering – assignation non supervisée

Le clustering consiste à distinguer au sein des données des groupes homogènes représentant des populations différentes. Contrairement aux méthodes de classification, les analyses de clustering ne nécessitent pas de populations de référence, ni de connaître *a priori* le nombre de populations présent dans l'échantillonnage. Les populations sont identifiées grâce au déséquilibre de liaison induit par le mélange d'individus provenant de plusieurs populations. Des groupes d'individus sont créés de façon à réduire au minimum le déséquilibre de liaison intra-groupe, ces groupes peuvent alors être considérés comme autant de populations. Dans un deuxième temps, les individus sont assignés aux différents groupes. Ces méthodes de clustering sont particulièrement utiles lorsque l'on ne dispose pas de populations de référence, ou lorsqu'il est difficile d'identifier des populations *a priori*, comme c'est le cas chez les chênes blancs européens. Elles permettent de plus d'effectuer une assignation "en aveugle" sans autre *a priori* que les données génétiques (Duminil *et al.*, 2006). Depuis le premier article décrivant le fonctionnement du logiciel Structure en 2000 (Pritchard *et al.*, 2000) de nombreux logiciels apparaissent (Tableau 3), chacun implémentant des méthodes statistiques plus efficaces afin de réduire les temps de calcul (par exemple FASTSTRUCT implémente le modèle de base de Structure dans un environnement non bayésien (Chen *et al.*, 2006)) ou des modèles différents permettant de répondre à des problèmes biologiques plus spécifiques dont deux exemples sont détaillés ci-dessous.

### 3.2.4. Délimitation spatiale des populations

Différentes méthodes permettent d'inclure les coordonnées spatiales des individus en plus des génotypes. Cette démarche permet de déterminer conjointement le nombre de populations présentes dans les données, la localisation des discontinuités génétiques dans l'espace et l'assignation des individus à l'une des populations. Elle est particulièrement utile lorsque les populations présentent une faible différenciation génétique. Dans ces cas l'ajout d'une information supplémentaire (l'organisation spatiale des individus) peut aider à résoudre la structure des populations (Guillot *et al.*, 2005a). De nombreuses méthodes d'assignation spatialement explicites ont été développées ces dernières années (Tableau 3) en relation avec l'émergence de la génétique du paysage ("landscape genetic" : Manel *et*

*al.*, 2003 ; Holderegger & Wagner, 2006 ; Storfer *et al.*, 2007) , nouveau domaine de recherche intégrant la génétique des populations, l'écologie du paysage et les statistiques spatiales pour étudier le rôle des caractéristiques du paysage dans la diversité génétique et la structure des populations.

### 3.2.5. Détection des hybrides

NEWHYBRIDS a été développé afin d'assigner les individus aux différentes catégories d'hybrides entre deux espèces (Anderson & Thompson, 2002). Au lieu d'estimer la proportion du génome d'un individu provenant de chaque espèce (comme c'est le cas du modèle *Admixture* de Structure, voir Encadré 3 pour le principe de fonctionnement de Structure), le programme estime une probabilité d'appartenance à chaque classe hybride (F1, F2 et rétrocroisements) pour chaque individu.

### 3.2.6. Conclusion

Les méthodes d'assignations disponibles se différencient donc principalement par la nécessité de disposer ou non de populations de référence. Pour une application à des espèces difficilement distinguables par leur morphologie, comme les chênes, les méthodes d'assignation non supervisées paraissent donc être un choix judicieux. Dans la partie suivante, j'ai comparé différentes méthodes non supervisées dans le cas de deux espèces de chênes.



Tableau 3 : Quelques exemples de logiciels d'assignation génétique (adapté de Excoffier & Heckel, 2006 et Manel *et al.*, 2005).

Logiciel	Description	Spécificité	Inférence <sup>1</sup>	Questions <sup>2</sup>	Référence
GENECLASS 2	Détection de migrants de première génération. Test d'assignation à partir de populations de référence.	Inclus de nombreuses méthodes. Test d'exclusion d'individu d'une population donnée.	Fréq., Max., Bayes.	Orig., Migr.	Piry <i>et al.</i> , 2004
Structure	Permet de tester le nombre de populations présentes dans un jeu de données multilocus. Assignation des individus à une ou plusieurs populations.	Le nombre de population est fixe dans le modèle mais il peut être déterminé en testant différentes valeurs. Comprend de nombreux modèles dont le modèle <i>Admixture</i> permettant d'identifier des individus hybrides ainsi que le modèle <i>Linkage</i> pouvant tirer partie de locus liés physiquement.	Bayes.	Orig., Migr., Délim., Hyb.	Pritchard <i>et al.</i> , 2000
FASTRUCT	Identique à Structure mais assigne les individus à une seule population.	Modèle de base de Structure ( <i>No Admixture</i> ) implémentation non bayésienne pour plus de rapidité de calcul.	Max.	Orig., Migr., Délim.	Chen <i>et al.</i> , 2006
STRUCTURAMA	Détermination du nombre de populations le plus vraisemblable. Assignation des individus à une population.	Le nombre de populations est intégré comme variable aléatoire dans le modèle permettant une détermination directe. Modèle <i>No Admixture</i> de Structure.	Bayes.	Orig., Migr., Délim.	Huelsenbeck & Andolfatto, 2007
BAPS	Détermination directe du nombre de populations par le modèle et assignation des individus à une ou plusieurs populations (modèle <i>No Admixture</i> ou <i>Admixture</i> ). Assignation de groupes d'individus.	Les individus particuliers peuvent être isolés dans plusieurs clusters spécifiques. Comprend de nombreuses versions implémentant différents modèles, par exemple prise en compte de la liaison physique entre marqueurs ou des coordonnées spatiales des individus.	Bayes.	Orig., Migr., Délim., Hyb., Stock.	Corander <i>et al.</i> , 2004
NEWHYBRIDS	Assignation des individus à différentes classes d'hybrides entre deux espèces.	Les hybrides peuvent être détectés jusqu'à n générations (n fixé par l'utilisateur), nécessite beaucoup de marqueurs pour n>1.	Bayes.	Hyb.	Anderson & Thompson, 2002
GENELAND	Détection de la structure des populations à partir de génotypes individuels géo référencés.	Détermine le nombre de populations directement. Intègre une incertitude sur les coordonnées spatiales. Package R avec une interface graphique dans la dernière version.	Bayes.	Orig., Migr., Délim.	Guillot <i>et al.</i> , 2005b
TESS	Identique à GENELAND mais inclus un modèle spatial différent.	Inclus un modèle <i>Admixture</i> .	Bayes.	Orig., Migr., Délim., (Hyb.)	Chen <i>et al.</i> , 2007

<sup>1</sup> Fréq. : fréquentiste, Max. : maximum de vraisemblance, Bayes. : Bayésienne

<sup>2</sup> Orig. : origine des individus, Migr. : migration, Délim. : délimitation des populations, Hyb. : hybridation, Stock. : analyse de stock génétique

### 3.3. Comparaison de méthodes d'assignation

#### 3.3.1. Introduction

L'analyse de données génétiques par plusieurs méthodes conduit parfois à des résultats différents voire contradictoires lorsque les données contiennent peu d'information. Les hypothèses sous-jacentes des méthodes d'analyses ont alors davantage de poids que les données elles mêmes et il devient difficile de tirer des conclusions claires à partir des résultats obtenus. Le premier objectif de cette partie est de comparer les résultats d'assignation génétique obtenus par différentes méthodes dans le cas d'une population comprenant deux espèces de chênes. La comparaison concerne quelques méthodes d'assignation non supervisée ainsi que plusieurs modèles pour chacune de ces méthodes. J'ai vérifié la concordance générale des résultats obtenus par ces méthodes. Le second objectif de cette partie est de vérifier la congruence entre les groupes d'individus identifiés génétiquement et l'identification morphologique des espèces.

#### 3.3.2. Matériels et méthodes

##### 3.3.2.1. La Petite Charnie

La Petite Charnie est une parcelle de chênes de la Sarthe comprise dans le massif de La Charnie. Elle a été analysée pour 16 microsatellites (6 microsatellites analysés lors de la thèse de Réjane Streiff (Streiff, 1998 ; Streiff *et al.*, 1998), et 10 microsatellites supplémentaires analysés par Jessica Lavabre (Lepais *et al.*, 2006b)). Cette parcelle est un bon modèle pour tester l'efficacité des méthodes d'assignation car elle a été étudiée avec un nombre important de marqueurs génétiques, comprend seulement deux espèces en effectifs relativement équilibrés (chênes pédonculés  $n = 154$ , et chênes sessiles  $n = 112$ , et 5 individus à la morphologie intermédiaire, total : 271 individus) et chaque arbre a été géo localisé dans la parcelle. De plus, des mesures de morphologie foliaire sont disponibles pour tous ces arbres (Bacilieri *et al.*, 1995) permettant une analyse indépendante du statut spécifique des chênes.

##### 3.3.2.2. Méthodes d'assignation génétique comparées

- Structure

Le logiciel Structure comprend plusieurs modèles de populations dont les résultats s'interprètent différemment (voir l'Encadré 3 pour une description du fonctionnement du logiciel Structure). Deux modèles sont utilisés dans cette comparaison : le modèle *Admixture* et le modèle *Linkage*.

Le modèle *Admixture* prend en compte le fait que les individus peuvent être issus d'un mélange de plusieurs populations. Les individus sont caractérisés par la proportion de leur génome appartenant à chaque population. Ce modèle tient compte de l'existence d'éventuels hybrides et permet de les identifier.

Parmi les 10 marqueurs génétiques utilisés, 3 sont localisés sur le groupe de liaison 10 et 2 sont situés sur le groupe de liaison 2 (voir chapitre 2). Le déséquilibre de liaison entre marqueurs proches n'étant pas dû au mélange de populations, il risque de biaiser l'assignation génétique s'il n'est pas spécifiquement intégré dans le modèle (Falush *et al.*, 2003). J'ai ainsi utilisé le modèle *Linkage* qui prend en compte l'espacement entre marqueurs génétique et donc le déséquilibre de liaison causé par leur proximité physique. La distance entre les marqueurs microsatellites a été extraite à partir de la carte génétique disponible chez le chêne pédonculé (Barreneche *et al.*, 1998) et complétée depuis par d'autres marqueurs (Scotti-Saintagne *et al.*, 2004b). Dans le cas où la liaison physique des marqueurs provoquerait un biais d'assignation, on devrait constater des différences importantes dans les résultats générés par les modèles *Admixture* et le modèle *Linkage*.

- Méthodes spatialement explicites : Geneland et Genepatch

J'ai utilisé deux méthodes intégrant le positionnement géographique des individus dans la parcelle. L'idée ici était de tester si l'ajout de l'information géographique allait modifier les résultats des assignations, ce qui signifierait que les données génétiques contiennent relativement peu d'informations concernant l'espèce des individus (leur localisation spatiale ayant davantage de poids dans le résultat de l'analyse).

- Méthode spécialisée dans la détection d'hybride : NewHybrids

Ce logiciel permet d'identifier des individus appartenant à une des deux espèces pures ou à différentes catégories d'hybrides. Je me suis intéressé ici à la capacité de cette méthode à identifier les arbres appartenant à une espèce pure. J'ai comparé l'utilisation de deux paramètres concernant la distribution a priori des paramètres (Uniform Prior et Jeffrey's Prior) afin d'estimer si les résultats étaient sensibles au choix des paramètres de départ.

### 3.3.2.3. Comparaison des résultats

Tous les logiciels utilisés, excepté NewHybrids, présentent les résultats sous forme de tableau dont les lignes représentent les individus et deux colonnes représentant la proportion de chaque espèce dans le génome de l'individu (Structure) ou la probabilité d'appartenance de l'individu à chacune des deux espèces (Geneland et Genepatch). Ces deux colonnes somment à 1 pour chaque individu. J'ai extrait de ces résultats la probabilité d'appartenir à l'espèce 1 (chêne pédonculé), les valeurs proches de 1 correspondent à des individus de l'espèce pédonculé alors que des valeurs proches de 0 correspondent à des individus sessiles.

Les résultats de NewHybrids forment un tableau à 6 colonnes, correspondant à la probabilité de l'individu d'appartenir à chacune des deux espèces ainsi que chacune des classes d'hybrides (hybrides de première génération F1, et de seconde génération : F2, rétrocroisement avec l'une des deux espèces). Afin de comparer ces résultats avec les autres méthodes, j'ai extrait la colonne correspondant à la probabilité d'appartenance à l'espèce 1 (chêne pédonculé).

Ces différents résultats d'assignation sont représentés graphiquement deux à deux et un coefficient de corrélation est calculé pour chaque couple de résultats au niveau individuel.

#### 3.3.2.4. *Correspondance entre détermination morphologique et assignation génétique*

Afin de s'assurer que les deux groupes identifiés par l'assignation génétique correspondent bien aux deux espèces, j'ai comparé l'assignation génétique par le logiciel Structure (modèle *Admixture*) avec ceux issu de l'analyse de la morphologie des feuilles des chênes de la Petite Charnie. Afin d'identifier l'espèce de chaque individu, un indice de discrimination morphologique (Kremer *et al.*, 2002a) est calculé pour chaque individu :

$$ID = 357 - (97 - PL) + (205 \times NV)$$

où PL représente la longueur du pétiole en millimètre et NV correspond au nombre de veines secondaires. Cet indice prend une valeur négative pour les chênes sessiles et une valeur positive pour les chênes pédonculés (Kremer *et al.*, 2002a).

Pour l'assignation génétique, les arbres sont considérés comme purs si plus de 90% de leur génome provient d'un des deux clusters, les autres individus étant considérés comme hybrides (voir la partie 4.5 pour plus de détails sur les seuils d'assignation).

### 3.3.3. Résultats et discussions

#### 3.3.3.1. *Comparaison des méthodes d'assignation*

L'ensemble des méthodes d'assignation testées donnent des résultats très semblables avec des coefficients de corrélation supérieurs à 0.98 (excepté l'analyse avec NewHybrids avec *Jeffrey's Prior* dont les corrélations avec les autres méthodes vont de 0.92 à 0.96). La distribution des valeurs d'appartenance à l'espèce pédonculée (courbes bimodales localisées sur la diagonale de la Figure 15) montre que le chêne pédonculé est plus représenté dans la parcelle que le chêne sessile (les valeurs proches de 1 sont plus fréquentes que les valeurs proches de 0).

Les deux modèles de Structure testés (dont un tient compte de la liaison physique entre les marqueurs génétiques) donnent les mêmes résultats ce qui signifie que le déséquilibre de liaison causé par la proximité physique des marqueurs ne biaise pas l'assignation.

L'ajout de l'information spatiale des individus ne change pas l'assignation des individus à l'espèce, mais modifie légèrement la valeur d'assignation pour les individus intermédiaires (hybrides). En effet, ces méthodes ne sont pas adaptées pour détecter des hybrides (modèle *No Admixture*) et ces individus intermédiaires vont être assignés à l'espèce dont ils sont la plus proche géographiquement.

Enfin, l'assignation avec le logiciel NewHybrids (*Uniform Prior*) conduit à des assignations aux espèces identiques, bien que des différences substantielles existent en fonction du choix de la distribution *a priori* des paramètres (*Uniform Prior* versus *Jeffrey's Prior*). En utilisant une distribution des paramètres *a priori* correspondant à l'option *Jeffrey's Prior*, tous les individus sessiles sont identifiés comme des rétrocroisements alors que le paramètre *Uniform Prior* classe les individus

de façon identique à Structure. Ce résultat pourrait être lié à une plus forte sensibilité du modèle aux allèles rares lorsque les paramètres à priori suivent le principe de Jeffrey (Anderson, communication personnelle).

Malgré les différences conceptuelles importantes entre ces méthodes, les assignations sont fortement concordantes en particulier pour assigner les individus à l'une des deux espèces.

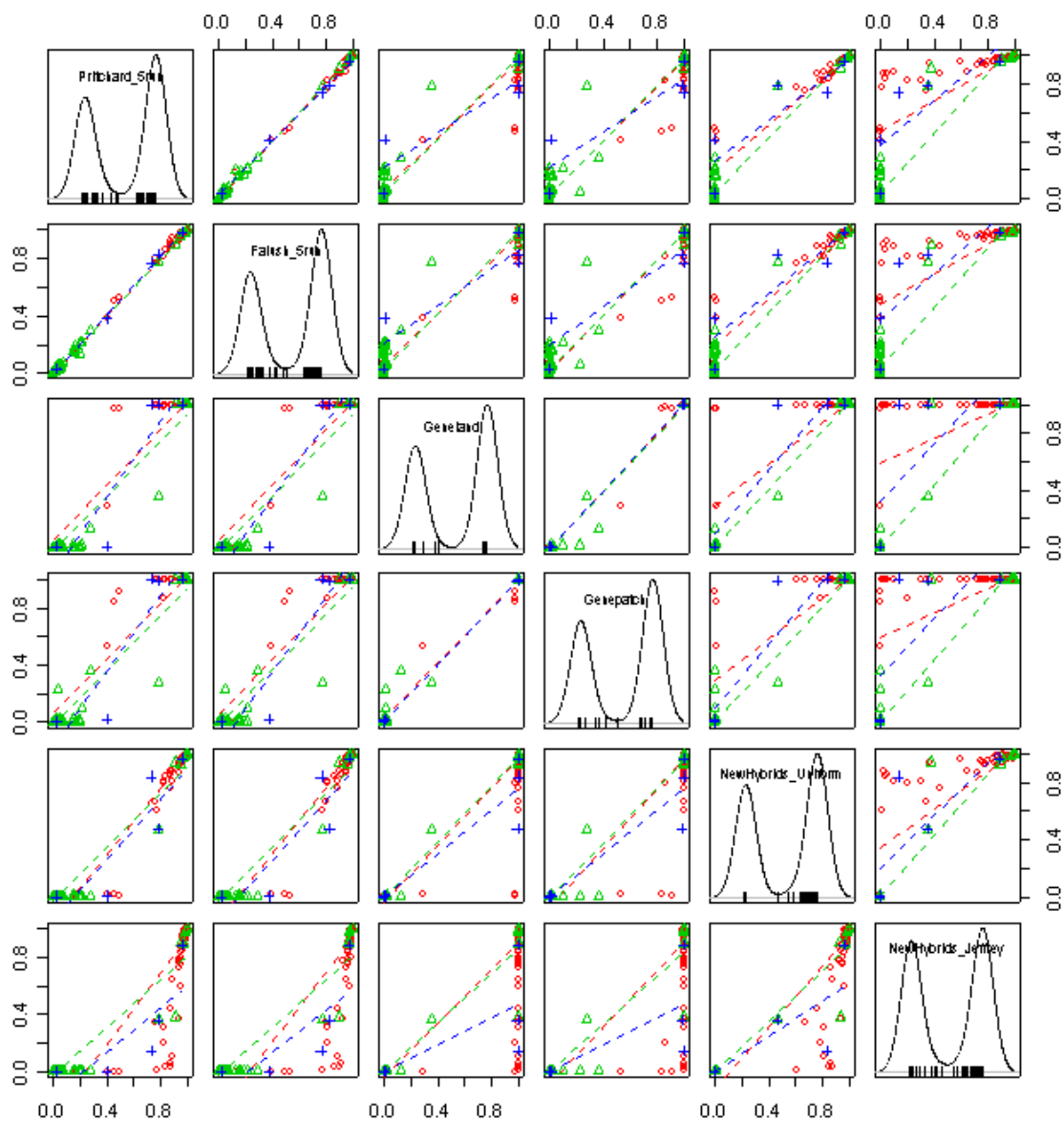


Figure 15 : Scatterplot montrant les corrélations deux à deux des assignations individuelles entre méthodes d'assignation différentes pour les individus de La Petite Charnie. En partant du haut les méthodes utilisées sont : Structure modèle *Admixture*, Structure modèle *Linkage*, GENELAND, GENEPATCH (TESS), NEWHYBRIDS avec Prior Uniformes et NEWHYBRIDS avec Prior de Jeffrey. En ce qui concerne les comparaisons deux à deux, chaque graphique représente en abscisse les valeurs d'assignation par une méthode et en ordonnée celles obtenues par une autre méthode. Les lignes en pointillé représentent sous forme graphique la corrélation des méthodes pour chaque espèce (ligne verte : chêne sessile, ligne rouge : chêne pédonculé et ligne bleue : chêne intermédiaires).

### 3.3.3.2. Correspondance entre détermination morphologique et assignation génétique

L'assignation génétique montre que la parcelle contient 150 chênes pédonculés, 104 chênes sessiles et 24 hybrides (total : 278 arbres). Le coefficient d'assignation génétique et l'indice morphologique sont très corrélés (Pearson's product-moment correlation :  $r = 0.89$ ,  $t = 32.48$ ,  $df = 267$ ,  $p\text{-value} < 0.001$ ), ce qui démontre une bonne concordance entre les groupes identifiés par l'assignation génétique et les espèces mise en évidence par la morphologie. Les chênes assignés génétiquement comme sessiles ont des valeurs négatives pour l'indice morphologique et inversement pour les chênes assignés pédonculés (Figure 16). Les chênes génétiquement hybrides (en rouge) ont des caractéristiques morphologiques de l'une ou l'autre espèce, alors que certains arbres caractéristiques d'une espèce montrent des morphologies intermédiaires. Ce résultat confirme que les espèces ont une grande variabilité morphologique intra-espèce. Les chênes présentant une morphologie intermédiaire ne sont pas forcément des hybrides et inversement.

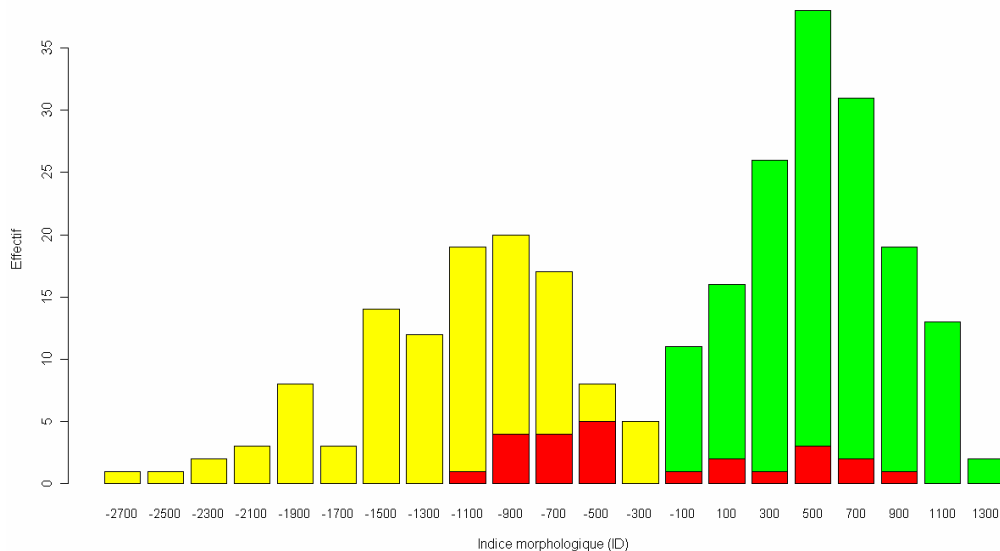


Figure 16 : Distribution de l'indice de morphologie foliaire en fonction de l'assignation génétique. La couleur indique les résultats de l'assignation génétique. En jaune, les individus assignés à l'espèce sessile, en vert, les individus assignés à l'espèce pédonculée et en rouge les individus hybrides.

### 3.3.4. Conclusion

Les résultats présentés dans cette partie montrent que 1) les différentes méthodes d'assignation non supervisée testées donnent des résultats très semblables malgré la diversité de leur approches et que 2) les méthodes d'assignations génétiques sont capables d'assigner les individus à des groupes homogènes correspondant aux espèces que l'on peut identifier par la morphologie foliaire.

Si l'identification des espèces se révèle très concordante entre données génétique et morphologique, ainsi qu'entre méthodes d'assignations, l'identification des hybrides est nettement plus difficile. D'une part, la morphologie foliaire intermédiaire d'un individu n'est pas un bon indicateur d'individu hybride. D'autre part, seule la prise en compte explicite de l'hybridation dans le modèle d'assignation génétique

permet l'identification des hybrides (ce qui réduit le choix à Structure et à NewHybrids). De nombreuses comparaisons entre Structure et NEWHYBRIDS montrent que les deux méthodes ont globalement la même performance pour étudier l'hybridation (Vaha & Primmer, 2006) et donnent des résultats concordants (Streiff *et al.*, 2005). Néanmoins, NewHybrids est spécialement adapté pour assigner les individus à une des espèces pures ou des classes hybrides, mais il ne peut être utilisé que pour l'hybridation entre deux espèces (E. Anderson, communication personnelle), ce qui limite grandement son utilité lorsque l'hybridation concerne plus de deux espèces.

L'article de Pritchard et collaborateurs (Pritchard *et al.*, 2000) est le premier à décrire une méthode d'assignation non supervisée ne reposant sur aucun *a priori* relatif aux populations échantillonnées. Cette propriété l'a rendu très attractif et on compte aujourd'hui de nombreuses applications de cette méthode avec plus 1780 citations de l'article original (juin 2008). Ce logiciel a plusieurs avantages qui en font un outil de choix pour l'étude du processus d'hybridation contemporaine dans les populations. D'une part, il intègre différents modèles dont deux sont explicitement adaptés pour la détection d'individus hybrides (modèles *Admixture* et *Linkage*). D'autre part, Structure étant publié depuis plus de 7 ans, de nombreuses auteurs l'ont utilisées en populations naturelles ou on testé l'efficacité de ses modèles grâce à des simulations (Vaha & Primmer, 2006). C'est donc le logiciel Structure qui va servir pour mener les analyses d'assignation. Un bref aperçu de son principe de fonctionnement est donné dans l'encadré 3.

Structure est facile à utiliser grâce à une documentation complète (Pritchard *et al.*, 2007) et à une interface graphique ergonomique. Néanmoins, l'interprétation des résultats est parfois ardue et leur fiabilité difficile à juger en l'absence de tests statistiques accompagnant l'assignation. Une solution consiste à utiliser des simulations dérivées des données expérimentales afin de tester les limites de la méthode et de pouvoir déterminer la précision et le degré d'interprétation spécifique applicable aux données analysées, c'est l'objectif de la fin de ce chapitre.

Dans la partie suivante, je décris une méthode d'utilisation du logiciel R permettant le traitement en chaîne d'une série d'analyses par Structure. J'applique cette technique afin de tester le nombre de marqueurs génétiques et le nombre d'individus nécessaires pour assigner efficacement les individus à leur espèce. Dans la dernière partie de ce chapitre, des simulations d'individus de quatre espèces de chênes ainsi que leurs hybrides permettent de fournir quelques critères objectifs pour assigner les individus, tester la puissance de Structure à identifier les hybrides et quantifier l'erreur de l'estimation du taux d'hybridation.

### Encadré 3 : Principe du fonctionnement du logiciel Structure

Supposons  $N$  individus échantillonnés dans un nombre de populations  $K$  inconnu. Le modèle caractérise chaque population par un jeu de paramètres inconnus à estimer :  $P$  les fréquences de chaque allèle à chaque locus. Les individus sont assignés de façon probabiliste à une ou plusieurs populations en fonction de leur génotype. Cette assignation probabiliste est effectuée en cherchant à minimiser le déséquilibre de liaison entre locus ainsi que l'écart à Hardy-Weinberg au sein des populations. Dans ce modèle, le nombre réel  $Z$  de population n'est pas un paramètre du modèle mais une variable inconnue fixe. Ainsi, pour déterminer le nombre de populations présent dans le jeu de données, il faut comparer les probabilités du modèle pour plusieurs valeurs de  $K$  et déterminer le nombre de populations  $K$  optimal pour lequel le modèle donne la meilleure vraisemblance en fonction des hypothèses d'équilibres intra-population et des données génétiques.

Structure inclut plusieurs types de modèles pour décrire les populations. Le modèle *No Admixture* suppose que chaque individu provienne d'une seule population, chaque population étant définie par ses fréquences alléliques. Le deuxième modèle de population proposé par Structure est le modèle *Admixture*, dans lequel chaque individu peut avoir pour origine plusieurs populations. Dans les deux cas, les résultats donnent la probabilité d'avoir observé les données pour chaque valeur du nombre de populations  $K$  testé, permettant d'estimer le nombre de populations le plus probable. Néanmoins ces deux modèles donnent des résultats qui ont un sens différent au niveau des individus.

- Le premier modèle considère que chaque individu appartient à une seule population et le résultat au niveau individuel est la probabilité d'appartenir à chacune des  $K$  populations.

- Dans le second modèle, les individus peuvent avoir pour origine plusieurs populations et ils sont caractérisés par la proportion de leur génome appartenant à chaque population. Ainsi un hybride entre deux populations aura la moitié de son génome appartenant à une population et l'autre moitié appartenant à une seconde population. Le modèle *Admixture* est donc particulièrement adapté pour étudier l'hybridation entre espèces.

- Un troisième modèle de population, *Use Population Information*, permet d'introduire dans l'analyse une information a priori sur l'appartenance des individus à l'une des populations. Ce modèle n'a que peu d'intérêt ici car il est équivalent aux méthodes d'assignations supervisées : il nécessite des populations de référence.

- Enfin, un dernier modèle, appelé modèle de *Linkage* a été développé afin de traiter le cas particulier de locus lié physiquement (Falush *et al.*, 2003). Dans ce cas, l'information sur la distance chromosomique entre locus est intégrée au modèle ce qui permet d'estimer l'origine de chaque portion de chromosome. La prise en compte des effets du déséquilibre de liaison, lié ici à la liaison physique des marqueurs, permet en théorie de détecter des événements d'hybridation plus anciens.

En plus des différents modèles concernant les populations, Structure peut utiliser deux modèles concernant les fréquences alléliques. Le premier modèle *Independent Allele Frequency* considère que les fréquences alléliques ne sont pas corrélées entre les populations. Ce modèle suppose que les populations ont divergé depuis suffisamment longtemps et que leurs fréquences alléliques sont très différentes. Ce modèle est utilisé lorsque des marqueurs génétiques très différenciés (du type marqueurs diagnostic) sont utilisés. Le second modèle de fréquences alléliques est le modèle *Correlated Allele Frequency* qui suppose que les fréquences alléliques entre populations sont très semblables. Ce modèle est le plus utilisé car les populations naturelles (ou les espèces proches) partagent souvent les mêmes allèles qui diffèrent seulement en fréquence.



## 3.4. Tests de sensibilité pour l'identification de deux espèces

### 3.4.1. Introduction

Comme nous l'avons vu dans la partie précédente, l'assignation génétique permet d'identifier l'espèce des individus d'une parcelle à partir de leur seul génotype multilocus. En utilisant toute l'information génétique disponible (tous les individus de la parcelle analysés pour 16 microsatellites) quelques arbres restent non assignés à l'une des deux espèces et sont considérés comme hybrides. Néanmoins, il n'est pas exclu que ces individus ne soient pas assignés par manque d'information génétique, l'ajout de quelques marqueurs génétiques permettrait peut-être de les assigner à l'une des deux espèces. Ceci nous amène à poser la question suivante : le nombre de marqueurs génétiques et d'individus utilisés pour l'assignation génétique est-il suffisant? L'effort d'échantillonnage et d'analyse nécessaire à l'utilisation d'une technique d'analyse génétique peut être testé par des techniques de ré-échantillonnage. En utilisant une partie des données, on peut estimer le nombre de locus et d'individus nécessaires pour pouvoir assigner les individus à leur espèce. Dans cette partie, je teste dans un premier temps l'effet du nombre de marqueurs microsatellites sur le taux d'assignation. Dans un second temps, je teste l'effet du nombre d'individus par espèce sur la fiabilité des résultats ainsi que sur le taux d'assignation. Ces résultats démontrent qu'un petit nombre de marqueurs microsatellites et d'individus par espèces suffisent à obtenir, pour les populations de chênes étudiées, une assignation génétique fiable.

### 3.4.2. Matériels et méthodes

#### 3.4.2.1. Données génétiques

- La Petite Charnie

Les données génétiques de la parcelle de la Petite Charnie (décrites précédemment, partie 4.3.2.1) ont été utilisées pour ces analyses de ré-échantillonnage.

- Populations ONF

Un échantillonnage couvrant 9 populations classées de l'ONF a été réalisé et génotypé par Erwan Guichoux pour 10 microsatellites communs avec La Petite Charnie (Lepais *et al.*, 2006b). D'après les déterminations morphologiques, cet échantillonnage comprend 295 chênes pédonculés et 486 chênes sessiles (soit 781 individus).

#### 3.4.2.2. Technique de ré-échantillonnage et assignation automatisée

R est un logiciel de calcul scientifique et de représentation graphique sous licence libre qui gagne en réputation parmi la communauté scientifique. Une grande communauté de développeurs existe et de nombreux programmes additionnels spécifiques à des types d'analyses très variés sont mis en ligne fréquemment (dont quelques packages spécifiques pour la génétique des populations). Comme tout logiciel de programmation en ligne de commande, R peut gérer la lecture et l'écriture dans un fichier extérieur ainsi que l'appel à des programmes exécutables en mode DOS externes.

Grâce à ces fonctions, il est possible d'automatiser la lecture et le tirage aléatoire de données, la mise en forme du fichier de données, l'envoi des données pour l'analyse par Structure, ce qui constitue la phase d'analyse (voir encadré 4). Concrètement, R enregistre le fichier de données dans le répertoire de Structure et lance son exécution pour finalement reprendre la main une fois l'analyse de Structure terminée. La lecture de la série de fichiers de résultats - afin d'en extraire les résultats essentiels - constitue la seconde partie du traitement des résultats (voir encadré 4). J'utilise cette méthode pour tester Structure avec un grand nombre de jeux de données différents de façon automatique. Deux types de ré-échantillonnages ont ainsi été effectués : l'un concerne le nombre de microsatellites et l'autre le nombre d'individus. Ces ré-échantillonnages ont pour but de déterminer les nombres minima et optima de locus et d'individus nécessaires à une bonne assignation génétique à l'espèce.

- Test de l'effet du nombre des marqueurs génétiques

J'ai utilisé la procédure de tirage aléatoire d'une sous-partie des données disponibles décrite précédemment afin de tester l'effet du nombre de locus employé et de leurs caractéristiques dans le taux de bonne assignation. Les données génétiques de la Petite Charnie, comprenant 16 marqueurs microsatellites, ont été utilisées dans cette analyse. Trois types de tirages ont été réalisés :

Le jeu de données entier a été utilisé pour tirer au hasard un nombre croissant de locus.

Pour chaque nombre de locus tiré au sort ( $i$  de 3 à 16), j'ai réalisé  $k = 100$  tirages aléatoires différents. Chacun des 1400 jeux de données ainsi obtenu a été soumis à l'analyse par Structure et le taux de bonne classification est moyenné pour chaque nombre de locus utilisés.

Le jeu de données de base se réduit aux 10 microsatellites sélectionnés pour étudier l'hybridation chez les chênes Lepais *et al.*, 2006b. La même procédure de tirage aléatoire a été utilisée, générant 800 jeux de données comprenant de 3 à 10 locus. Le taux d'assignations correctes a été calculé pour chaque analyse ainsi que sa moyenne par nombre de locus.

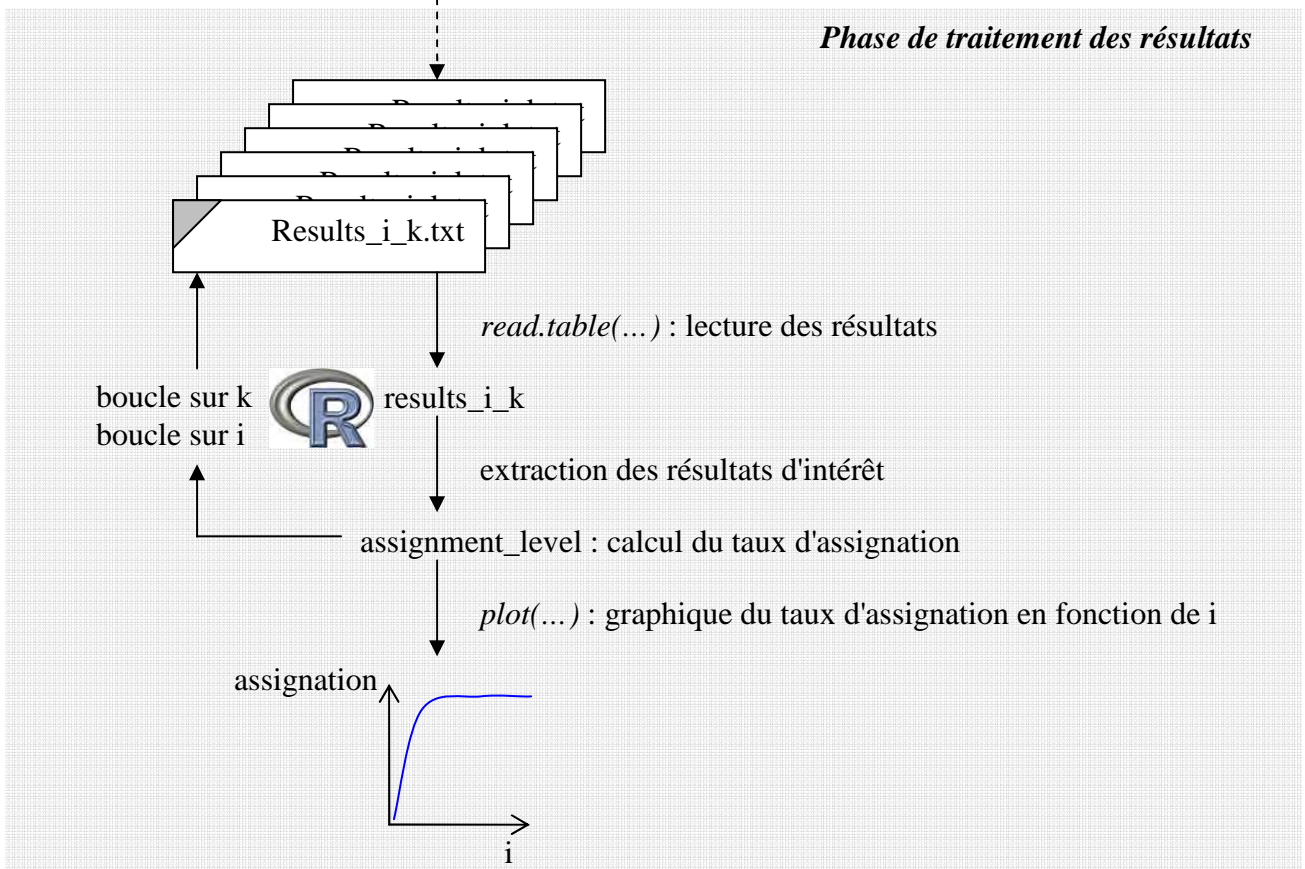
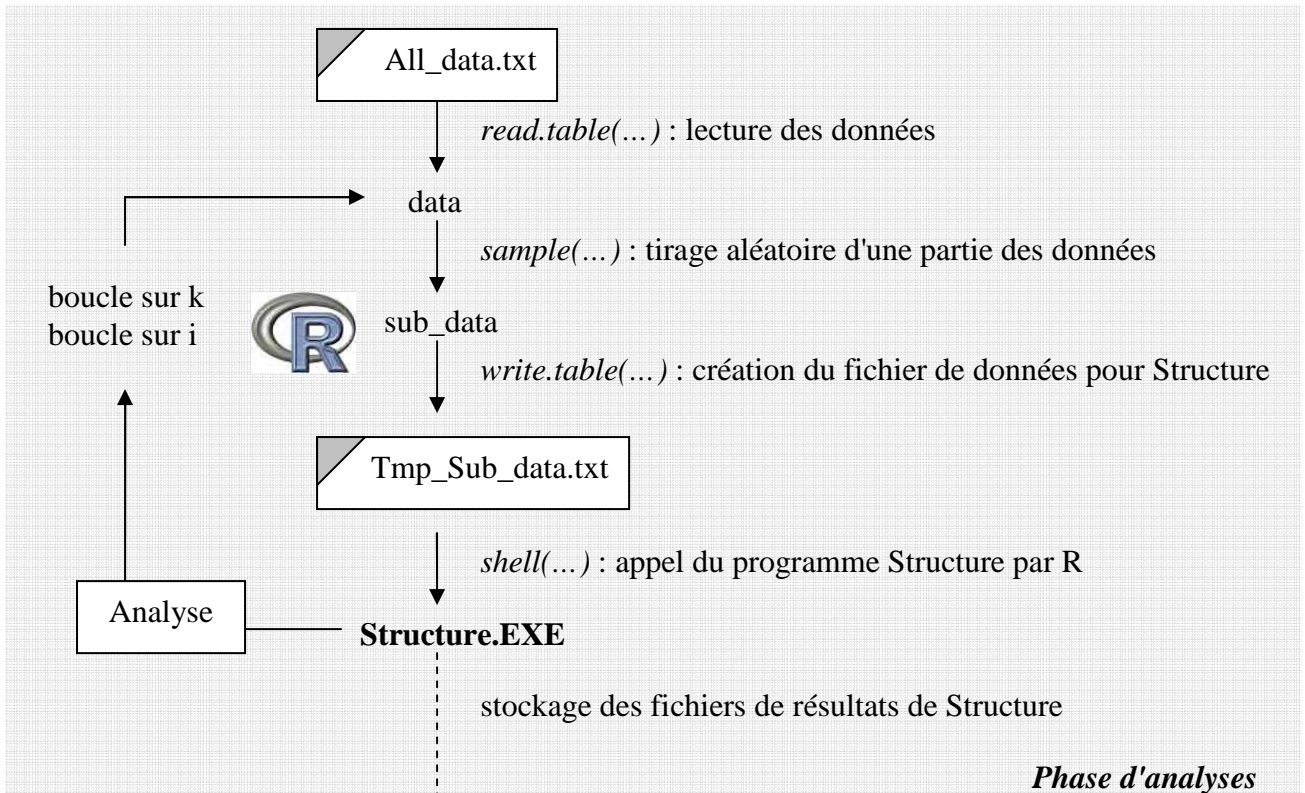
Enfin, j'ai voulu savoir s'il était possible de réduire encore plus le nombre de locus nécessaire en utilisant systématiquement les trois locus les plus différenciés entre les deux espèces, les autres locus étant tirés aléatoirement parmi les locus restants (moins différenciés). De cette façon, une autre série de 800 jeux de données a été obtenue, allant de 3 à 10 locus, chacun comprenant obligatoirement les trois locus supposés plus puissants (Qr11, Qr96 et Qr112).

Toutes les données ont été analysées avec les mêmes paramètres du logiciel Structure : le modèle de population *Admixture* et pour les fréquences alléliques le modèle *Allele Frequency Correlated* ont été utilisés. Un temps de chauffe (*Burnin period*) de 50000 cycles, qui permet d'atteindre la convergence du modèle, précède les 100000 répétitions de la chaîne MCMC au cours desquels sont enregistrés les résultats, comme cela est préconisé dans le manuel du logiciel (Pritchard *et al.*, 2007). Avec l'ensemble des locus utilisés, le taux d'assignation est de 90% (l'assignation génétique et

morphologique concorde pour 90% des individus), environ 10% des individus ne sont pas assignés à une espèce et sont considéré hybrides. Pour plus de clarté, j'ai considéré dans un premier temps les individus qui s'assignaient à l'une des deux espèces, laissant pour le moment de côté les individus hybrides.

**Encadré 4 : Algorithme de ré-échantillonnage et automatisation du logiciel Structure**

$i$  : nombre de locus ou d'individus à échantillonner ;  $k$  : nombre de ré-échantillonnages par nombre de locus ou d'individus



- Test de l'effet du nombre d'individus par espèces

Pour tester l'effet du nombre d'individus, j'ai procédé d'une façon similaire sauf qu'au lieu d'échantillonner les locus, j'ai échantillonné les individus parmi le jeu de données complet, le nombre de locus étant maintenu constant.

Les 10 microsatellites communs à la Petite Charnie et aux populations ONF (Erwan Guichoux) ont servi de données de base. Le nombre d'individus par espèce tiré au sort est compris entre 10 et 260 (parmi les 454 chênes pédonculés et 600 chênes sessiles) et pour chaque nombre d'individus par espèce, 100 jeux de données sont tirés au sort et analysés.

La même expérience de tirage aléatoire d'un nombre croissant d'individus est répétée à la Petite Charnie seule mais en incluant les 16 microsatellites disponibles dans le but de voir si l'utilisation d'un plus grand nombre de marqueurs microsatellites permet d'échantillonner un nombre réduit d'individus par espèce. Ici, le nombre d'individus ré-échantillonnés va de 10 à 100 individus par espèce avec toujours 100 tirages aléatoires par nombre d'individus par espèces.

A chaque analyse, certains paramètres sont mesurés, en premier lieu, la convergence des chaînes de Markov. Avec des jeux de données comprenant un petit nombre d'individus, le logiciel Structure risque de ne pas avoir suffisamment de données pour détecter les deux espèces. Dans ce cas, tous les individus sont assignés en proportion égale aux deux clusters et les résultats ne sont donc pas exploitables. Ce phénomène est facilement détectable par le fait que chaque cluster contient 50% de chaque espèce (proportion moyenne de l'espèce pédonculée dans le cluster1 = 0.5, proportion moyenne de chêne pédonculé dans la cluster2 : 0.5). Ainsi, un critère permettant d'identifier la convergence des chaînes de Markov peut être définie par la distribution des espèces dans chaque cluster. Si la convergence n'a pas eu lieu, les deux clusters comprendront une proportion égale de chaque espèce. Par exemple les chênes pédonculés seront également répartis entre les deux clusters soit :

$$\text{Moyenne du cluster1} - \text{Moyenne du cluster2} = 0.51 - 0.49 = 0.02$$

Si la convergence est atteinte, un cluster comprendra la grande majorité d'une espèce. Dans ce cas, les chênes pédonculés vont être majoritairement classés dans un cluster :

$$\text{Moyenne du cluster1} - \text{Moyenne du cluster2} = 0.90 - 0.10 = 0.80$$

Le critère de convergence que j'ai utilisé se focalise sur la distribution des chênes pédonculés entre les deux clusters :

- si : Moyenne du cluster1 - Moyenne du cluster 2 < 0.2 → non convergence
- si : Moyenne du cluster1 - Moyenne du cluster2 ≥ 0.2 → convergence

Le second paramètre mesuré est le taux d'assignation pour les analyses dont les chaînes de Markov ont convergé.

### 3.4.3. Résultats et discussions

#### 3.4.3.1. Effet du nombre de marqueurs génétiques

En ne tenant pas compte des hybrides, le taux d'assignation est maximal avec 16 locus (Figure 17). Il augmente rapidement avec le nombre de locus utilisé (courbe noire) pour un faible nombre de locus puis à partir de 8 locus, il augmente lentement et semble atteindre un plateau. Un nombre de 8 locus semble correspondre à un seuil pour lequel la variance du taux d'assignation faiblit fortement ce qui indique une stabilité des résultats au cours des différentes analyses. Si on utilise les 10 locus sélectionnés (courbe bleue), le taux d'assignation augmente de la même façon avec le nombre de locus utilisé mais le plateau est atteint pour seulement 7 locus. Lorsque les trois locus les plus différenciés sont utilisés (courbe rouge), le taux d'assignation est immédiatement meilleur, même avec trois locus, et le plateau est atteint avec 5 à 6 microsatellites. Avec les 10 locus sélectionnés, un taux d'assignation correcte de 96% est obtenu, alors que le fait de rajouter 6 locus permet de classer seulement 4% d'individus supplémentaires. Finalement, ce résultat met en évidence l'effet du niveau de différenciation interspécifique dans le taux d'assignation génétique : les 16 locus présentent une différenciation de  $G_{ST} = 0.029$ , alors qu'elle est de  $G_{ST} = 0.036$  pour les 10 locus sélectionnés (avec  $G_{ST} = 0.092$  pour les trois locus les plus différenciés et  $G_{ST} = 0.016$  pour les 7 autres).

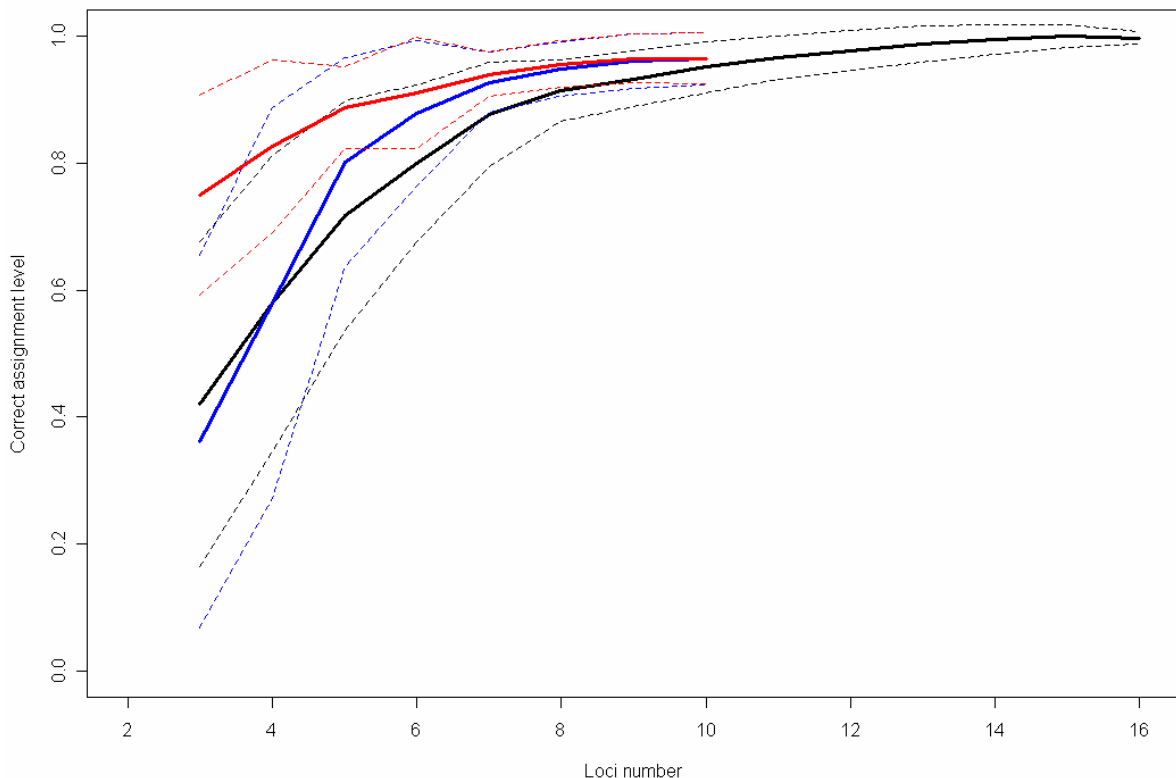


Figure 17 : Taux d'assignations correctes en fonction du nombre de locus utilisé. Résultats pour trois types de jeu de données : l'ensemble des 16 marqueurs disponibles (courbe noire), 10 locus sélectionnés (courbe bleue) et 10 locus sélectionnés avec les trois locus les plus différenciés systématiquement utilisées (courbe rouge).

### 3.4.3.2. Effet du nombre d'individus échantillonné par espèce.

- Convergence des chaînes de Markov

Avec moins de 30 individus par espèce, une grande majorité des analyses ne sont pas exploitables, c'est-à-dire que les chaînes de Markov n'ont pas convergé (Tableau 4). Dans ces cas, aucune structure n'est détectée. De manière globale, plus le nombre d'individus augmente, plus Structure produit des résultats exploitables (Tableau 4). Pour atteindre une fiabilité de 100%, il faut un minimum de 70 individus analysés pour 10 microsattellites ou 40 individus analysés pour 16 microsattellites. Au-delà de ces valeurs, toutes les analyses produisent des résultats fiables.

Tableau 4 : Pourcentage d'analyses convergentes (donc exploitables) en fonction du nombre de microsattellites et du nombre d'individus par espèces.

Nombre d'individus		10	20	30	40	50	60	70	80	90	100
Analyses convergentes (%)	10 SSR	2	39	64	89	96	99	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
	16 SSR	2	60	90	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>

- Taux d'assignations

Parmi les analyses dont la convergence est atteinte pour un nombre d'individus inférieur à 30 par espèce, les taux d'assignation peuvent atteindre 80% même si la variance est importante (Figure 18). Ainsi, même avec 30 individus par espèce pour 10 microsattellites ou 20 individus par espèces pour 16 microsattellites, il est possible de trouver des analyses exploitables donnant de très bons résultats d'assignation. Une augmentation des paramètres de longueur du temps de chauffe et des répétitions MCMC permettrait sans doute d'améliorer les résultats. A partir de 60 individus pour 10 microsattellites ou 40 individus pour 16 microsattellites, le taux d'assignation se stabilise autour de son maximum et l'écart-type du taux d'assignation est nettement réduit. Entre 100 et 260 individus pour 10 microsattellites, le taux d'assignation n'augmente presque plus, alors que l'écart-type se réduit doucement. Cette série d'analyses montre qu'un minimum de 60 individus par espèce est nécessaire afin d'assigner les individus correctement pour 10 microsattellites, alors qu'un nombre de 40 individus suffit lorsque 16 marqueurs microsattellites sont utilisés. Néanmoins, dans les cas où le nombre d'individus disponibles est limité, une augmentation des paramètres de longueur du temps de chauffe et des répétitions MCMC, conjointement à une sélection des analyses convergentes, peut permettre d'obtenir des assignations fiables pour un nombre réduit d'individus (30 pour 10 microsattellites ou 20 pour 16 microsattellites).

Ces résultats démontrent qu'il faut être particulièrement attentif lorsque plusieurs espèces (ou populations) en effectifs déséquilibrés sont analysées : si une espèce est représentée par quelques individus seulement, elle risque de ne pas être identifiée, ou sera mal assignée.

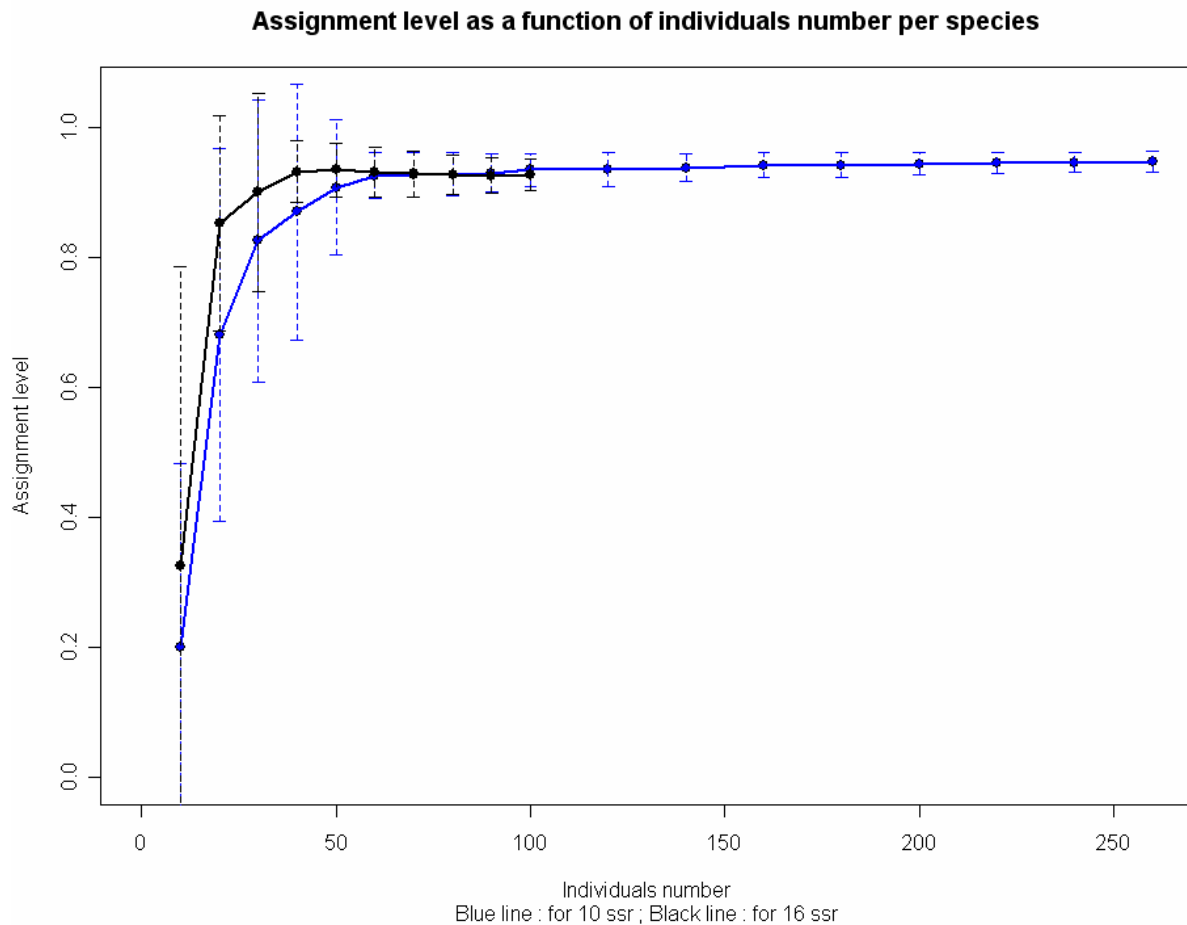


Figure 18 : Taux d'assignation en fonction du nombre d'individus par espèce. Les courbes correspondent au taux d'assignation et à son écart-type pour les analyses exploitables (convergence des chaînes de Markov). En bleu : 10 microsatellites sont utilisés ; en noir, 16 microsatellites utilisés.

### 3.4.4. Conclusion

Les résultats d'assignation issus des ré-échantillonnages montrent clairement que le génotype d'une cinquantaine d'individus par espèce pour 10 locus microsatellites fournit suffisamment d'information au logiciel Structure pour assigner les individus à leur espèce. Le taux d'assignation n'augmentant plus au-delà d'une certaine quantité d'information génétique fournie, les individus qui ne sont pas assignés à une espèce sont donc bien des individus hybrides.

Bien qu'étant des entités relativement peu différenciées, les espèces de chênes peuvent être bien assignées génétiquement pour un nombre d'individus facilement échantillonnable et un nombre de locus différencié relativement raisonnable.

Comme nous l'avons vu, certains individus ne sont jamais assignés à une espèce, quel que soit le nombre de locus microsatellites utilisé ou le nombre d'individus par espèce, ce qui laisse penser que ces individus sont des hybrides. Nous allons voir dans la dernière partie de ce chapitre s'il est possible de détecter des individus hybrides par la méthode d'assignation génétique et ce grâce à des simulations.



## 3.5. Identification des hybrides au sein d'un complexe d'espèces

### 3.5.1. Introduction

Le niveau de différenciation génétique entre espèces est un indicateur de la possibilité d'assignation des individus à leur espèce respective. Comme le montrent les résultats de la partie précédente, plus la différenciation génétique est élevée, plus l'assignation génétique sera aisée. Les hybrides sont des individus qui résultent de la reproduction entre des espèces différentes, ce qui implique que ces hybrides sont faiblement différenciés de leurs espèces parentales. En d'autres termes, les hybrides risquent d'être plus difficiles à assigner génétiquement.

Dans cette partie, j'ai testé la capacité de Structure à identifier des hybrides dans le cas où il y a quatre espèces de chênes pouvant s'hybrider. A partir de fréquences alléliques des quatre espèces de chênes, j'ai utilisé des simulations afin de créer des individus appartenant à l'une des espèces ou des individus hybrides. Connaissant la nature des individus simulés, il est possible (1) d'expérimenter la capacité de l'assignation génétique à identifier les individus hybrides, (2) de déterminer des critères objectifs permettant d'optimiser l'assignation génétique et (3) d'estimer l'erreur inhérente à cette analyse. Ce sont les objectifs de la dernière partie de ce chapitre.

### 3.5.2. Matériels et méthodes

#### 3.5.3. Données génétiques

Une analyse à l'aide du logiciel Structure intégrant différentes populations a été effectuée dans le but d'identifier des individus d'espèce pure pour chaque espèce afin de définir les compositions alléliques des espèces (voir chapitre 4). 65 individus ayant une probabilité supérieure à 0.90 d'appartenir à une espèce ont été utilisés pour calculer les fréquences alléliques de chaque espèce à chaque locus.

##### 3.5.3.1. Simulations des individus

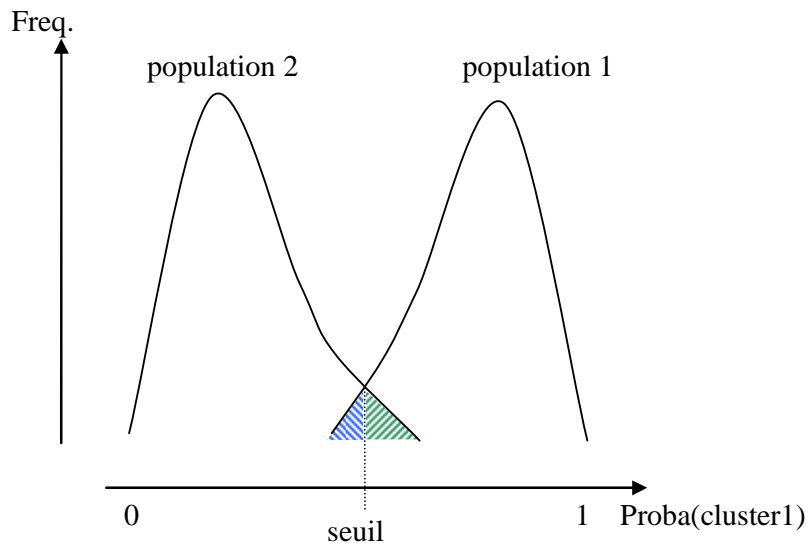
Les fréquences alléliques obtenues ont été utilisées afin de simuler des individus appartenant à une espèce pure ainsi que différentes classes d'hybrides (Nielsen *et al.*, 2003; Nielsen *et al.*, 2006). Un script a été réalisé avec le logiciel R permettant de simuler des individus purs, hybrides de première génération (F1) ou de seconde génération (croisement entre individus F1 et l'une des deux espèces parentales). Nous avons simulé 1000 individus de chaque espèce, 30 hybrides F1 pour chaque couple d'espèce ainsi que 60 rétrocroisements (hybrides de seconde génération) de chaque classe (F1 x espèce parente 1 et F1 x espèce parente 2). Les proportions d'hybrides simulées (environ 20% d'hybrides et 80% d'espèces pures) sont des proportions réalistes comparées à ce que l'on peut identifier en populations naturelles (Curtu *et al.*, 2007; Valbuena-Carabana *et al.*, 2005).

##### 3.5.3.2. Assignation

J'ai utilisé ces données simulées pour tester la puissance de Structure à assigner les individus à leur espèce ou à leur classe d'hybride. De plus, ces résultats permettent de déterminer des seuils d'assignation optimum.

### Encadré 5 : Efficacité, précision et erreurs d'assignation+

Supposons deux populations dont on connaît les fréquences alléliques à partir desquelles on simule 100 individus de chaque population par tirage aléatoire des allèles pondéré par leur fréquence. L'assignation génétique donne pour chaque individu la probabilité d'appartenir à chacun des deux clusters. Voici un schéma montrant la distribution des probabilités d'appartenance des individus au cluster 1 :



Nous voyons que les courbes de distribution se recoupent au milieu ce qui signifie que l'assignation ne sera pas parfaite. Soit les individus présentant des valeurs intermédiaires sont déclarés non assignables (ce qui peut représenter un nombre important d'individus), ou bien on détermine un seuil d'assignation à l'intersection des courbes (par exemple 0.5 dans cet exemple) : les individus présentant une valeur supérieure sont assignés à la population 1 alors que ceux qui ont une valeur inférieure au seuil sont assignés à la population 2. Cette dernière solution implique des erreurs d'assignation (parties hachurées) qui peuvent être quantifiées.

Connaissant la nature des individus simulés, il est possible de calculer deux estimateurs de la justesse d'une assignation génétique (Vaha & Primmer, 2006). Soit le tableau qui dénombre les individus assignés en rapport avec les individus simulés :

Simulé \ Assigné	Pop1	Pop2	total simulé
Pop1	95	5	100
Pop2	8	92	100
total assigné	103	97	

L'efficacité d'assignation permet d'estimer le nombre d'individus assignés à sa bonne population et se calcule ainsi :

Efficacité(pop1) =  $95/100 = 0.95$  ; Efficacité(pop2) =  $92/100 = 0.92$

La précision quantifie la composition des individus assignés à une population, ainsi :

Précision(pop1) =  $95/103 = 0.92$  ; Précision(pop2) =  $92/97 = 0.95$

Enfin, la performance globale correspond au produit des deux estimateurs. Ici pour les deux populations, la performance est égale à  $0.95 \times 0.92$  soit 0.87.

Ces paramètres sont très utiles afin de trouver un seuil optimal pour assigner les individus ainsi que pour apprécier l'interprétation de ces résultats au regard de l'information contenue dans les données génétiques.

### 3.5.3.3. Estimation de la performance de l'assignation

L'assignation des individus est testée par deux paramètres : l'efficacité d'assignation et la précision (Vaha & Primmer, 2006 ; voir Encadré 5). L'efficacité représente le nombre d'individu correctement assigné à sa classe divisé par le nombre d'individus simulés appartenant à cette classe (par exemple : Efficacité assignation chêne pédonculé = nombre d'individus assignés comme chêne pédonculé / nombre de chênes pédonculés simulés). La précision est le rapport du nombre d'individu correctement assigné à sa classe par le nombre d'individus total assigné dans cette même classe (Précision d'assignation chêne pédonculé = nombre d'individus correctement assigné comme chêne pédonculé / nombre d'individus assignés à la classe chêne pédonculé). Finalement, la performance globale de la méthode d'assignation est le produit de ces deux paramètres.

### 3.5.3.4. Estimation du taux d'hybridation

Afin d'apprécier la précision du taux d'hybridation estimé par l'assignation génétique, j'ai comparé le taux d'hybridation observé à celui réellement simulé. L'erreur calculée en pourcentage permet de mettre en évidence un éventuel biais sur l'estimation de ce taux.

## 3.5.4. Résultats et discussions

### 3.5.4.1. Coefficient d'admixture par classe et choix de seuils d'assignation

En observant les distributions des coefficients d'admixture (CA), on constate des différences flagrantes entre les différentes classes d'individus (Figure 19). Les individus d'espèces pures ont des coefficients d'admixture proche de 1 (Figure 19A ; moyenne =  $0.98 \pm 0.03$  SD, médiane = 0.98, min = 0.59, max = 0.99). Les hybrides de première génération ont une moyenne d'admixture proche de 0.5 avec une variance élevée (Figure 19B ; moyenne =  $0.46 \pm 0.17$  SD, médiane = 0.49, min = 0.02, max = 0.86). Les hybrides de seconde génération ont une moyenne d'admixture proche de 0.75 avec là aussi une variance élevée et des distributions légèrement différentes en fonction du couple d'espèce considéré (Figure 19C ; moyenne =  $0.76 \pm 0.17$  SD, médiane = 0.78, min = 0.03, max = 0.99). Globalement (Figure 19D) les distributions des coefficients d'admixture pour ces trois classes d'individus se chevauchent ce qui indique qu'un même coefficient d'admixture peut correspondre à différentes classes d'hybrides. Afin d'assigner les individus dans des classes, des seuils sont déterminés afin de discrétiser une information continue. Pour les espèces pures un seuil de 0.90 apparaît raisonnable, d'autant que ce même seuil a servi à définir les fréquences alléliques permettant d'effectuer les simulations (voir matériels et méthodes). Les distributions des coefficients d'admixture pour les hybrides de première et seconde génération se recoupent à environ 0.65, ce qui constitue le seuil optimum minimisant les erreurs d'assignation.

Nous allons tester deux types d'assignations. Une première, optimiste, essayant de distinguer les deux générations d'hybrides :

$$0.35 \leq CA < 0.65 \rightarrow \text{hybride F1}$$

$0.65 \leq CA < 0.90 \rightarrow$  rétrocroisement

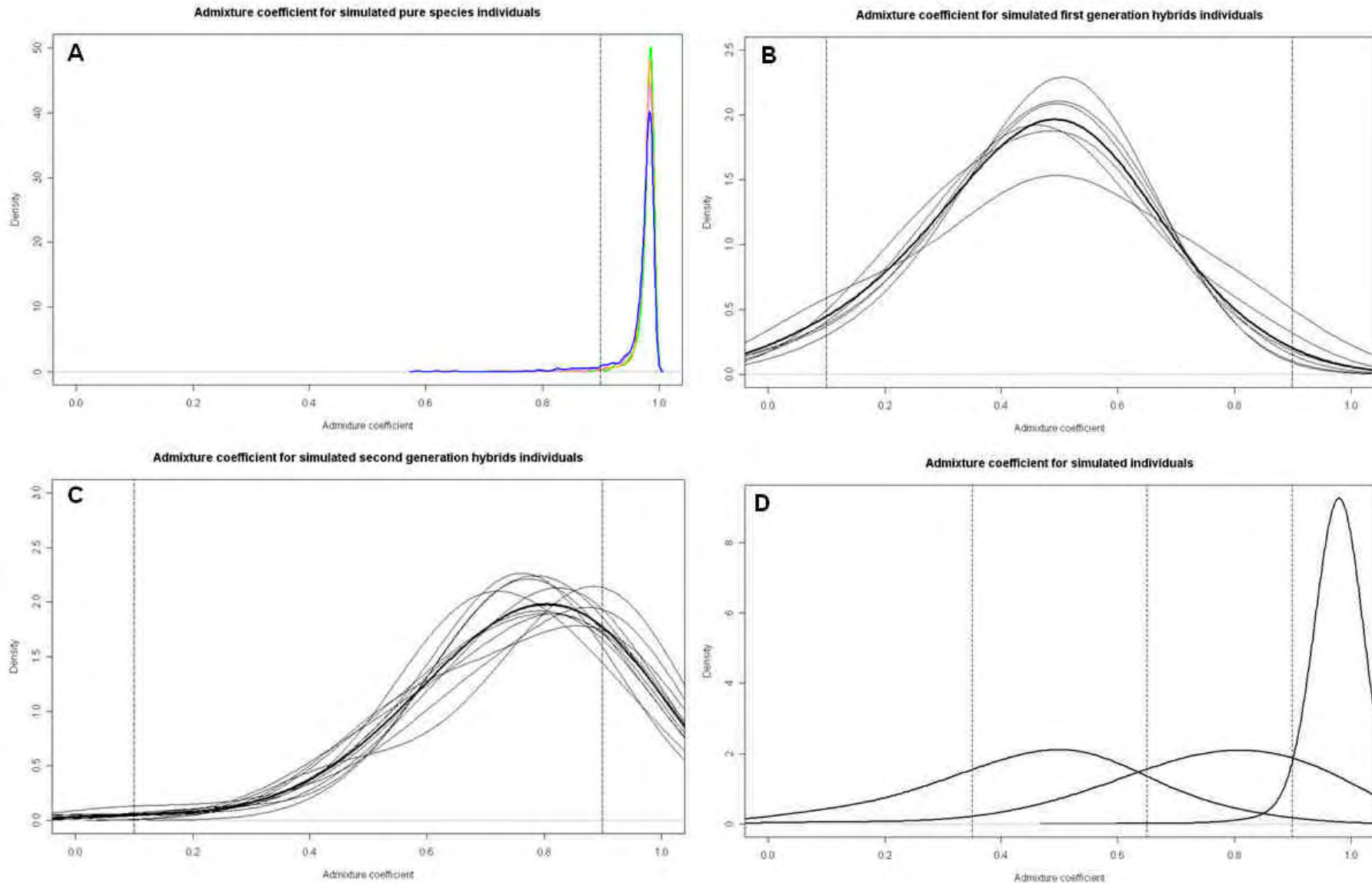
$CA \geq 0.90 \rightarrow$  espèce pure

Une seconde, plus réaliste, qui se contentera d'identifier les espèces pures et les hybrides :

$CA < 0.90 \rightarrow$  hybride

$CA \geq 0.90 \rightarrow$  espèce pure

Enfin, 2% des individus présentant un coefficient d'admixture élevé pour l'un des clusters (inférieur à 0.90) mais des coefficients inférieurs à 0.1 pour les trois autres clusters sont classés comme individus d'espèce pure du cluster majoritaire.



**Figure 19** : Distribution des coefficients d'admixture pour chaque type d'individus simulés. A : individus appartenant à une des quatre espèces, B : individus hybrides de première génération, C : individus hybrides de seconde génération (rétrocroisements) et D : courbes moyenne pour les espèces pures, les hybrides F1 et les rétrocroisements. Les lignes fines correspondent à un type d'hybride, la ligne épaisse correspond à la moyenne. Les lignes verticales correspondent aux seuils d'assignation choisis.

Tableau 5 : Répartition des individus simulés (lignes) dans les différentes classes d'assignation (colonnes). Les assignations correctes sont indiquées en gras.

Simulé\Assigné	Ped	Ses	Toz	Pub	Hyb PedSes	Hyb PedToz	Hyb PedPub	Hyb SesToz	Hyb SesPub	Hyb TozPub	total
Ped	<b>996</b>				3	1					1000
Ses		<b>988</b>			7			4	1		1000
Toz			<b>992</b>			3		2		3	1000
Pub				<b>972</b>			4		14	10	1000
F1_PedSes		3			<b>27</b>						30
bc_PedSes	19	1			<b>38</b>	2					60
bc_SesPed		20			<b>38</b>			2			60
F1_PedToz			1			<b>28</b>				1	30
bc_PedToz	15				2	<b>40</b>	3				60
bc_TozPed	1		16			<b>41</b>		1		1	60
F1_PedPub				1			<b>29</b>				30
bc_PedPub	17				4	1	<b>38</b>				60
bc_PubPed				11	1		<b>44</b>		1	3	60
F1_SesToz			1					<b>29</b>			30
bc_SesToz		16			2	1		<b>38</b>	3		60
bc_TozSes			23			2		<b>32</b>		3	60
F1_SesPub							1		<b>29</b>		30
bc_SesPub		19	1		1		4		<b>35</b>		60
bc_PubSes				29			2		<b>27</b>	2	60
F1_TozPub								1	2	<b>27</b>	30
bc_TozPub			16			1		4	1	<b>38</b>	60
bc_PubToz			1	25				1	2	<b>31</b>	60
total	1048	1047	1051	1038	123	120	121	118	115	119	4900

Tableau 6 : Efficacité, précision et performance globale de l'assignation génétique.

Classe	Efficacité	Précision	Performance
Ped	99.60%	95.04%	94.66%
Ses	98.80%	94.36%	93.23%
Toz	99.20%	94.39%	93.63%
Pub	97.20%	93.64%	91.02%
Hyb_PedSes	68.67%	83.74%	57.50%
Hyb_PedToz	72.67%	90.83%	66.01%
Hyb_PedPub	74.00%	91.74%	67.88%
Hyb_SesToz	66.00%	83.90%	55.37%
Hyb_SesPub	60.67%	79.13%	48.01%
Hyb_TozPub	64.00%	80.67%	51.63%

#### 3.5.4.2. Identification des deux générations d'hybrides et des espèces pures

En appliquant les seuils définis précédemment il n'est pas possible d'identifier de façon fiable les hybrides de première et seconde génération (taux d'assignation moyen respectif : 55% et 45%). Les résultats détaillés sont présentés dans les tableaux fournis en annexe (Annexe 2).

Ces mauvais résultats laissent penser que les 10 microsattellites utilisés ne sont pas assez informatifs pour distinguer les différentes classes d'hybrides simulées. Ce résultat n'est pas étonnant si on considère que pour arriver à une telle précision avec un taux de différenciation de l'ordre de  $F_{ST} = 0.10$  comme c'est le cas ici, il faudrait plus de 48 microsattellites pour obtenir une performance correcte (Vaha & Primmer, 2006). Nous allons donc tester si les données sont suffisamment puissantes pour identifier les espèces pures des individus hybrides, sans distinguer les deux générations d'hybrides.

#### 3.5.4.3. Identification des hybrides et des espèces pures

Si on utilise un seuil de 0.90 pour identifier les espèces des hybrides, l'efficacité d'assignation des espèces pures est bonne (Tableau 5 et Tableau 6). Elle est inférieure pour les hybrides avec des taux de bonnes assignations allant de 61% à 73%. La précision est bonne pour les espèces pures mais inférieure pour les hybrides. Les classes pures contiennent une grande proportion d'individus purs alors que les classes hybrides contiennent une proportion non négligeable d'individus purs et d'hybrides appartenant à d'autres espèces. Il faut tout de même remarquer que les hybrides mal assignés sont en général assignés à l'une des deux espèces parentale (Tableau 5). Ainsi seulement 2.6 % des hybrides pédonculé x sessile sont assignés à une autre classe hybride (précisément, 2 sont assignés comme hybride pédonculé x taurin alors que 2 autres sont assignés comme hybride sessile x taurin). Pour les hybrides pubescent x taurin, dont les espèces parentales sont les moins différenciées, le taux d'erreur d'assignation des hybrides est de 8% (8% des hybrides pubescent x taurin sont assignés à d'autres classes hybride). En général, la performance de la méthode reste bonne pour identifier les espèces pures et les hybrides (de 90 à 97% des F1 sont correctement assignés, mais seulement 52 à 68% des rétrocroisements sont correctement assignés à leur classe hybride, voir Tableau 5). De façon plus globale avec les 10 microsattellites utilisés présentant une différenciation génétique globale ( $F_{ST}$ ) entre espèce de 0.07, les hybrides de première génération sont correctement assignés dans plus de 90% des cas alors que 33% des rétrocroisements sont assignés par erreur à l'une des espèces parentales.

Ces résultats peuvent directement être comparés à des simulations effectuées chez divers mammifères (Randi, 2008). En utilisant de 20 à 30 microsattellites dont la différenciation globale ( $F_{ST}$ ) entre espèces s'étend de 0.10 à 0.20, 100% des hybrides de première génération entre le chien et le loup ou entre le chat domestique et le chat sauvage sont correctement assignés alors que 20 % des rétrocroisements sont assignés à l'une de leur espèce parentale. Chez la perdrix, l'utilisation de 8 microsattellites présentant une forte différenciation ( $F_{ST} = 0.5$ ) permet d'assigner tous les hybrides de première génération ainsi que 90 % des rétrocroisements (Randi, 2008).

Comme nous l'avons vu, des erreurs d'assignation au niveau individuel sont possibles même si elles restent rares et cantonnées aux marges des distributions des coefficients d'admixture. Alors que l'espèce ou la paire d'espèces dans le cas d'un hybride sera déterminée avec une grande précision, l'incertitude principale réside dans la difficulté de délimiter espèce pure, hybride de seconde génération et de première génération. Qu'en est-il de l'estimation du taux d'hybridation dans l'ensemble de la population ?

#### 3.5.4.4. Estimation du pourcentage d'hybrides

L'estimation du pourcentage d'hybrides dans le jeu de données est biaisée en faveur des espèces pures (Tableau 7). Ainsi, le pourcentage de chaque espèce pure est surestimé d'environ 5% alors que le taux d'hybrides est sous-estimé d'environ 20%. Ce résultat s'explique par le fait que certains hybrides de seconde génération sont assignés à une espèce pure. Une légère augmentation du seuil d'assignation pourrait permettre de minimiser cette erreur (par exemple 0.92 au lieu de 0.90 pour assigner un individu à une espèce pure) et augmenterait la précision d'assignation des hybrides. Cependant cette modification entraînerait une perte d'efficacité d'assignation des individus appartenant à une espèce et conduirait globalement à une performance d'assignation équivalente. Enfin, il est difficile de conclure précisément si ces résultats obtenus grâce à des individus simulés reflètent de manière aussi fine les résultats que l'on peut obtenir en populations naturelles.

Tableau 7 : Pourcentages de chaque espèce et hybrides simulés et estimés.

Classe	% simulé	% estimé	erreur
Ped	20.41%	21.39%	4.80%
Ses	20.41%	21.37%	4.70%
Toz	20.41%	21.45%	5.10%
Pub	20.41%	21.18%	3.80%
Hyb_PedSes	3.06%	2.51%	-18.00%
Hyb_PedToz	3.06%	2.45%	-20.00%
Hyb_PedPub	3.06%	2.47%	-19.33%
Hyb_SesToz	3.06%	2.41%	-21.33%
Hyb_SesPub	3.06%	2.35%	-23.33%
Hyb_TozPub	3.06%	2.43%	-20.67%

#### 3.5.5. Conclusion

L'analyse d'individus d'origine connue *a priori*, car simulés à partir de fréquences alléliques des espèces étudiées, a permis de montrer l'intérêt et les limites inhérentes à l'assignation génétique. Les caractéristiques génétiques des espèces étudiées (grande diversité, faible différenciation) et le nombre relativement limité de marqueurs microsatellites utilisé ne permettent pas de distinguer les différentes générations d'hybrides. La performance d'assignation des chênes appartenant à une espèce est élevée, mais elle est moindre lorsqu'il s'agit d'identifier des individus hybrides.



Bien que certains hybrides (en particulier les rétro croisements) soient assignés à une espèce, ils sont dans la grande majorité des cas assignés à l'une des deux espèces parentales mais sont en revanche rarement assignés à une autre classe hybride. Pour les analyses en population, ce type d'erreur pourra être minimisé en analysant les groupes d'individus hybrides. Enfin, ces simulations pourront être utiles afin de comparer les individus hybrides identifiés en populations naturelles aux individus hybrides simulés (chapitre 4).

### 3.6. Conclusion générale

Les méthodes d'assignation génétiques récentes permettent, avec des données génétiques suffisamment informatives, d'identifier l'espèce d'un individu inconnu à partir de son génotype. La corrélation étroite entre assignation morphologique et génétique a été illustrée par l'exemple de la distinction entre les chênes pédonculés et sessiles à la Petite Charnie. Les individus génétiquement intermédiaires (hybrides) présentent en moyenne une morphologie intermédiaire et les espèces une grande variabilité morphologique (voir Viscosi *et al.*, 2008b et Annexe 3).

Bien que peu différenciés génétiquement, les individus peuvent être assignés aux espèces à partir d'un nombre réduit de locus microsatellites. J'ai montré que ce nombre dépendait de la différenciation interspécifique des locus employés : un sous-ensemble de 5 à 6 microsatellites différenciés permet d'atteindre des taux d'assignation élevés.

Le nombre d'individus présents dans le jeu de données conditionne aussi la fiabilité de l'analyse d'assignation. Pour des effectifs trop faibles (inférieurs à 30 individus par espèce), les résultats risquent de ne pas être fiables pour un nombre limité de locus. Lorsque le nombre d'individus est le facteur limitant, il est toujours possible d'analyser un plus grand nombre de locus, ce qui va augmenter la fiabilité ainsi que l'assignation génétique. En pratique, il est possible d'assigner correctement les individus aux espèces pédonculées et sessiles à partir de 30 individus par espèce pour 10 locus ou 20 individus et 16 locus. Dans le cadre d'application pour la traçabilité et le contrôle de l'espèce à partir de bois ou autres matériaux forestiers, une base de génotypes de référence incluant toutes les espèces appartenant au complexe (donc au-delà des quatre espèces étudiées ici : *Quercus faginea* ou *Quercus frainetto* par exemple) pourrait permettre des assignations génétiques en routine. Disposant de génotype de références pour les espèces, des assignations génétiques supervisées, plus simples et rapides, pourront être appliquées.

Pour l'identification des hybrides, j'ai démontré grâce aux simulations de deux générations d'hybridation entre quatre espèces de chênes qu'il était possible d'identifier les hybrides des espèces pures sans cependant pouvoir distinguer les deux générations d'hybrides (F1 et rétrocroisements). La performance est relativement moyenne car certains hybrides de seconde génération sont confondus avec une espèce pure, ce qui se traduit au niveau de l'estimation du taux d'hybride en population par une sous-estimation de l'hybridation. Les erreurs d'assignation au niveau individuel sont néanmoins à relativiser car l'assignation de l'espèce ou du couple d'espèce dans le cas d'hybrides est fiable.

Ces résultats sont à pondérer car ils résultent de génotypes recréés à partir des fréquences alléliques. D'une part, ces fréquences ne sont qu'une estimation (à partir de 65 individus), d'autre part, les phases

gamétiques (générant du déséquilibre de liaison) sont perdues. J'ai resimulé les individus en créant des gamètes directement à partir des génotypes adultes mais les mêmes taux d'assignation ont été obtenus. Dans tout les cas, c'est une reproduction aléatoire qui a été simulée, or une partie de l'information liée au déséquilibre de liaison peut manquer comparativement à une population réelle où les gamètes ne s'associent pas au hasard. De plus, les données simulées qui ont été testées avaient toutes des effectifs d'individus équilibrées ce qui n'est pas le cas dans les populations naturelles.

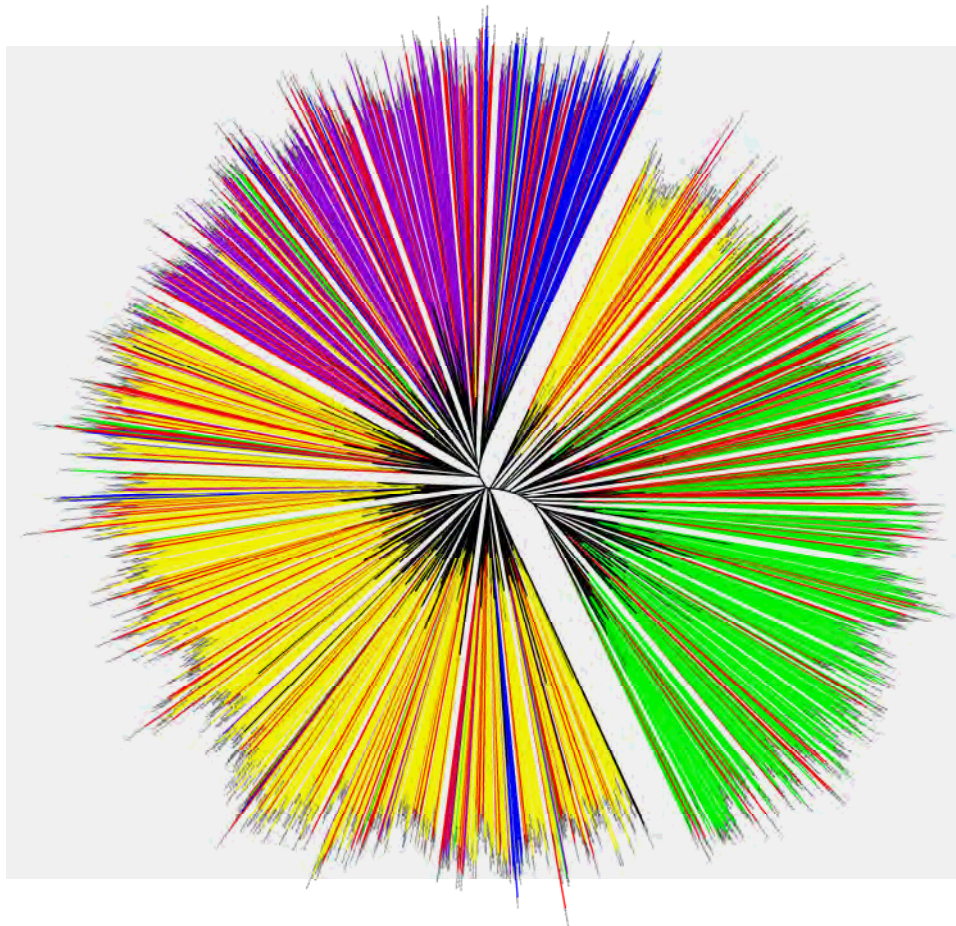
La différenciation génétique n'est pas le critère le plus efficace pour sélectionner les marqueurs les plus informatifs car le nombre d'allèles et la diversité génétique jouent un rôle important dans leur pouvoir discriminant (Banks *et al.*, 2003 ; Rosenberg, 2005). Ainsi, pour réduire encore le nombre de marqueurs génétiques nécessaire à l'identification des espèces, des indices tels que la différenciation standardisée (Hedrick, 2005 ; Smith *et al.*, 2001) ou d'autres indices de pouvoir discriminant (Banks & Eichert, 2000 Topchy *et al.*, 2004) devront être utilisés dans la phase de sélection des locus. De nouveaux marqueurs génétiques, tels que les microsatellites isolés de gènes (EST-SSR) pourraient s'avérer plus efficaces pour discriminer les espèces que les microsatellites génomiques. Des comparaisons entre microsatellites génomiques et génétiques chez diverses espèces végétales montrent que les microsatellites isolés de séquences transcrites sont plus conservés entre espèces proches, ce qui permet une meilleure transférabilité (Pashley *et al.*, 2006) ainsi que des analyses plus faciles et rapides (profils alléliques plus clairs : Leigh *et al.*, 2003 ; Chagné *et al.*, 2004 ; moins d'allèles nuls : Rungis *et al.*, 2004). Bien que généralement moins polymorphes que les microsatellites génomiques (Ellis & Burke, 2007), les microsatellites géniques peuvent présenter davantage de différenciation entre populations ou espèces car il est plus probable qu'ils soient soumis à une forme de sélection (Yatabe *et al.*, 2007). Enfin, ces marqueurs sont beaucoup moins coûteux à développer que les microsatellites génomiques (Ellis & Burke, 2007). Ces avantages en font des marqueurs de choix pour de futures études de l'hybridation entre espèces de chênes. Le grand nombre de séquences EST disponibles chez les chênes permet déjà d'identifier et de développer de nombreux microsatellites géniques dont certains présentent une forte différenciation entre espèces (Erwan Guichoux, communication personnelle). De plus leur relative facilité d'analyse devrait permettre de les multiplexer en grand nombre lors du génotypage, en offrant une compréhension beaucoup plus fine de l'hybridation chez les chênes.



# **PARTIE B – DÉLIMITATION DES ESPÈCES, ASSIGNATION GÉNÉTIQUE EN POPULATIONS**



# Chapitre 4- Hybridation en populations naturelles



Page précédente : Arbre des distances génétiques entre 2000 chênes. En rouge, les hybrides. .

## 4.1. Introduction

Le rôle de l'hybridation dans l'évolution des espèces fait débat depuis plus de 70 ans et cette question reste un sujet de recherche très actif encore de nos jours. Un grand nombre de travaux sur des populations naturelles ont montré que les conséquences de l'hybridation pouvaient être très variées en fonction des mécanismes d'isolement reproducteur (pré- et postzygotiques) en jeux. La disparition de barrières prézygotiques peut mener à une homogénéisation totale des espèces si elles n'ont pas développé de mécanismes d'isolement postzygotique. Chez d'autres espèces, il semble se produire un équilibre entre hybridation et sélection pour les formes parentales (ou contre les hybrides) rendant possible le maintien des espèces. Parfois les flux de gènes interspécifiques se révèlent asymétriques, voire unidirectionnels, ce qui conduit au processus d'introgession défini par Anderson en 1938 par le terme introgressive hybridation (Anderson & Hubricht, 1938). C'est particulièrement le cas lorsqu'une espèce est introduite ou migre dans l'aire d'une espèce taxonomiquement proche. Si l'isolement reproducteur est faible entre ces espèces, l'espèce introduite va se reproduire avec l'espèce préexistante, formant essentiellement des hybrides de première génération qui, s'ils sont eux même viables et fertiles, vont se reproduire avec l'espèce la plus fréquente, formant des rétrocroisements successifs. Différentes contraintes dans cette dynamique (isolement pré- ou postzygotique partiel, stérilité des hybrides, sélection pour ou contre les hybrides...) peuvent conduire à différents équilibres ou à l'extinction de l'une des deux espèces. Des dynamiques d'hybridations très variées ont été mises en évidence chez les plantes et les animaux. Les traits d'histoire de vie des arbres, caractérisés par un temps de génération long, un taux d'allofécondation élevé, des aires de distribution étendues et fragmentées ainsi que des distances de dispersion du pollen importantes, rendent ces espèces particulièrement sujettes à l'hybridation.

Chez les chênes, de nombreuses études ont montré que des événements d'hybridation avaient eu lieu à la suite de la migration postglaciaires (Petit *et al.*, 1997 ; Petit *et al.*, 2002a; Petit *et al.*, 1997) et que l'hybridation est un phénomène contemporain chez ces espèces (Bacilieri *et al.*, 1996b; Curtu *et al.*, 2007; Streiff *et al.*, 1999; Valbuena-Carabana *et al.*, 2007). Des résultats de croisements contrôlés (Rushton, 1977; Steinhoff, 1993) et des analyses en population montrent de plus que cette introgression semble unidirectionnelle dans le sens chêne pédonculé vers chêne sessile (croisement sessile (pollen) x pédonculé (ovule) conduisant à une introgression de gènes de pédonculé dans le sessile). L'hybridation chez ces espèces pourrait donc avoir des conséquences importantes ; en particulier, il a été suggéré que l'une de ces espèces puisse migrer par pollen en quelques générations dans un peuplement de l'autre espèce (Petit *et al.*, 2003). Néanmoins, cette hypothèse a récemment été remise en cause par Muir *et al.* qui proposent l'hypothèse de partage de polymorphisme ancestral pour expliquer les résultats de l'analyse de plusieurs populations européennes grâce à des marqueurs microsatellites (Muir & Schlotterer, 2005). Ces auteurs émettent trois hypothèses pouvant expliquer la faible, mais néanmoins significative, différenciation génétique entre les chênes sessiles et pédonculés :



- Un **faible flux de gènes** entre les deux espèces pourrait réduire la divergence génétique causée par la dérive génétique.
- La **sélection** à un nombre limité de locus pourrait maintenir l'intégrité des espèces malgré des **flux de gènes importants**. Cette hypothèse prédit une grande hétérogénéité génomique de la différenciation interspécifique. La majeure partie du génome ne présente pas de différenciation à cause de l'effet homogénéisant des flux de gènes, mais certaines régions génomiques, liées à des locus sélectionnés responsables du maintien des espèces, présenteraient un fort niveau de différenciation génétique.
- Un événement de spéciation récent pourrait causer une faible différenciation génétique du fait que les espèces partagent les mêmes allèles à cause d'un ancêtre commun récent. Cette hypothèse explique la faible différenciation par un **partage de polymorphisme ancestral** récent sans flux de gènes entre espèces.

Laissant de côté de nombreuses études démontrant l'existence de l'hybridation chez de nombreuses espèces de chênes, Muir *et al.* (Muir & Schlotterer, 2005) observent que le nombre d'hybrides est très faible dans leur échantillonnage et concluent grâce à l'analyse d'un seul jeu de données microsatellites que l'hypothèse du partage de polymorphisme ancien est la plus probable. Cette conclusion hâtive a poussé Lexer *et al.* (Lexer *et al.*, 2006) à publier une réponse rappelant l'ensemble des travaux non pris en compte par Muir, incluant l'analyse récente de la structure génomique de la différenciation interspécifique (Scotti-Saintagne *et al.*, 2004b) qui conforte d'avantage l'hypothèse de la sélection sur un nombre limité de locus permettant le maintien des espèces malgré un fort flux de gènes. Néanmoins, la controverse reste ouverte puisque Muir *et al.* publient simultanément une contre réponse "*Moving beyond single locus studies to characterize hybridization between oaks*" (Muir & Schlotterer, 2006), sans même tenir compte de la faiblesse de leur propre échantillonnage (un nombre limité d'individus échantillonnées dans plusieurs population) et de leurs marqueurs génétiques (microsatellites). Il est pourtant essentiel d'adopter une méthode comparative permettant d'incorporer le maximum d'informations disponibles pour statuer sur l'existence et le rôle évolutif potentiel de l'hybridation dans l'évolution des espèces (Arnold, 2006). Ce type de débat accompagne l'histoire de la biologie évolutive. En effet, différents chercheurs étudiant des espèces similaires peuvent conclure à différents processus pour expliquer une même observation. Ceci est particulièrement vrai en ce qui concerne les phénomènes d'hybridation (Arnold, 2006). Les zoologistes ont historiquement étudié les zones hybrides dans l'optique de comprendre le processus d'isolement reproducteur et de spéciation alors que les botanistes ont très tôt reconnu la porosité des barrières reproductives et le rôle de l'hybridation comme phénomène évolutif (Anderson & Stebbins, 1954). L'hybridation est donc depuis longtemps source de controverse entre zoologistes et botanistes, le débat chez les chênes européen en est un exemple très récent.

Des méthodes d'assignation génétique destinées à estimer la proportion d'hybrides en forêt, par échantillonnage exhaustif de parcelles, ont été utilisées à plusieurs reprises (Curtu *et al.*, 2007; Valbuena-Carabana *et al.*, 2007; Valbuena-Carabana *et al.*, 2005). Elles permettent une estimation plus fiable du phénomène d'hybridation que dans un échantillonnage classique en population, où seule une partie des arbres sont analysés. Dans les deux cas, un pourcentage non négligeable d'hybrides a été estimé. Seule une espèce éloignée, *Quercus cerris*, n'a montré aucun hybride dans une parcelle analysée contenant quatre autres espèces proches qui s'hybrident (*Quercus robur*, *Q. petraea*, *Q. pubescens* et *Q. frainetto* ; Curtu *et al.*, 2007). De plus, les hybrides retrouvés ne semblaient pas limité à des hybrides de première génération. Le taux d'hybrides semble être variable entre parcelles, montrant que des facteurs liés aux caractéristiques des parcelles pourraient influencer la dynamique d'hybridation (Valbuena-Carabana *et al.*, 2007). Ces analyses montrent que l'hybridation n'est pas un phénomène si rare, rendant l'hypothèse de partage de polymorphisme ancestral (Muir & Schlotterer, 2005) peu probable pour expliquer les résultats observés (Lexer *et al.*, 2006 ; Arnold, 2006).

Ces nouvelles méthodes incluent l'utilisation de microsatellites et de méthodes d'assignation dans le but d'identifier au niveau individuel le statut spécifique ou le caractère hybride des individus. Néanmoins, l'utilisation de ces méthodes doit se faire avec certaines précautions, spécialement lorsqu'il s'agit d'estimer le taux d'hybridation entre espèces. Un biais d'échantillonnage peut fortement fausser une telle estimation dans le cas où les individus ne sont pas échantillonnés au hasard ou exhaustivement. D'autre part, il faut avoir échantillonné un nombre suffisant d'individus par espèces, sans quoi l'assignation et l'analyse correspondante ignorera une partie des espèces présentes. Par exemple, dans la parcelle de Briouant (Haute Garonne) comprenant quatre espèces de chênes, nous avons échantillonné quelques arbres dont la morphologie foliaire était typique du chêne sessile. Néanmoins, lorsque cette parcelle est analysée seule, cette espèce n'est pas correctement identifiée par assignation car elle est présente en effectif trop réduit (voir chapitre 3 : nombre d'individus minimum pour l'assignation génétique). Il est nécessaire d'analyser cette parcelle conjointement avec une autre qui contient un nombre plus important d'individus de chêne sessile pour rendre compte de la composition en espèces de la parcelle de Briouant. Dans ce cas, l'interprétation était facilitée par la présence évidente de chênes sessiles typiques. Néanmoins, dans le cas où une espèce est présente sous forme d'hybrides uniquement, sa présence sera très difficile à percevoir à l'œil. Il est alors important d'inclure plusieurs parcelles comprenant le maximum d'espèces afin d'identifier les compositions exactes des peuplements. Ceci est particulièrement important chez les arbres qui sont caractérisés par des flux de pollen se propageant sur de longues distances.

Le but de ce chapitre est d'analyser conjointement plusieurs populations et parcelles comprenant plusieurs espèces de chênes, l'ensemble fournissant des effectifs suffisants pour les quatre espèces étudiées. Une assignation génétique permet d'identifier les espèces et les hybrides composant chacun de ces peuplements. Le taux d'hybrides est comparé ; les différences en fonction de la stratégie d'échantillonnage adoptée sont étudiées. Une analyse plus fine des processus d'introgession est enfin menée dans les

parcelles échantillonnées de manière exhaustive afin d'étudier plus précisément la dynamique d'introggression entre ces espèces.

## 4.2. Matériels et méthodes

### 4.2.1. Echantillonnages

#### 4.2.1.1. Populations

- Aurignac

90 chênes comprenant environ 30 individus par espèce (pédonculé, sessile et pubescent) sont inclus dans l'analyse.

- Transect pyrénéen

12 populations de chênes ont été échantillonnées pour 288 individus (moyenne par population : 24, min = 10, max = 34) le long de deux vallées pyrénéennes (Luz et Ossau). Les populations comprennent un mélange d'espèces mais l'échantillonnage concerne seulement une partie des chênes sessiles. Ces populations ont été analysées par Florian Alberto pour 16 microsatellites (Alberto et al., en préparation). Nous n'utiliserons ici que les 10 locus communs à toutes les populations.

- Population ONF

Neuf populations de forêts domaniales ont été échantillonnées pour 526 chênes (moyenne par population = 58 individus, min = 47, max = 82). Les peuplements comprennent des mélanges de chêne pédonculé et chêne sessile en différentes proportions. Le génotypage des individus pour les 10 microsatellites ont été réalisé par Erwan Guichoux.

#### 4.2.1.2. Parcelles

- Briouant

Nous avons échantillonné tous les chênes adultes (dbh>50mm) d'une portion du bois de Briouant. L'échantillonnage a permis d'obtenir des feuilles pour 834 chênes.

- Petite Charnie

276 chênes précédemment analysés par Réjane Streiff (Streiff *et al.*, 1998) pour six microsatellites ont été génotypés pour 10 microsatellites (Jessica Lavabre). Ils ont été inclus dans l'analyse afin d'augmenter l'effectif des chênes sessiles.

- La Paguères

Une parcelle située en Haute-Garonne, la Paguères, a fait l'objet d'un échantillonnage pour 220 chênes, génotypés à 10 locus microsatellites par Jessica Lavabre (Lavabre, 2005).

### 4.2.2. Génotypage

L'ADN a été extrait et 10 microsatellites ont été analysés par une méthode de PCR multiplex (Lepais *et al.*, 2006b).

### 4.2.3. Assignment des individus

Le logiciel *Structure* version 2.1 (Falush *et al.*, 2003; Pritchard *et al.*, 2000) est utilisé sur l'ensemble des populations. Les modèles *Admixture* et *Correlated Allele Frequency among Populations* sont utilisés avec 100000 itérations pour le temps de chauffe (Burnin Period) et 500000 répétitions MCMC après le temps de chauffe. Le nombre de population  $K$  est fixé à 4 afin de refléter la présence de quatre espèces dans le jeu de données. Cette valeur de  $K=4$  aboutissait e effet à la solution la plus vraisemblable selon la démarche indiquée dans le manuel de *Structure* (résultats non présentés). Dix analyses successives sont réalisées afin de s'assurer de la stabilité des résultats ; la moyenne des résultats obtenus pour ces répétitions est calculée pour chaque individu. Les individus sont ensuite classés dans une des classes d'hybrides ou d'espèce suivant les seuils déterminés lors des simulations (voir chapitre 3). Les individus ayant une probabilité supérieure à 0.90 d'appartenir à un cluster sont déclarés membres de l'espèce correspondante. Les hybrides correspondent aux individus qui ont un score inférieur à 0.90 pour tous les clusters. Les deux clusters donnant les meilleurs scores permettent de déterminer les deux espèces ayant engendré l'hybride. Enfin, quelques individus présentant une valeur inférieure à 0.90 pour un cluster mais des valeurs inférieures à 0.10 pour les trois autres clusters sont assignés à l'espèce correspondant au cluster majoritaire. Le nombre d'individus appartenant à chaque espèce ou chaque classe d'hybrides est reporté pour chaque population.

### 4.2.4. Analyse phylogénétique

Afin d'avoir une vision globale des relations génétiques entre les espèces présentes dans les différents peuplements et leurs hybrides, j'ai utilisé une méthode basée sur les distances génétiques. Les groupes sont constitués à partir des assignations de *Structure*. Chaque groupe contient un minimum de 10 individus assignés à une même espèce ou classe d'hybride et échantillonnés dans la même parcelle ou population. J'ai utilisé le logiciel *Populations* (version 1.2.28 ; Langella O., 1999, CNRS UPR9034; <http://bioinformatics.org/~tryphon/populations/>) pour calculer les distances génétiques de Cavalli-Sforza et Edwards ( $D_S$  ; Cavalli-Sforza & Edwards, 1967) entre chaque paire de groupes. Cette matrice de distance est utilisée pour construire un arbre par la méthode de Neighbor Joining avec 1000 permutations des locus afin d'estimer les valeurs de bootstrap permettant d'apprécier la fiabilité des nœuds de l'arbre. Enfin, l'arbre non enraciné est tracé grâce au package APE ("Analyses of Phylogenetics and Evolution", (Paradis *et al.*, 2004) du logiciel R.

### 4.2.5. Caractérisation de l'hybridation dans les parcelles

#### 4.2.5.1. Participation des espèces à l'hybridation

Pour chaque parcelle échantillonnée de façon exhaustive (Briouant, Petite Charnie) ou continue (Paguères), j'ai calculé la moyenne des assignations individuelles de *Structure* afin de rendre compte de la proportion des espèces en présence dans chaque population. J'ai comparé la proportion de chaque cluster parmi les arbres assignés à une espèce (quelle que soit son espèce) et parmi les arbres assignés comme hybrides. Cette analyse a pour objectif de mettre en évidence des différences éventuelles de

représentativité des espèces (1) entre les arbres d'espèce pure et (2) entre les arbres hybrides. Une comparaison de moyenne a été effectuée dans chaque parcelle (intra-parcelle), afin de tester une différence dans la composition génomique globale en espèce des arbres hybrides et non hybrides.

#### 4.2.5.2. Quantification relative des hybrides

J'ai adopté une approche probabiliste permettant de tester l'abondance ou la rareté d'une catégorie donnée d'hybrides en fonction des effectifs des deux espèces parentales présentes dans la parcelle. Le nombre d'hybrides attendu est calculé à partir des fréquences des quatre espèces sous l'hypothèse de panmixie. Sous cette hypothèse la fréquence attendue en hybrides de première génération est égale au produit des fréquences des deux espèces :

$$f_{F1} = f_{Esp1} \times f_{Esp2}$$

La fréquence des hybrides de seconde génération est obtenue en multipliant la fréquence des hybrides de première génération et la fréquence d'une des deux espèces parentale :

$$f_{bc\_Esp1} = f_{F1} \times f_{Esp1}$$

et

$$f_{bc\_Esp2} = f_{F1} \times f_{Esp2}$$

La fréquence globale des hybrides est la somme des fréquences des hybrides de première et de seconde génération :

$$f_{hyb} = f_{F1} + f_{bc\_Esp1} + f_{bc\_Esp2}$$

Le nombre d'hybrides attendu entre chaque couple d'espèces est ainsi calculé à partir des fréquences des espèces présentes dans chaque parcelle étudiée. L'écart entre le nombre d'hybrides attendu et observé pour chaque parcelle est testé par un test de Fisher exact.

#### 4.2.5.3. Dynamique d'introgession

- Distribution des valeurs d'admixture

Comme l'ont montré les résultats de Structure avec les hybrides obtenus par simulation (chapitre 3), l'assignation des individus au-delà de la première génération d'hybrides est impossible avec la résolution dont je dispose. Une analyse plus globale, prenant en compte l'ensemble des individus assignés à une même classe hybride, devrait permettre d'étudier les éventuels croisements préférentiels entre espèces. J'ai comparé les distributions des valeurs des clusters pour chaque type d'hybride en incluant les différents F1 et rétrocroisements obtenus par simulation ainsi que les hybrides des différentes populations. Cette analyse a deux objectifs : (1) comparer les distributions obtenues pour les classes d'hybrides simulées et observées en population afin de caractériser la nature des hybrides présents dans les parcelles; (2) tester une différence de distribution des valeurs des clusters pour les individus hybrides de chaque parcelle pour mettre en évidence d'éventuelles différences dans la dynamique d'introgession selon les lieux.

- Analyse d'admixture

Outre les résultats d'assignation de Structure, j'ai utilisé directement les génotypes des individus afin d'estimer la participation relative des espèces dans la constitution des hybrides. Une méthode de maximum de vraisemblance (Wang, 2003) permet d'estimer les paramètres d'un modèle d'admixture simple dans lequel au moins deux populations divergent d'une population ancestrale, évoluent indépendamment pendant un certain nombre de générations, avant de se retrouver en contact et de former une population hybride (**Figure 20**). Après l'événement d'hybridation, les populations parentales et hybrides évoluent indépendamment pendant un certain nombre de générations avant d'être échantillonnées. Ce modèle suppose que la structure génétique est principalement modelée par l'hybridation et la dérive. Cette méthode implémentée dans le logiciel LEADMIX permet, entre autres paramètres, d'estimer la proportion de chaque espèce ayant contribué à la formation de la population hybride ( $p1$  ;  $p2 = 1 - p1$  dans le cas de deux espèces). Chaque classe d'hybride dans chaque population est soumise à cette analyse d'admixture, ce qui permet d'estimer la contribution génétique relative de chacune des deux espèces parentales. Chaque espèce est représentée par 65 individus assignés avec une haute probabilité par Structure. Les populations hybrides analysées correspondent aux classes comprenant au moins 10 individus dans une population. Ainsi, la proportion d'admixture ( $p1$ ) est estimée dans 10 populations hybrides :

- Briouant : hybrides pédonculé x sessile (N = 27), pédonculé x pubescent (N = 38), pédonculé x tauzin (N = 45), sessile x tauzin (N = 41) et pubescent x tauzin (N = 85)
- Petite Charnie : hybrides pédonculé x sessile (N = 15), pédonculé x pubescent (N = 13) et pédonculé x tauzin (N = 13)
- Paguères : hybrides pédonculé x sessile (N = 15), et pédonculé x pubescent (N = 18).

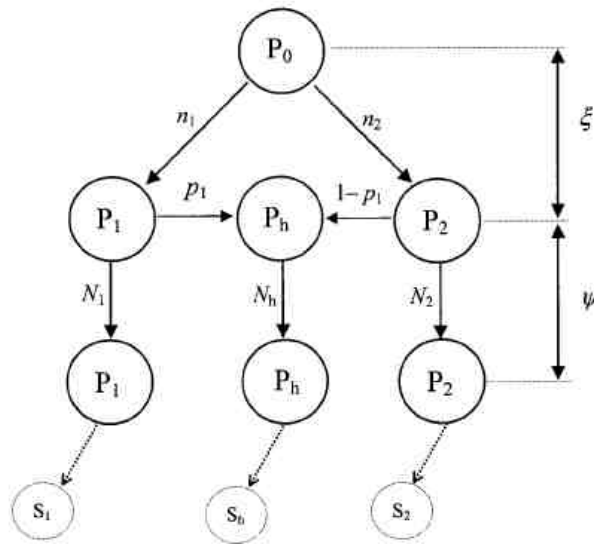


Figure 20 : Schéma du modèle implémenté dans LEADMIX (d'après Wang, 2003) et des paramètres estimés : il y a  $(\xi + \psi)$  générations une population ancestrale  $P_0$  se sépare en deux populations  $P_1$  et  $P_2$  qui évoluent indépendamment pendant  $\xi$  générations avant de former, il y a  $\psi$  générations, une population hybride  $P_h$  qui combine une proportion  $p_1$  de gène provenant de la population  $P_1$  et une proportion  $1-p_1$  de gènes provenant de la population 2. Ces trois populations évoluent indépendamment pendant  $\psi$  générations après l'événement d'admixture (hybridation), un échantillonnage de chaque population ( $S_1$ ,  $S_h$ , et  $S_2$ ) est alors effectué. L'analyse de marqueurs génétiques permet d'estimer chaque paramètre du modèle dont le coefficient d'admixture  $p_1$  ainsi que la dérive génétique ( $n_1$ ,  $n_2$ ,  $N_1$ ,  $N_h$ ,  $N_2$ ) subie par chaque population.

- Introgression et abondance des espèces

La relation entre les caractéristiques génétiques des hybrides et les effectifs des espèces parentales est testée par corrélation. Deux estimateurs de la caractéristique des hybrides sont pris en compte : la moyenne de la valeur du cluster de l'espèce parentale 1 déterminé par *Structure* (15 groupes d'hybrides) ainsi que la participation de l'espèce parentale 1 ( $p_1$  estimé sur 10 groupes d'hybrides). L'effectif relatif de l'espèce parentale 1 est calculé en rapportant l'effectif de l'espèce 1 dans la parcelle à la somme des effectifs des deux espèces parentales présentes dans la parcelle. Une analyse de corrélation est utilisée pour examiner les relations entre les fréquences relatives des espèces et leur implication dans la composition génétique des hybrides.

### 4.3. Résultats

#### 4.3.1. Assignment génétique des individus

##### 4.3.1.1. Pourcentage d'hybrides général

Les 10 répétitions de l'analyse d'assignment génétique pour un nombre de population  $K = 4$  donnent des résultats très stables. Entre ces 10 répétitions, les coefficients d'admixture individuels ont un écart type moyen de 0.0006 (max : 0.007). De plus, les quatre populations imposées ( $K = 4$ ) correspondent bien aux quatre espèces de chênes présentes dans les populations (Figure 21).

Les hybrides représentent de 15 à 30 % des individus adultes analysés (Tableau 8). Les deux parcelles échantillonnées exhaustivement présentent 20 et 30% d'hybrides (Tableau 8). L'échantillonnage non

exhaustif de la Paguères contient 24% d'hybrides adultes. Les échantillonnages en populations (non exhaustifs) ont des taux légèrement plus faibles allant de 15 à 20 %. Ce qui pourrait s'expliquer par la sélection d'individus morphologiquement caractéristiques d'une espèce lors de l'échantillonnage (biais d'échantillonnage).

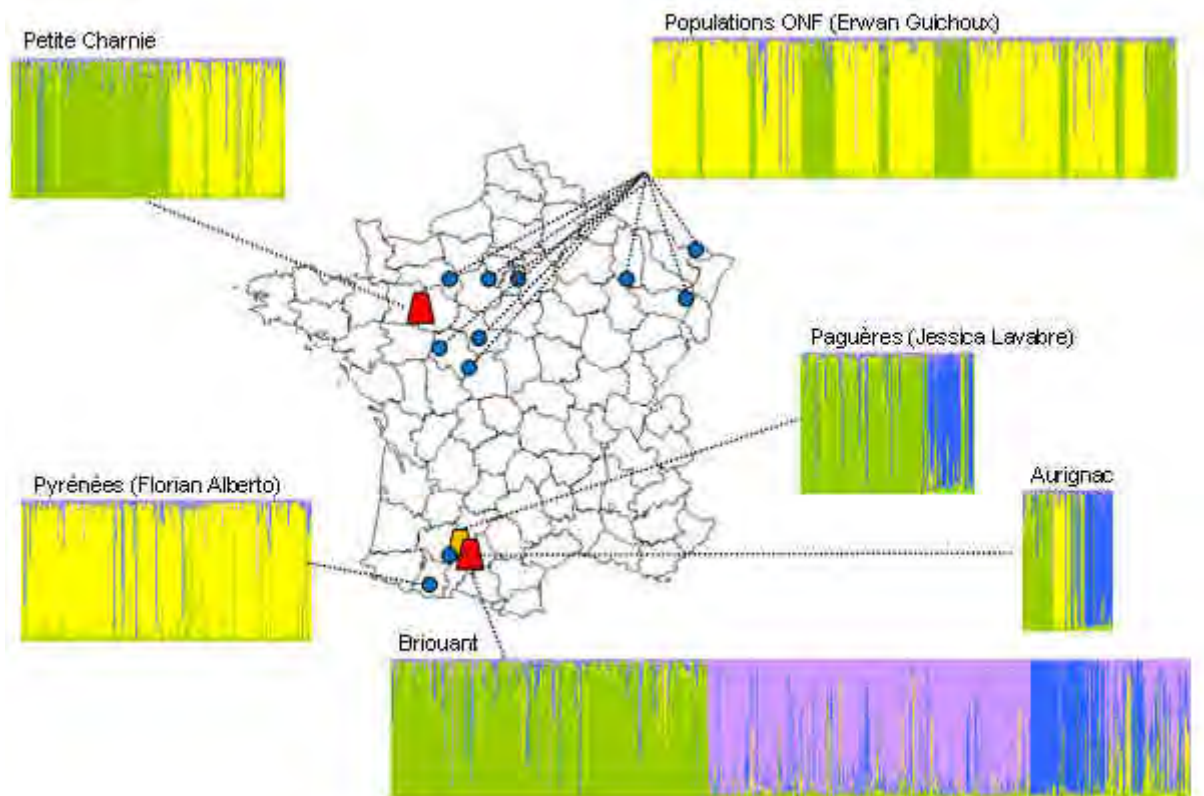


Figure 21 : Clustering génétique par STRUCTURE. Chaque barre verticale représente un individu dont la composition génétique est représentée par les différentes couleurs correspondant aux quatre clusters (vert : pédonculé, jaune : sessile, bleu : pubescent et violet : tauzin).

Tableau 8 : Pourcentage d'hybrides assignés génétiquement comme appartenant à une espèce ou comme hybride dans les différentes populations et parcelles.

Echantillonnage	Type	N	Pure	Hybride
Briouant	parcelle, exhaustif	817	70.38%	29.62%
Petite Charnie	parcelle, exhaustif	271	79.70%	20.30%
Paguères	parcelle, non exhaustif	174	76.44%	23.56%
Aurignac	Populations	87	85.06%	14.94%
Transect pyrénéen	Populations	288	79.51%	20.49%
Populations ONF	Populations	526	84.22%	15.78%

N : nombre d'individus, Pure : pourcentage d'individus assignés à une espèce et Hybride : pourcentage d'individus assignés comme hybride.

#### 4.3.1.2. Aurignac

Cet échantillonnage ponctuel dans la région des parcelles Briouant et Paguères a été effectué en privilégiant les individus au phénotype typique des espèces pures. On retrouve néanmoins un taux d'hybrides de l'ordre de 15% (Tableau 8). Environ 30 chênes sont assignés aux espèces pédonculées et



pubescentes et seulement 16 à l'espèce sessile (Tableau 9). Cette dernière espèce est particulièrement présente chez les hybrides avec 4 chênes hybrides pédonculé x sessile et 4 chênes hybrides sessile x pubescent (Tableau 9). On retrouve la présence du chêne tauzin à l'état d'hybride (4 individus) alors qu'aucun chêne tauzin pur n'a été échantillonné dans ces populations.

Tableau 9 : Détail des résultats d'assignation génétique dans les différents peuplements

Population	N	Ped	Ses	Pub	Toz	hyb PedSes	hyb PedPub	hyb PedToz	hyb SesPub	hyb SesToz	hyb PubToz
Briouant	817	241 (29.5%)	5 (0.6%)	83 (10.2%)	246 (30.1%)	27 (3.3%)	38 (4.7%)	45 (5.5%)	6 (0.7%)	41 (5.0%)	85 (10.4%)
Petite Charmie	271	130 (48.0%)	85 (31.4%)	1 (0.4%)	-	15 (5.5%)	13 (4.8%)	13 (4.8%)	8 (3.0%)	6 (2.2%)	-
Paguères	174	101 (58.0%)	1 (0.6%)	31 (17.8%)	-	15 (8.6%)	18 (10.3%)	-	3 (1.7%)	-	5 (2.9%)
Aurignac	87	28 (32.2%)	16 (18.4%)	30 (34.5%)	-	4 (4.6%)	1 (1.1%)	2 (2.3%)	4 (4.6%)	-	2 (2.3%)
Transect pyrénéen	288	1 (0.3%)	224 (77.8%)	4 (1.4%)	-	30 (10.4%)	3 (1.0%)	1 (0.3%)	9 (3.1%)	15 (5.2%)	1 (0.3%)
Populations ONF	526	118 (22.4%)	323 (61.4%)	2 (0.4%)	-	23 (4.4%)	12 (2.3%)	3 (0.6%)	23 (4.4%)	22 (4.2%)	-

Les chiffres représentent les effectifs de la classe correspondant à la colonne et entre parenthèse le pourcentage par rapport au nombre d'individus dans la parcelle. N : nombre d'individus, Ped : pédonculé, Ses : sessile, Pub : pubescent, Toz : tauzin et hyb XY : hybride entre les deux espèces X et Y.

#### 4.3.1.3. Populations pyrénéennes

Plusieurs populations pyrénéennes ont été échantillonnées en cherchant à prélever seulement des arbres à morphologie typique du chêne sessile. L'échantillonnage se révèle efficace car cette espèce représente en effet la majorité des arbres (77.8%, Tableau 9). On retrouve tout de même un chêne pédonculé et 4 chênes pubescents. Environ 20% des arbres adultes sont des hybrides (Tableau 8).

Tableau 10 : Pourcentage des différentes espèces et hybrides dans les populations du transect pyrénéen. Le nom des populations indique la vallée (L : vallée de Luz, et O : vallée d'Ossau) ainsi que leur altitude (altitude = nombre x 100).

Pop	N	Pure	Hybride	Ped	Ses	Pub	Toz	hyb PedSes	hyb PedPub	hyb PedToz	hyb SesPub	hyb SesToz	hyb PubToz
L1	20	80.0	20.0	0.0	80.0	0.0	0.0	10.0	5.0	0.0	5.0	0.0	0.0
L3	24	66.7	<b>33.3</b>	0.0	66.7	0.0	0.0	16.7	0.0	4.2	0.0	<b>12.5</b>	0.0
L4	40	95.0	5.0	0.0	95.0	0.0	0.0	2.5	0.0	0.0	0.0	2.5	0.0
L8	22	86.4	13.6	0.0	77.3	9.1	0.0	4.5	4.5	0.0	0.0	4.5	0.0
L12	32	81.3	18.8	0.0	75.0	6.3	0.0	9.4	0.0	0.0	6.3	3.1	0.0
L14	23	69.6	<b>30.4</b>	0.0	69.6	0.0	0.0	8.7	0.0	0.0	<b>8.7</b>	<b>13.0</b>	0.0
L16	23	56.5	<b>43.5</b>	0.0	56.5	0.0	0.0	<b>30.4</b>	0.0	0.0	4.3	4.3	4.3
O1	28	82.1	17.9	0.0	82.1	0.0	0.0	14.3	0.0	0.0	0.0	3.6	0.0
O4	22	95.5	4.5	0.0	95.5	0.0	0.0	0.0	0.0	0.0	0.0	4.5	0.0
O8	10	80.0	20.0	10.0	70.0	0.0	0.0	10.0	0.0	0.0	0.0	10.0	0.0
O12	34	79.4	20.6	0.0	79.4	0.0	0.0	8.8	0.0	0.0	5.9	5.9	0.0
O16	10	60.0	<b>40.0</b>	0.0	60.0	0.0	0.0	20.0	10.0	0.0	10.0	0.0	0.0
Total	288	79.5	20.5	0.3	77.8	1.4	0.0	10.4	1.0	0.3	3.1	5.2	0.3

N : nombre d'individus par population. Les autres chiffres correspondent à des pourcentages.

Sur l'ensemble des populations, le taux d'hybride varie de 4.5 à 43.5% (Tableau 10). Il semble y avoir un plus grand nombre d'hybrides pour les populations d'altitude (Pearson's product-moment correlation,  $t = 2.48$ ,  $df = 10$ ,  $p\text{-value} = 0.03$ ,  $r = 0.62$ ). Néanmoins, cette corrélation pourrait être liée à un biais d'échantillonnage. En altitude, les effectifs réduits en chêne pourraient conduire à échantillonner davantage d'individus à la morphologie atypique du chêne sessile comparé aux populations de faible altitude où le grand nombre d'individus permet de choisir des individus typique de la morphologie sessile. Les hybrides pédonculé x sessile sont les plus représentés (10.4% de l'ensemble des individus). Les hybrides sessile x tauzin sont aussi relativement fréquents à toutes les altitudes alors que les hybrides sessile x pubescent se retrouvent principalement dans les populations de haute altitude (Tableau 10). Les hybrides pédonculé x pubescent et pédonculé x tauzin sont rares et cantonnés aux populations de faible altitude de la vallée de Luz.

#### 4.3.1.4. Populations ONF

Cet échantillonnage correspond à diverses forêts domaniales comprenant du chêne pédonculé et sessile, gérées par l'Office National des Forêts. Le taux d'hybrides est d'environ 16% (Tableau 8), mais avec un biais d'échantillonnage privilégiant les formes typiques. Outre les chênes pédonculé (22.4%) et sessile (61.4%), on retrouve 2 chênes pubescents (tableau 2). Les hybrides concernent ces trois espèces (11.1%) ainsi que quelques hybrides faisant intervenir l'espèce tauzin (4.8%). Alors que seulement deux individus pubescents ont été échantillonnés, les hybrides sessile x pubescent et pédonculé x pubescent sont relativement nombreux (4.4% et 2.3% respectivement).

Tableau 11 : Pourcentage des différentes espèces et hybrides dans les populations ONF.

Pop	N	Pure	Hybride	Ped	Ses	Pub	Toz	hyb PedSes	hyb PedPub	hyb PedToz	hyb SesPub	hyb SesToz	hyb PubToz
REN	52	88.5	11.5	9.6	78.8	0.0	0.0	1.9	0.0	0.0	1.9	<b>7.7</b>	0.0
LOC	47	80.9	<b>19.1</b>	2.1	78.7	0.0	0.0	4.3	0.0	0.0	<b>6.4</b>	<b>8.5</b>	0.0
SEN	53	88.7	11.3	5.7	83.0	0.0	0.0	3.8	0.0	3.8	3.8	0.0	0.0
RUS	54	87.0	13.0	13.0	74.1	0.0	0.0	0.0	3.7	1.9	1.9	<b>5.6</b>	0.0
CHA	82	86.6	13.4	37.8	48.8	0.0	0.0	<b>6.1</b>	3.7	0.0	2.4	1.2	0.0
FON	79	81.0	<b>19.0</b>	35.4	43.0	2.5	0.0	2.5	<b>5.1</b>	0.0	<b>7.6</b>	3.8	0.0
DAR	53	83.0	17.0	47.2	35.8	0.0	0.0	<b>11.3</b>	3.8	0.0	1.9	0.0	0.0
HAG	50	86.0	14.0	18.0	68.0	0.0	0.0	2.0	0.0	0.0	<b>8.0</b>	4.0	0.0
BIT	56	76.8	<b>23.2</b>	16.1	60.7	0.0	0.0	<b>7.1</b>	1.8	0.0	<b>5.4</b>	<b>8.9</b>	0.0
Total	526	84.3	15.7	22.4	61.4	0.4	0.0	4.4	2.3	0.6	4.4	4.2	0.0

N : nombre d'individus par population. Les autres chiffres correspondent à des pourcentages.

En fonction des populations, le pourcentage d'hybrides varie de 11.3% à 23.2% (Tableau 11). Les pourcentages de chaque type d'hybrides varient considérablement d'une population à l'autre. Ainsi les hybrides pédonculé x sessile représentent de 0 à 11.3% des individus des populations RUS et DAR respectivement. Sur l'ensemble des populations, les hybrides pédonculé x pubescents ou x tauzin sont rares alors que le chêne sessile semble s'hybrider plus fréquemment avec ces deux espèces. En effet, alors que les peuplements échantillonnés sont essentiellement constitués de chênes sessiles et pédonculés, on constate qu'il y a le même nombre d'hybrides pédonculé x sessile que d'hybrides sessile x pubescent et sessile x tauzin (4.4%, 4.4% et 4.2% au total).

#### 4.3.1.5. Briouant

Ce bois, où coexistent les quatre espèces, est caractérisé par une forte proportion d'hybrides chez les adultes (30%). Alors que les effectifs en chêne pédonculé et tauzin sont équilibrés (environ 240 arbres ; Tableau 9), les chênes pubescents sont plus rares (83 arbres) et le chêne sessile est présent sporadiquement avec seulement 5 individus de cette espèce. L'ensemble des couples d'espèces forment des hybrides adultes (Tableau 9). Le pourcentage des différentes classes d'hybrides montre que les hybrides chênes pubescents x tauzins sont les plus nombreux alors que les hybrides chênes sessiles x pubescents sont nettement plus rares (10.4% contre 0.7%).

#### 4.3.1.6. La Petite Charnie

La Petite Charnie est caractérisée par 20% d'hybrides parmi les adultes (Tableau 8). Cette parcelle est principalement composée de chênes pédonculé (131 individus) et sessile (78), un individu a été assigné à l'espèce pubescent (Tableau 9), espèce non décrite jusqu'à présent dans ce bois. Malgré le faible nombre de chênes pubescents et l'absence d'individus tauzin, on identifie une quantité non négligeable d'hybrides avec ces deux espèces (14.8% au total).

#### 4.3.1.7. *Paguères*

Ce bois présente un taux d'hybridation moyen de 24% (Tableau 8). Le peuplement adulte se compose essentiellement de chênes pédonculés (101 arbres) et de chênes pubescents (31 individus). L'échantillonnage non exhaustif de cette parcelle comprend un seul chêne sessile, néanmoins cette espèce intervient de façon non négligeable chez les hybrides (pédonculé x sessile et sessile x pubescent ; Tableau 9). Comme pour la Petite Charnie, alors qu'aucun chêne tauzin n'a été échantillonné, ni même décrit dans cette parcelle, on retrouve cinq hybrides pubescent x tauzin parmi les adultes.

#### 4.3.2. Relations génétiques entre espèces, hybrides et populations

Les individus assignés à une espèce ou une classe d'hybrides pour chaque parcelle ou population sont regroupés en autant de populations afin de calculer des distances génétiques entre chaque paire de populations. Ces distances sont utilisées pour tracer un arbre phylogénétique représentant les relations génétiques entre espèces, hybrides et populations (Figure 22).

La distance génétique entre populations appartenant à une même espèce (distance intra-spécifique) est globalement faible car ces populations sont groupées sur une même branche secondaire de l'arbre et sont séparées par des longueurs de branche relativement courtes (groupes d'espèces pures entourés par des traits pleins colorés ; Figure 22).

Les populations hybrides appartenant à une même classe (donc ayant les deux mêmes espèces parentales) se regroupent globalement dans une même partie de l'arbre. Cette observation générale peut être illustrée par les populations d'hybrides pédonculé x sessile qui sont toutes connectées à un nœud commun (groupe I, Figure 22).

Entre les branches correspondant aux espèces pédonculé et pubescent, on trouve les populations d'hybrides pédonculé x pubescent et pédonculé x tauzin qui sont connectées en divers points de la branche principale de l'arbre (groupe II, Figure 22). Les hybrides pédonculé x pubescent échantillonnés dans les populations ONF et à la Paguères dérivent d'une branche commune qui se connecte à la branche principale de l'arbre. Les groupes d'hybrides échantillonnés dans les parcelles et dont les effectifs sont plus importants, se regroupent par parcelle et non par classe d'hybride : les hybrides pédonculé x pubescent et pédonculé x tauzin de la Petite Charnie ont un nœud en commun tout comme ces mêmes hybrides à Briouant. Ce résultat confirme bien la présence de pubescents et tazuins sous forme d'hybrides à la Petite Charnie car ils se regroupent avec leurs équivalents présents dans d'autres parcelles ou populations.

Les hybrides sessile x tauzin des populations pyrénéennes et ONF semblent génétiquement proches (groupe III, Figure 22) et localisés entre les deux espèces parentales. Remarquons que cette classe d'hybrides à Briouant est séparée par une faible distance de l'espèce tauzin. De la même façon, les hybrides pubescent x tauzin de Briouant sont séparés par une faible distance génétique de l'espèce tauzin. La localisation de ces deux types d'hybrides à Briouant traduit bien leur composition génétique de ces hybrides qui semble liée à l'effectif important de cette parcelle en chênes tazuins.

Enfin, la seule population d'hybrides sessile x pubescent (ONF) se branche à proximité des populations de chênes sessiles.

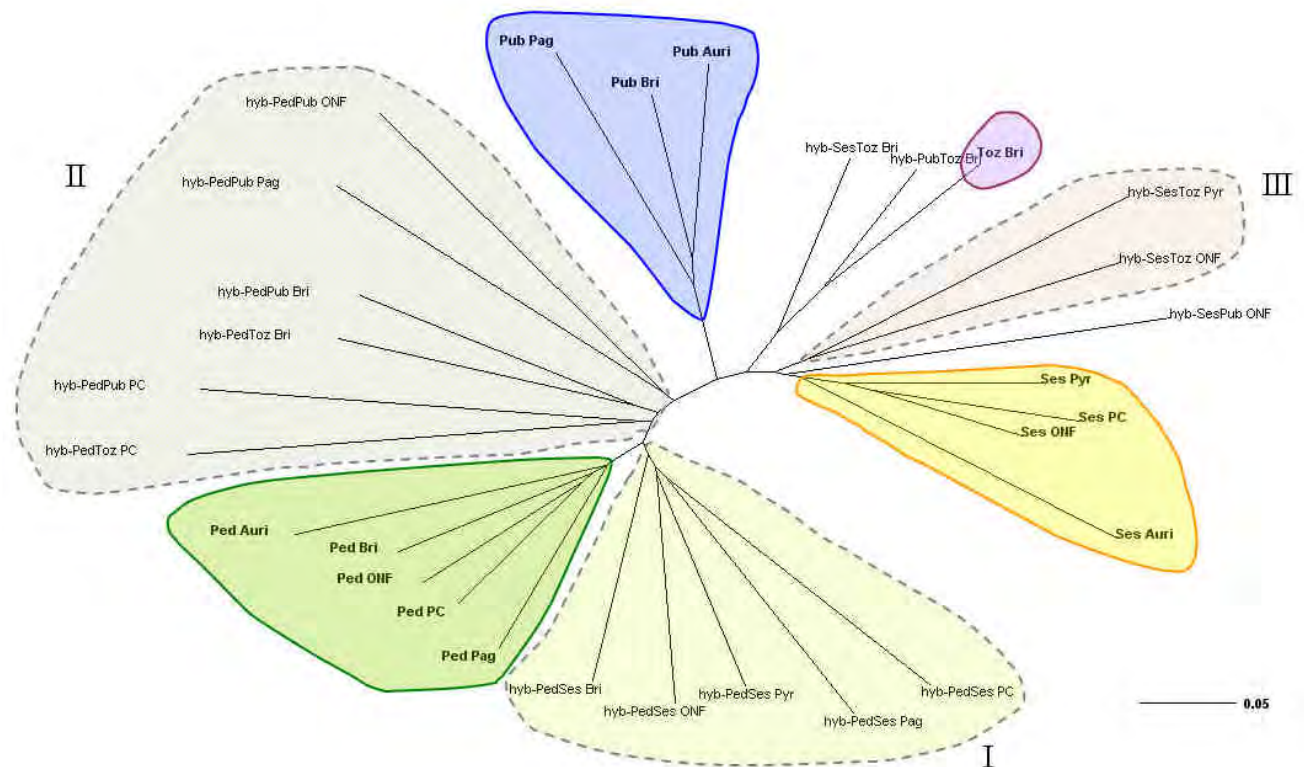


Figure 22 : Arbre phylogénétique (Neighbor Joining) basé sur la distance génétique de Cavalli Sforza et Edwards (Cavalli-Sforza & Edwards, 1967) entre les espèces et les hybrides identifiés par Structure. Légende : Ped : pédonculé, Ses : sessile, Pub : pubescent, Toz : tauzin, Bri : Briouant, PC : Petite Charnie, Pag : Paguères, Auri : populations de la région d'Aurignac, ONF : populations ONF, Pyr : populations pyrénéennes.

### 4.3.3. Caractéristiques génétiques des hybrides dans les parcelles

#### 4.3.3.1. Participation des espèces à l'hybridation

La comparaison de la composition moyenne en espèces entre les arbres appartenant à une espèce et les hybrides montre quelques différences entre les parcelles étudiées (Tableau 12).

A Briouant, le chêne pédonculé est moins représenté parmi les hybrides que parmi les arbres assignés à une espèce pure (Tableau 12). Inversement, les chênes pubescents et sessiles apparaissent plus représentés parmi les hybrides que parmi les individus appartenant à une espèce pure. La proportion de chênes tauzins est quant à elle statistiquement non différente entre la classe espèce pure et hybride.

On constate généralement les mêmes tendances à Paguères. Le chêne pédonculé est moins représenté parmi les hybrides que parmi les arbres assignés à une espèce pure (Tableau 12) alors que les chênes sessiles et tauzins sont plus représentés parmi les hybrides. L'espèce pubescente est également représentée chez les arbres purs et les arbres hybrides.

A la Petite Charnie les deux espèces majoritaires sont autant représentées parmi les arbres assignés à une espèce et les hybrides alors que les chênes pubescents et tauzins sont très rares au niveau des espèces pures mais significativement représentées au niveau des arbres hybrides (Tableau 12).

Tableau 12 : Comparaison des moyennes de chaque cluster de **Structure** entre les arbres d'espèce pure et les hybrides dans chaque parcelle.

Population	Espèce	Clust.Ped	Clust.Ses	Clust.Pub	Clust.Toz
Briouant	Pures	<b>0.41</b>	<b>0.02</b>	<b>0.15</b>	0.41
	Hybrides	<b>0.27</b>	<b>0.12</b>	<b>0.24</b>	0.37
	<i>p value</i>	<0.001	<0.001	<0.001	0.195
Petite Charnie	Pures	0.58	0.38	<b>0.03</b>	<b>0.02</b>
	Hybrides	0.47	0.32	<b>0.10</b>	<b>0.11</b>
	<i>p value</i>	0.118	0.355	<0.001	<0.001
Paguères	Pures	<b>0.72</b>	<b>0.03</b>	0.23	<b>0.02</b>
	Hybrides	<b>0.47</b>	<b>0.14</b>	0.32	<b>0.07</b>
	<i>p value</i>	<0.001	<0.001	0.168	<0.001

#### 4.3.3.2. Quantification relative des hybrides

- Briouant

La comparaison de la composition génétique globale des arbres d'espèces pures et des hybrides (Tableau 12) montre que le chêne pédonculé est sous-représenté chez les hybrides alors que les chênes sessiles et pubescents sont sur-représentés parmi les hybrides. Une fois les effectifs des espèces pris en compte (Figure 23), les hybrides pédonculé x sessile, sessile x tauzin et pubescent x tauzin sont plus nombreux qu'attendu sous l'hypothèse de panmixie. Les hybrides pédonculé x pubescent et sessile x pubescent ont des effectifs proches de ceux attendus en panmixie alors que les hybrides pédonculé x tauzins sont trois fois moins nombreux qu'attendu.

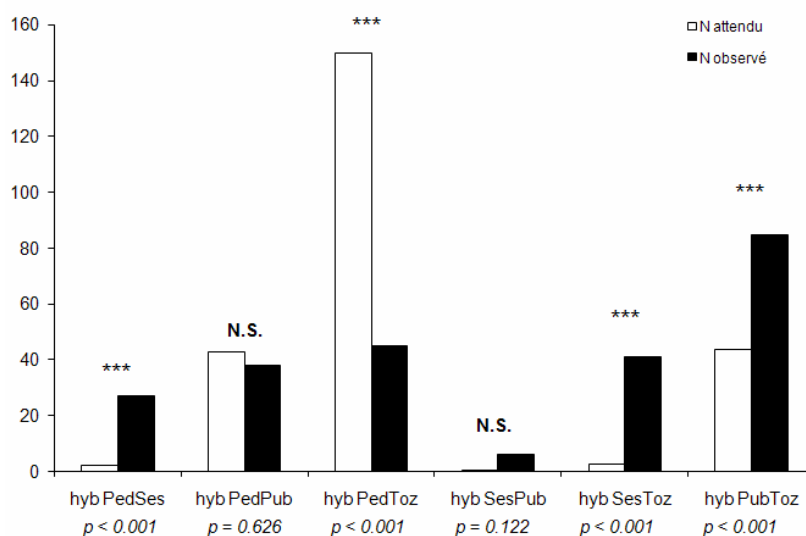


Figure 23 : Nombre d'hybrides attendus et observés à Briouant

- La Petite Charnie

Il n'y a pas de différence significative au niveau de la représentativité des deux espèces majoritaires entre les arbres purs et les hybrides (Tableau 12) alors que les espèces pubescentes et tauzines sont surreprésentées parmi les hybrides. Relativement au nombre d'arbre d'espèces pures, le taux d'hybrides pour les espèces rares dans cette parcelle est élevé (Figure 24) : les hybrides pédonculé x pubescent, pédonculé x tauzin, sessile x pubescent et sessile x tauzin sont en nombre non négligeable. D'autre part, les hybrides pédonculé x sessile sont moins nombreux qu'attendu sous l'hypothèse de croisement interspécifiques libres.

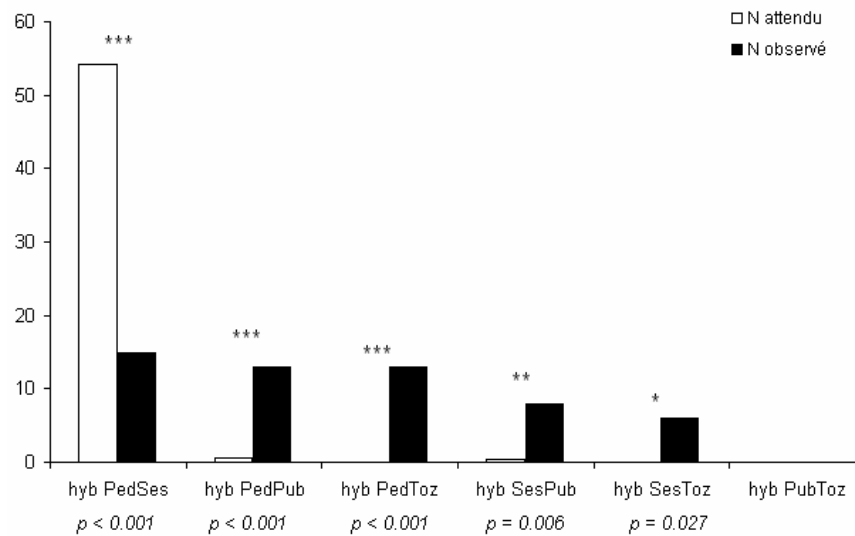


Figure 24 : Nombre d'hybrides attendus et observés à la Petite Charnie

- Paguères

Au niveau global, le chêne pédonculé est sous-représenté parmi les hybrides par rapport à la composition des arbres d'espèces pures (Tableau 12). Le chêne pubescent contribue de façon égale au pool hybride et pur alors que les chênes sessiles et tauzins sont en plus grande proportion parmi les hybrides. Rapporté aux effectifs des individus d'espèce pure, le nombre d'hybrides pédonculé x sessile est élevé dans cette parcelle (Figure 25), alors que les hybrides pédonculé x pubescent sont moins nombreux qu'attendu sous l'hypothèse de croisements aléatoires. La comparaison entre effectifs attendus et observés n'est pas significative pour les hybrides sessile x pubescent et pubescent x tauzin.

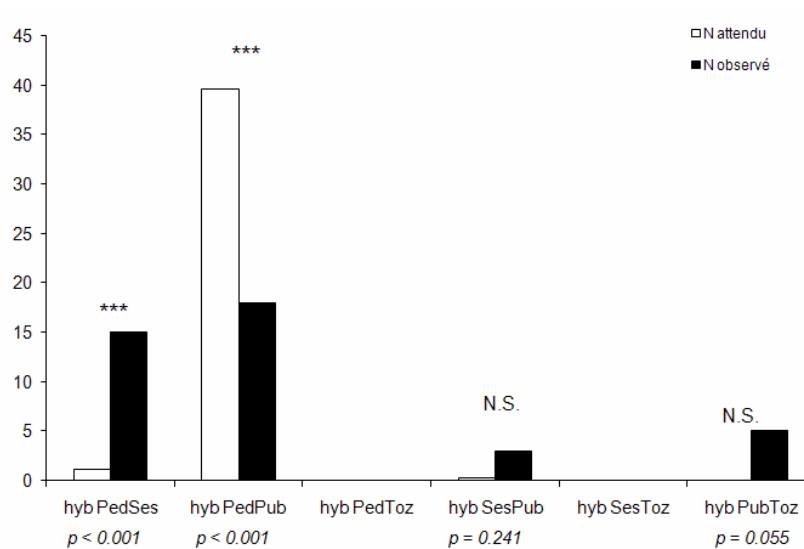


Figure 25 : Nombre d'hybrides attendus et observés à la Paguères

#### 4.3.3.3. Introgression et directionnalité de l'hybridation

- Distribution des valeurs d'admixture

Lorsque deux espèces sont présentes en effectif suffisant dans les populations, leurs coefficients d'admixture se répartissent de façon bimodale (Figure 26, cas des hybrides pédonculé x sessile à la Petite Charnie (A3), pédonculé x pubescent à Paguères (B2), pédonculé x tauzin et pubescent x tauzin à Briouant (C1 et F1)). A l'inverse, lorsqu'une espèce est absente ou très rare sous forme pure, on observe qu'elle est souvent représentée sous forme d'hybrides qui montrent des coefficients d'admixture proche de l'autre espèce (cas pédonculé x pubescent ou pédonculé x tauzin à la Petite Charnie (B3 ou C3) par exemple). Ces résultats semblent indiquer des effets probables de l'effectif des espèces pures dans les populations et de croisements préférentiels entre espèces. Ainsi, la comparaison des distributions des valeurs de cluster de Structure entre parcelle et pour les différents hybrides devrait permettre de comprendre les directions d'hybridation préférentielles.



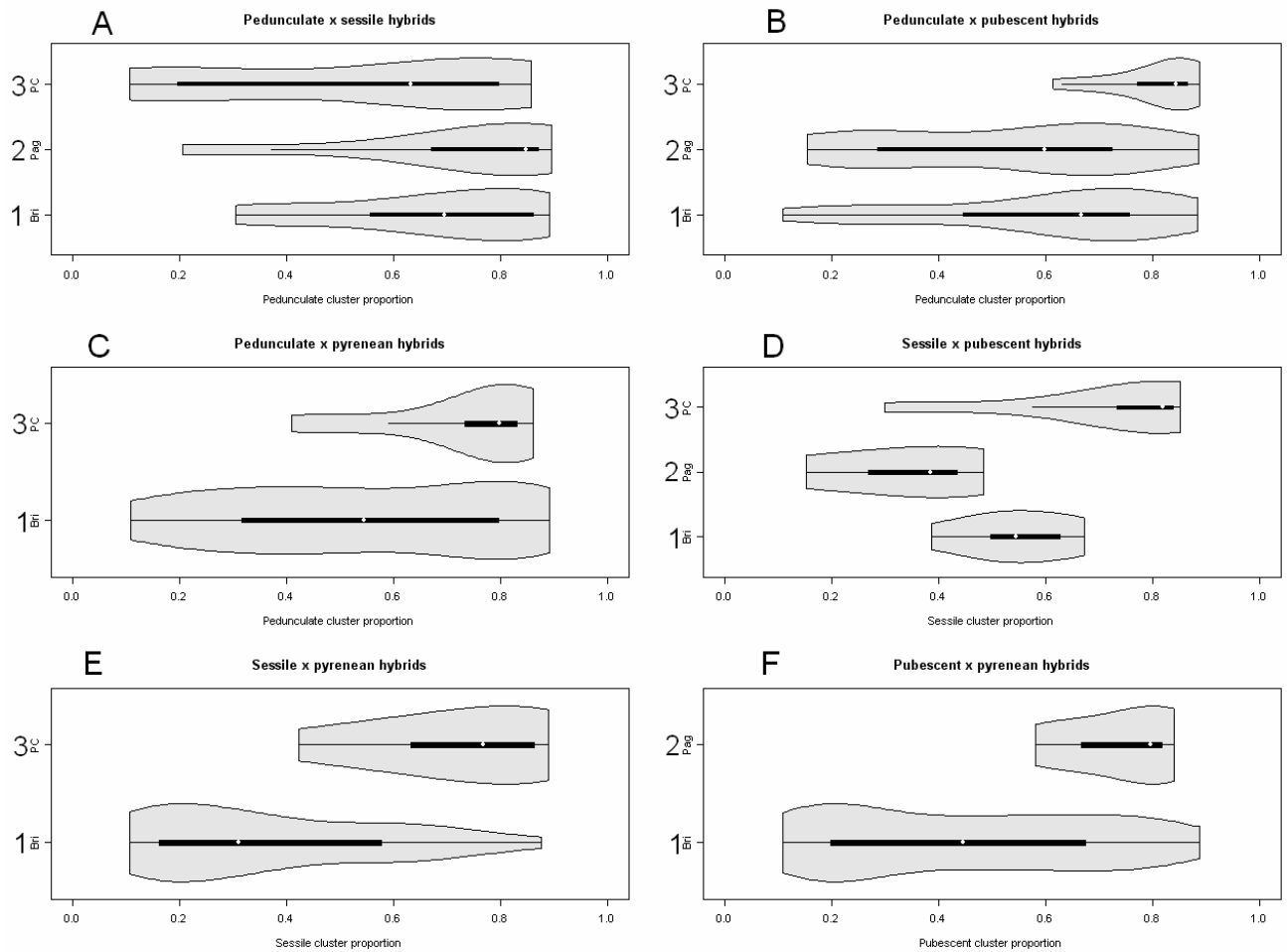


Figure 26 : Distribution des coefficients d'admixture des hybrides (pédonculé x sessile (A), pédonculé x pubescent (B), pédonculé x tauzin (C), sessile x pubescent (D), sessile x tauzin (E) et pubescent x tauzin (F)) dans chaque population (1 : Briouant, 2 : Paguères et 3 : Petite Charnie).

La simulation d'hybrides entre deux espèces permet d'obtenir une distribution des coefficients d'admixture (coefficient d'admixture = valeurs du cluster de *Structure* pour l'une des espèces parentales) de référence que l'on peut comparer à ceux obtenus chez les hybrides identifiés en population (Figure 27). Les distributions des hybrides issus de rétrocroisement (Figure 27, modalités 1 et 3) représentent des cas où l'hybridation s'effectue de façon stricte dans une seule direction (par exemple seule l'espèce 1 se croise avec les hybrides de première génération pour donner les rétrocroisements représentés dans la modalité 1, Figure 27). De façon similaire, la distribution des hybrides F1 (modalité 2, Figure 27) représente les cas où l'hybridation est limitée à la première génération d'hybridation ou bien s'effectue sans direction préférentielle à l'issue de la première génération (les hybrides ne tendent pas vers une espèce ou l'autre mais se répartissent uniformément entre les deux espèces).

La comparaison de la distribution des coefficients d'admixture pour les classes d'hybrides identifiés dans les différentes parcelles (Figure 27) met en évidence qu'il n'existe pas de croisements préférentiels systématiques entre espèces : les hybrides ont des caractéristiques différentes entre les parcelles étudiées.

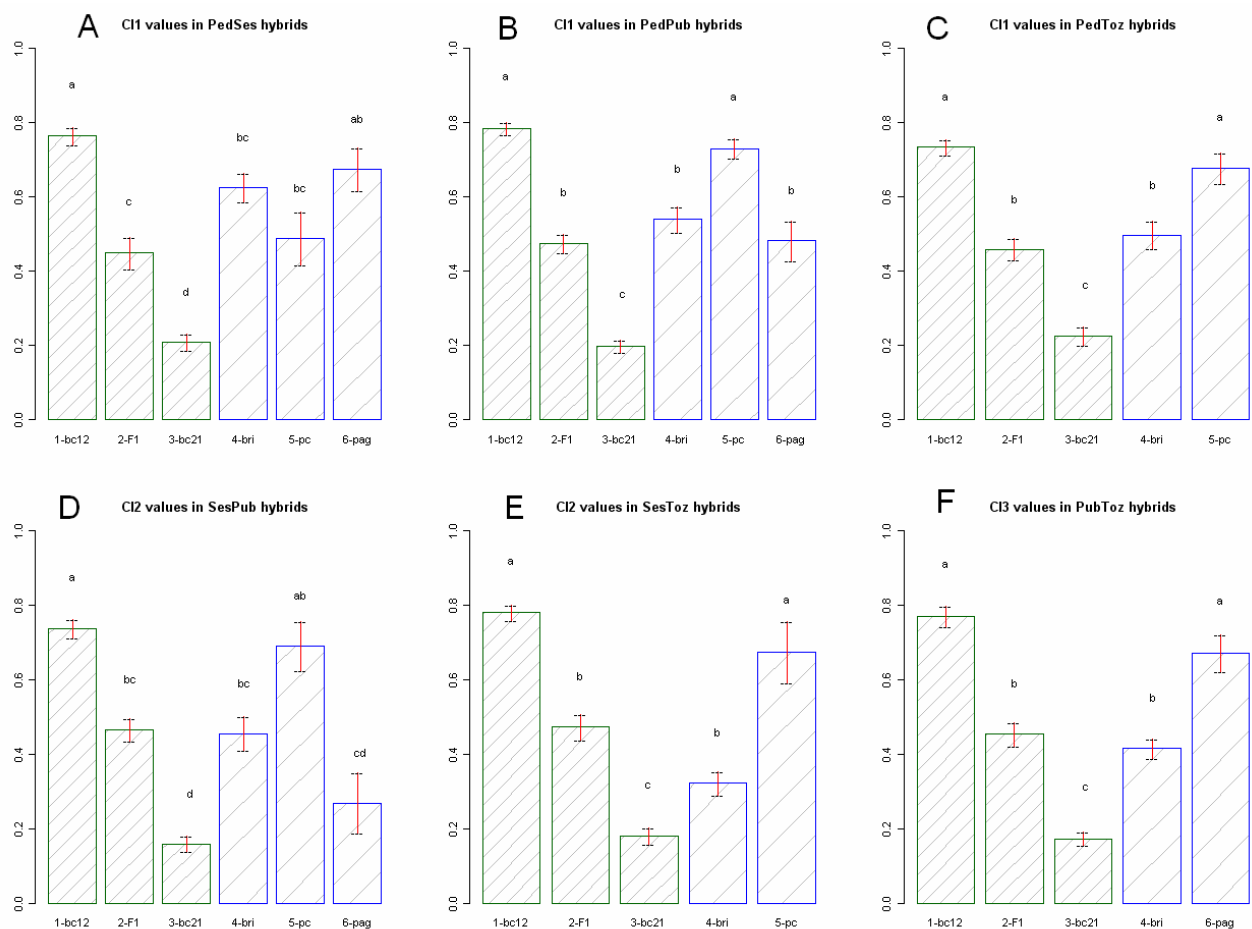


Figure 27 : Test de comparaison multiple non paramétrique de Behrens-Fisher entre les moyennes des coefficients d'admixture des différentes classes d'hybrides simulés (1 à 3, en vert) et d'hybrides entre parcelles (4 à 6, en bleu). A : hybrides pédonculé x sessile, B : hybrides pédonculé x pubescent, C : hybrides pédonculé x tauzin, D : hybrides sessile x pubescent, E : hybrides sessile x tauzin et F : hybrides pubescent x tauzin.

A Briouant (modalité 4, Figure 27), la plupart des hybrides ont des valeurs d'admixture moyennes intermédiaires, comparables à des hybrides F1 ou à un mélange équilibré de rétrocroisements issu des deux espèces parentales (cas des hybrides pédonculé x pubescent, pédonculé x tauzin, sessile x pubescent et pubescent x tauzin, modalité 4 des graphiques B, C, D et F, Figure 27). Seuls les hybrides pédonculé x sessile et sessile x tauzin montrent des caractéristiques d'assignation légèrement plus proche d'une des espèces parentales. C'est le cas des hybrides pédonculé x sessile qui tendent vers le type pédonculé sans toutefois avoir des caractéristiques identiques à des rétrocroisements F1 x pédonculé (Figure 27A). De façon similaire les hybrides sessile x tauzin semblent non statistiquement différents des hybrides F1 (Figure 27E), néanmoins leur moyenne d'admixture est intermédiaire entre des hybrides F1 et des rétrocroisements F1 x tauzin. Ces deux classes d'hybrides montrent un déséquilibre vers l'une des deux espèces parentales sans néanmoins pouvoir être assimilé uniquement à des rétrocroisements (les modalités 1 et 4 pour le graphique A et les modalités 3 et 4 pour le graphique E sont statistiquement différentes). Ces hybrides semblent donc contenir un mélange d'hybrides de première génération et de rétrocroisements

avec une seule des espèces parentale, indiquant un sens préférentiel dans les croisements à l'issue de la première génération d'hybride.

Dans la parcelle de la Petite Charnie (modalité 5), seuls les hybrides pédonculé x sessile ont une distribution des coefficients d'admixture intermédiaires, comparable à celle des individus F1 simulés (Figure 27A). Les hybrides pédonculé x pubescent (B), pédonculé x tauzin (C), sessile x pubescent (D) et sessile x tauzin (E) se rapprochent des caractéristiques des individus issus de rétrocroisement entre l'hybride F1 et l'espèce la plus fréquente (respectivement, pédonculé et sessile). Il faut remarquer néanmoins que les moyennes des coefficients d'admixture pour ces hybrides sont légèrement inférieures et leur écart-type supérieurs comparativement aux rétrocroisements simulés, indiquant probablement l'existence d'un petit nombre d'hybrides de première génération.

A la Paguères (modalité 6), les hybrides pédonculé x pubescent (Figure 27B) ont une moyenne de coefficient d'admixture équivalente aux hybride F1, indiquant une absence de directionnalité dans les croisements entre les deux espèces. Pour toutes les autres classes d'hybrides, les moyennes des coefficients d'admixture tendent vers l'espèce parentale la plus fréquente. Ainsi, les hybrides pédonculé x sessile, sessile x pubescent et pubescent x tauzin se rapprochent des caractéristiques des espèces pédonculées et pubescentes, espèces fortement dominantes en effectif. Ces classes d'hybrides ont des moyennes cependant moins tranchées que les rétrocroisements simulés indiquant qu'ils ne sont pas constitué seulement de rétrocroisements mais probablement de quelques hybrides de première générations ou autres rétrocroisements.

- **Analyse d'admixture**

La contribution génétique ( $pI$  estimé par le modèle d'admixture implémenté dans LEADMIX) des deux espèces parentales de chaque type d'hybride est reportée dans le Tableau 13. La précision de l'estimation de la contribution génétique des espèces dépend fortement du nombre d'individus hybrides contenu dans la classe d'hybride analysée : l'intervalle de confiance à 95% est plus restreint pour les hybrides de Briouant (largeur de l'intervalle : de 0.12 à 0.19) que pour les hybrides comprenant moins de 20 individus (Petite Charnie et Paguères : largeur de l'intervalle : de 0.20 à 0.27).

La contribution génétique des espèces à la formation des hybrides (Tableau 13) reflète globalement la composition des peuplements. Dans tous les cas où l'une des deux espèces contribue de manière significativement supérieure, il s'agit de l'espèce la plus fréquente. Certaines valeurs de contribution ne diffèrent pas significativement entre les deux espèces malgré des effectifs fortement déséquilibrés, néanmoins, bien que certains intervalles de confiance incluent la valeur de 0.50, l'estimation moyenne de la contribution génétique des espèces aux hybrides est concordante avec le rapport des effectifs des espèces parentales.

Même si le sens de la contribution des espèces aux populations hybrides concorde avec les effectifs des espèces, les valeurs de ces contributions indiquent systématiquement une relativement forte contribution des espèces qui sont les plus rares (voire absentes) des parcelles. Malgré le faible effectif du chêne sessile à Briouant (moins de 1% des individus assignés à cette espèce), il contribue pour 30 à 50% à la formation

des hybrides pédonculé x sessile (Tableau 13). De la même façon, à la Petite Charnie, les espèces pubescentes et tauzines qui sont pratiquement absentes à l'état pur contribuent pour 20 à 40% et 25 à 44% à la constitution des hybrides pédonculé x pubescent et pédonculé x tauzin, respectivement. Ces résultats montrent que les hybrides entre espèces abondantes et rares (voire absentes) sont pour une bonne proportion des hybrides de première génération.

Tableau 13 : Estimation et intervalle de confiance à 95% du paramètre **p1** du modèle d'admixture (correspondant à la participation de la première espèce indiquée pour nommé les hybrides). Les valeurs de **p1** dont l'intervalle de confiance ne comprend pas 0.5 sont indiquées en gras.

Population	Hybride	N	<i>p1</i>
Briouant	Ped/Ses	27	<b>0.60</b> (0.50 - 0.69)
	Ped/Pub	38	<b>0.58</b> (0.51 - 0.65)
	Ped/Toz	45	0.53 (0.47 - 0.59)
	Ses/Toz	41	0.43 (0.36 - 0.51)
	Pub/Toz	85	<b>0.41</b> (0.34 - 0.47)
Petite Charnie	Ped/Ses	15	0.49 (0.37 - 0.60)
	Ped/Pub	13	<b>0.71</b> (0.60 - 0.80)
	Ped/Toz	13	<b>0.64</b> (0.56 - 0.75)
Paguères	Ped/Ses	15	0.62 (0.48 - 0.75)
	Ped/Pub	18	0.55 (0.44 - 0.66)

- Introgression et abondance des espèces

La caractéristique génétique des hybrides est fortement dépendante de l'effectif relatif des deux espèces parentales. Les deux estimateurs que nous avons utilisés pour caractériser les hybrides montrent tous les deux une forte corrélation positive avec l'effectif relatif des espèces parentales (Figure 28 ; Pearson's product-moment correlation :  $t = 7.8$ ,  $df = 13$ ,  $r = 0.91$ ,  $p < 0.0001$  pour le coefficient d'admixture et  $t = 6.7$ ,  $df = 8$ ,  $p < 0.001$  pour *p1*). Lorsque la fréquence relative d'une espèce augmente, la proportion du génome de cette espèce parmi les hybrides augmente (pente de la droite de régression = 0.3). Même lorsque l'une des deux espèces parentales est en effectif très réduit, les hybrides présentent une proportion non négligeable du génome de l'espèce la plus rare (ordonnée à l'origine : 0.38). Une analyse plus fine de cette relation pour la moyenne du coefficient d'admixture indique que trois types d'hybrides s'éloignent de la droite de régression, indiquant une introgression dont la directionnalité n'est pas seulement liée aux

effectifs relatifs des espèces parentales. Ainsi, les hybrides sessile x pubescent de Briouant présentent des caractéristiques plus proche de l'espèce sessile, une fois pris en compte le déséquilibre des effectifs parentaux (Figure 28, point 1, résiduelle : 0.15). Les hybrides pédonculé x pubescent à la Petite Charnie tendent davantage vers l'espèce pédonculé (Figure 28, point 3, résiduelle : 0.10) alors que ces hybrides à Paguères présentent inversement un excès de gènes pubescent (Figure 28, point 2, résiduelle : 0.10).

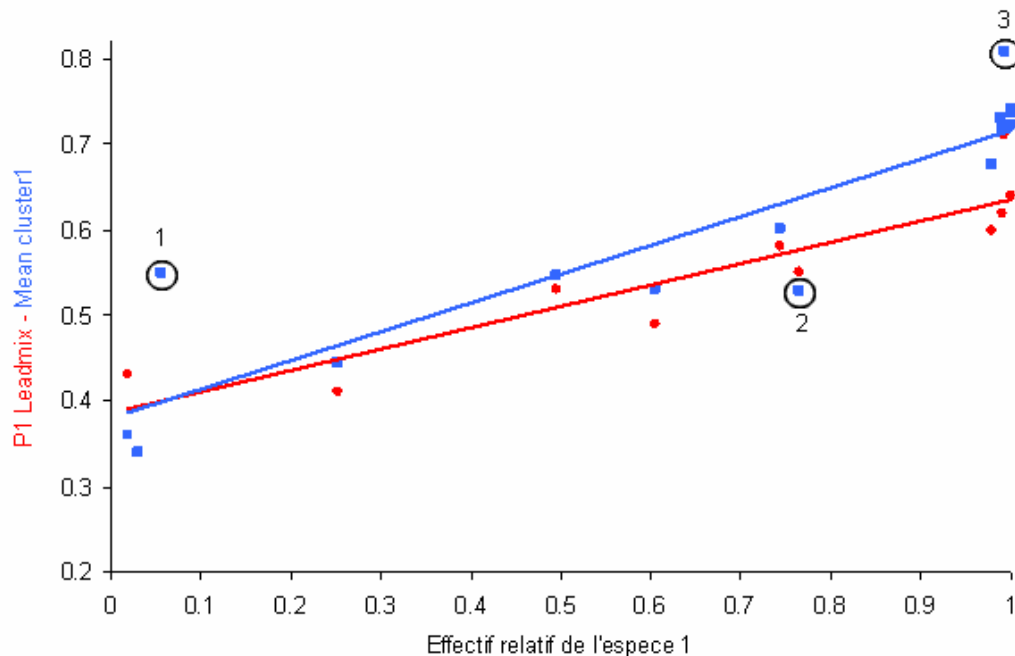


Figure 28 : Relation entre la fréquence d'une espèce et sa participation dans la composition génétique des hybrides estimée par deux méthodes : la moyenne du coefficient d'admixture (en bleu ;  $r = 0.91$ ,  $t=7.8$ ,  $df=13$ ,  $p<0.001$ ) et la proportion de contribution de l'espèce (paramètre  $p_1$ , en rouge ;  $r = 0.91$ ,  $t=6.3$ ,  $df=8$ ,  $p<0.001$ ).

## 4.4. Discussion

L'analyse de l'hybridation entre espèces de chênes européens est habituellement restreinte à deux espèces (Muir *et al.*, 2000 ; Valbuena-Carabana *et al.*, 2005) ou à un nombre limité de parcelles ou de populations (Curtu *et al.*, 2007). Cette étude est la première à analyser l'hybridation entre les quatre espèces les plus fréquentes du complexe des chênes blancs européens dans différentes parcelles et populations. Cette démarche nous a permis d'étendre l'échantillonnage à un plus grand nombre d'espèces pouvant intervenir dans l'hybridation et de comparer la dynamique d'hybridation entre sites.

### 4.4.1. Taux d'hybridation comparé entre populations

La fréquence des hybrides diffère entre les peuplements étudiés, variant de 15 à 20% selon les populations analysées et de 20 à 30% selon les parcelles. La différence de fréquence d'hybrides entre les populations et les parcelles est vraisemblablement liée à un biais d'échantillonnage favorisant les individus à la morphologie typique d'une espèce (morphotype sessile privilégié dans les populations pyrénéennes, morphotypes sessile et pédonculé pour les populations ONF et morphotypes sessile, pédonculé et pubescent pour les populations d'Aurignac). Parmi les trois parcelles étudiées, la Petite Charnie et

Briouant ont été échantillonnées exhaustivement et présentent un pourcentage d'hybride supérieur à la parcelle de la Paguères dont l'échantillonnage n'est pas exhaustif. Ces différences montrent que lorsque l'échantillonnage n'est pas exhaustif, les arbres ayant une morphologie atypique sont évités, diminuant légèrement le pourcentage d'hybrides échantillonnés. Ceci confirme aussi que les chênes à morphologie intermédiaire sont plus introgressés. Seul un échantillonnage exhaustif d'une parcelle (Petite Charnie et Briouant) semble pouvoir fournir une estimation non biaisée du pourcentage d'hybridation. Un échantillonnage non exhaustif mais réellement aléatoire pourrait permettre de corriger ce biais dans le cas d'échantillonnage en populations.

Le pourcentage d'hybrides trouvé dans cette étude (20 à 30% dans les parcelles) est supérieur à ceux observés dans d'autres études. Par exemple, Valbuena et al. (Valbuena-Carabana *et al.*, 2007) trouvent de 6 à 22% d'hybrides entre le chêne sessile et tauzin dans trois parcelles situées en Espagne. Curtu *et al.* (Curtu *et al.*, 2007) analysent l'hybridation entre quatre espèces de chênes en Roumanie et trouvent de 1.7 à 16.2% d'hybrides entre paires d'espèce. Dans notre étude, le taux d'hybrides entre paires d'espèces varie de 6.4% pour les hybrides sessile x pubescent à 20.5% pour les hybrides pubescent x tauzin à Briouant. Néanmoins, le pourcentage global d'hybride est élevé car il combine les effectifs de 6 types d'hybrides différents puisque les 4 espèces s'hybrident mutuellement. Typiquement, chaque type d'hybride représente de 0.7% (hybrides sessile x pubescent) à 10.4% (hybrides pubescent x tauzin) des chênes adultes de Briouant. Tous les couples d'espèces étudiés forment des hybrides en population naturelle et l'hybridation semble intensive puisque la méthode d'assignation utilisée est conservative et sous-estime le pourcentage d'hybridation réel (voir chapitre 3).

#### 4.4.2. Caractéristiques des hybrides

Les coefficients d'admixture des hybrides identifiés dans les populations ne peuvent pas être expliqués par une hybridation limitée à une première génération. Alors qu'un petit nombre d'hybrides montre des coefficients d'admixture pouvant correspondre à des hybrides de première génération (typiquement coefficient d'admixture de 0.5), de nombreux hybrides présentent des coefficients d'admixture intermédiaires entre ceux des hybrides F1 et ceux des espèces pures. Conjointement et plus globalement, la comparaison des hybrides simulés et identifiés dans les parcelles montre que les populations naturelles contiennent un mélange de plusieurs générations d'hybrides. L'hybridation entre les chênes n'est pas confinée à la première génération, mais semble se poursuivre au moins jusqu'à la seconde génération. Les hybrides semblent donc être viables et fertiles, ils contribueraient eux-mêmes à la descendance, ce qui expliquerait le pourcentage élevé d'hybrides dans les populations.

#### 4.4.3. Hybridation hors zone de sympatrie

Notre analyse multi-population met en évidence de façon répétée la présence cryptique d'espèces uniquement sous forme d'hybrides, alors que ces espèces n'ont jamais été décrites à l'état pur dans les peuplements étudiés, ni même parfois dans la région correspondante. Nous pouvons être confiants dans la réalité de cette assignation car les simulations (chapitre 3) ont montré que seul un faible pourcentage

d'hybrides est assigné par erreur à une autre classe d'hybride. Par ailleurs, nous identifions un nombre équivalent d'hybrides pédonculé x sessile (hybrides attendus) et pédonculé x tauzin ou pédonculé x pubescent (hybrides non attendus) à la Petite Charnie, chacun représentant environ 5% des arbres de la parcelle, ce qui démontre que ces assignations ne sont pas le fait de biais d'analyse mais bien une réalité biologique. Le taux de faux positif peut être calculé explicitement à partir des simulations détaillées au chapitre 3. Sur 150 hybrides pédonculé x sessile simulés, seulement 4 (soit 2.6%) sont assignés à une autre classe hybride (Tableau 5 du chapitre 3). De plus, l'analyse phylogénétique basée sur les distances entre groupe d'individus montre que ces hybrides inattendus présentent les mêmes caractéristiques génétiques que les hybrides correspondants dans les autres parcelles. Le cas du chêne tauzin est le plus spectaculaire car cette espèce est absente à l'état pure à la Petite Charnie mais bien représentée parmi les hybrides. Les peuplements de chênes tauzins les plus proches sont localisés à plusieurs dizaines de kilomètre de cette parcelle, ce qui démontre que l'influence d'une espèce par hybridation peut avoir lieu bien au-delà de sa zone de distribution (hybridation allopatriques). Ce type d'hybridation à longue distance a été identifié chez plusieurs espèces d'arbres (Depamphilis & Wyatt, 1989, 1990) dont le pollen transporté par le vent ou par des insectes parfois sur de longues distances peut aller féconder des populations distantes de plusieurs dizaines de kilomètres. Il n'est pas étonnant alors de trouver chez ces espèces une introgression par du pollen provenant d'une espèce non sympatrique et ce type d'introgression ne semble pas limité aux seuls arbres car il se retrouve chez les iris par exemple (Arnold, 1992).

#### 4.4.4. Hybridation et introgression fréquence-dépendante

Nous avons étudié la dynamique d'hybridation, en tirant avantage des trois parcelles analysées de manière plus fine. Chacune des parcelles présente des compositions spécifiques différentes pouvant influencer la dynamique d'hybridation (Anderson & Hubricht, 1938; Burgess *et al.*, 2005; Nason *et al.*, 1992b). D'une façon générale, les espèces rares sont davantage représentées parmi les hybrides alors que les espèces dominantes vont participer plus faiblement à l'hybridation. Nous avons pu expliquer cette tendance générale par : (1) la fréquence de formation des hybrides, (2) la caractéristique génétique des hybrides, et (3) la relation entre ces facteurs et les fréquences en espèces dans les parcelles.

L'estimation du nombre d'hybrides attendu d'après les effectifs des espèces présentes dans les parcelles montre que les deux espèces majoritaires produisent systématiquement moins d'hybrides qu'attendu sous l'hypothèse de reproduction aléatoire entre individus. Ce résultat est à contraster avec l'excès d'hybrides observé pour les espèces absentes de la parcelle (tauzin à la Petite Charnie par exemple) ou représentées par quelques individus (sessiles à Briouant ou pubescent à la Petite Charnie). Le chêne pubescent à Briouant semble se reproduire sans contrainte avec le chêne pédonculé et il forme une plus grande proportion d'hybrides avec le chêne tauzin.

Ces résultats montrent que la formation d'hybrides est limitée entre les deux espèces dominant les peuplements (pédonculé et tauzin à Briouant, pédonculé et sessile à la Petite Charnie, pédonculé et pubescent à Paguères), évoquant l'existence de possibles barrières reproductives pré- ou postzygotiques

entre ces espèces. Une sélection différentielle entre les hybrides et les espèces parentales pourrait expliquer cette observation. Les espèces fréquentes étant potentiellement bien adaptées à leur milieu, les hybrides entre ses espèces pourraient présenter une valeur adaptative inférieure aux espèces parentales et donc être partiellement contre sélectionnés.

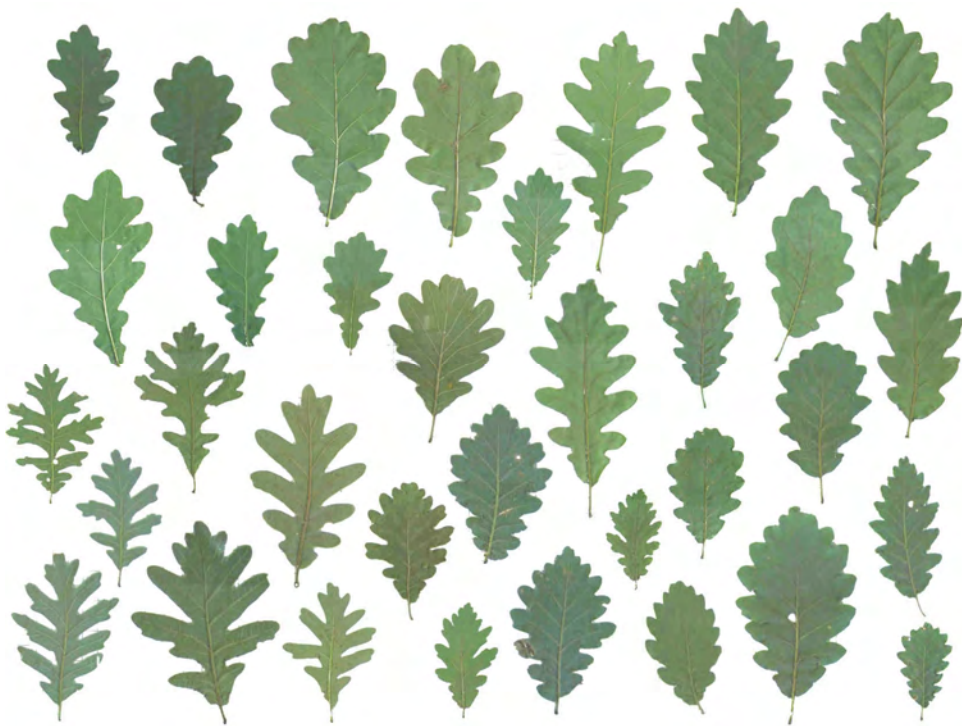
Cette analyse démontre que des hybrides sont formés entre tous les couples d'espèces présents dans les parcelles, et même au-delà par hybridation à longue distance. L'hybridation ne semble pas cantonnée à la première génération car on peut identifier des individus issus au moins d'une seconde génération d'hybridation, indiquant que certains hybrides sont fertiles et capables de fournir une descendance. Ceci explique le taux élevé d'hybrides identifiés dans les populations. Même lorsque la formation d'un hybride de première génération est un événement très rare, les populations peuvent contenir une forte proportion d'hybrides si les quelques hybrides formés sont viables et fertiles. En effet, l'ensemble de la descendance d'un hybride F1 représente autant d'individus hybrides de seconde génération. Ainsi chez des espèces telles que les épicéas ou les pins pignons, malgré un nombre limité d'hybrides de première génération, les hybrides de seconde générations et au delà sont nombreux (Lanner & Phillips, 1992; Perron & Bousquet, 1997).

La dynamique d'hybridation dépend donc d'une part de la viabilité et de la fertilité des hybrides, ce qui semble être démontré chez le chêne par les résultats présentés dans ce chapitre, et d'autre part par le comportement reproducteur de ces hybrides. En effet, la sélectivité du pistil hybride pour le pollen va grandement influencer la dynamique des flux de gènes au cours des différentes générations d'hybridations. Plusieurs scénarios de reproduction des hybrides peuvent être envisagés, chacun conduisant à des conséquences différentes au niveau du complexe d'espèces. Les hybrides peuvent être incompatibles avec les espèces parentales mais présenter des taux importants d'autofécondation. L'autogamie des hybrides peut conduire à la divergence de la population hybride et finalement à un isolement reproducteur complet des hybrides et des espèces parentales aboutissant à la spéciation hybride. Bien que fortement allogames, des résultats montrent que l'autogamie a néanmoins lieu chez certains chênes à un faible pourcentage (typiquement, moins de 5%, Bacilieri *et al.*, 1996b; Streiff *et al.*, 1999). Un scénario alternatif est la perte totale des barrières reproductrices des hybrides. Dans cet autre cas extrême, les hybrides constituent une véritable connexion génique entre espèces. Ce scénario peut même aboutir à une spéciation inverse si l'hybridation est telle que le maintien des espèces n'est plus assuré. Ceci semble peu probable pour les chênes car malgré une hybridation fréquente, l'intégrité des espèces est maintenue, probablement par une sélection divergente forte. Entre ces deux scénarios extrêmes, toutes les situations intermédiaires sont possibles. Cependant, ils mettent bien en évidence l'importance du comportement reproducteur des hybrides dans la dynamique d'hybridation ainsi que ses conséquences sur l'évolution du complexe d'espèces.





# Chapitre 5- Conséquences phénotypiques et génétiques de l'hybridation dans une parcelle



Page précédente : Exemple de variabilité morphologique parmi les chênes de Briouant.

## 5.1. Introduction - Caractéristiques générales du bois de Briouant

Les analyses statistiques classiquement réalisées en génétique des populations consistent à quantifier et à comparer la diversité génétique et sa structuration à l'intérieur ou entre populations définies a priori. La définition des populations est souvent réalisée de façon géographique (niveau intraspécifique) ou morphologique (niveau interspécifique).

Les résultats de l'assignation génétique au niveau individuel permettent de grouper les individus en fonction de leur espèce déterminée non pas par un a priori morphologique mais par une méthode purement fondée sur les principes de la génétique des populations et donc liée à une définition de l'espèce associée à sa reproduction. Dans ce chapitre, nous allons étudier quelques conséquences de l'hybridation au niveau **génétique** et **phénotypique** en utilisant l'assignation génétique comme classification a priori.

La première partie consistera en une description du **contexte historique et écologique** du bois de Briouant. Nous verrons que la stabilité des forêts à travers le temps à l'échelle du paysage permet de connaître l'origine des populations de chênes grâce à l'analyse de l'ADN chloroplastique à hérédité maternelle (Lepais *et al.*, 2006a ; voir ANNEXE B). Cependant, à l'échelle d'un bois, nous verrons que l'impact de la gestion impose de fortes contraintes démographiques.

La **morphologie foliaire** est le principal critère de reconnaissance des espèces, bien que les individus intermédiaires posent problème. L'étude de la morphologie foliaire de 300 individus va permettre (1) de montrer la concordance entre assignation génétique et détermination morphologique et (2) de comparer la morphologie des hybrides à celle des espèces parentales.

Nous allons ensuite quantifier la diversité **génétique** des espèces en incluant ou non leurs hybrides, en testant l'hypothèse que les hybrides vont augmenter la diversité génétique de leurs espèces et réduire la différenciation interspécifique. D'autre part, nous allons comparer la structuration génétique spatiale intra- et interspécifique. D'un point de vue théorique, l'intégration des hybrides dans les analyses devrait réduire la structuration génétique spatiale des espèces et augmenter la composante interspécifique de la structuration génétique.

### 5.1.1.1. Evolution à l'issue de la dernière glaciation

Lors de chaque épisode glaciaire qu'a connu le quaternaire, les espèces ont migré vers le sud pour trouver des conditions favorables à leur survie, dans des zones refuges localisées au sud des péninsules européennes (ibérique, italienne et balkanique) (Taberlet *et al.*, 1998). La recolonisation des espèces vers le nord à l'issue de la dernière période glaciaire (Hewitt, 1999) a laissé des traces bien visibles lorsque les populations sont étudiées avec des marqueurs génétiques appropriés (Comes & Kadereit, 1998). L'analyse de l'ADN chloroplastique, à hérédité maternelle chez les chênes et donc transmis par les graines, permet d'étudier l'histoire relativement ancienne des populations. Des analyses systématiques à travers l'Europe ont permis d'identifier plusieurs lignées chloroplastiques provenant de plusieurs refuges glaciaires (Petit *et al.*, 2002a; Petit *et al.*, 2002b, voir chapitre 1). L'étude de la diversité chloroplastique de la région autour

du canton d'Aurignac montre que les populations de cette zone proviennent majoritairement d'une lignée maternelle originaire de la péninsule ibérique (Lepais *et al.*, 2006a ; ANNEXE B). Ainsi, indépendamment de l'espèce, toutes les populations étudiées dans le canton d'Aurignac (zone centrale échantillonnée intensément Figure 29) présentent le même haplotype 10a originaire de la péninsule ibérique. Seul un chêne pubescent possède un haplotype différent (haplotype 1) provenant de refuges italiens. A l'échelle régionale (zone échantillonnée de façon plus lâche Figure 29), on retrouve l'haplotype 10a dans toutes les populations de la partie sud-ouest, alors que dans la zone nord-est on retrouve plusieurs haplotypes provenant des trois zones refuges supposées : haplotypes 10a, 10b et 11 appartenant à la même lignée originaire de la péninsule ibérique, l'haplotype 1 provenant de refuges italiens ainsi que l'haplotype 7 provenant de refuges balkaniques. Il y a donc une nette structuration géographique de ces haplotypes dans cette région. La présence exclusive d'un seul haplotype dans la partie sud-ouest de la zone d'étude quelle que soit l'espèce de chêne tend à montrer que des épisodes d'hybridations ont eu lieu pendant ou depuis l'établissement des peuplements de chênes dans la région. En outre, le chêne pubescent semble très lié à l'haplotype 1 à l'échelle de son aire de répartition ce qui laisse supposer que cette espèce était en partie restreinte aux refuges italiens (Petit *et al.*, 2002b). Cependant, à la limite ouest de sa répartition actuelle (vallée de la Garonne, de la Charente, ...) cette espèce possède des haplotypes de la lignée ibérique, qu'il partage avec les autres espèces présentes dans les mêmes peuplements (Petit *et al.*, 2002b). Une étude de la diversité chloroplastique sur les quatre espèces de chênes dans le sud de la France montre que les chênes pédonculés et sessiles partagent systématiquement le même haplotype au niveau local alors que le chêne pubescent partage l'haplotype local seulement dans certaines populations, indiquant que le processus d'hybridation avec cette espèce est soit en cours, soit localisé (Dumolin-Lapegue *et al.*, 1999). Les haplotypes ibériques "atypiques" des chênes pubescents pourraient avoir été gagnés par hybridation unidirectionnelle et récurrente avec les espèces de chênes préinstallées, conduisant au final à la migration de cette espèce par pollen.

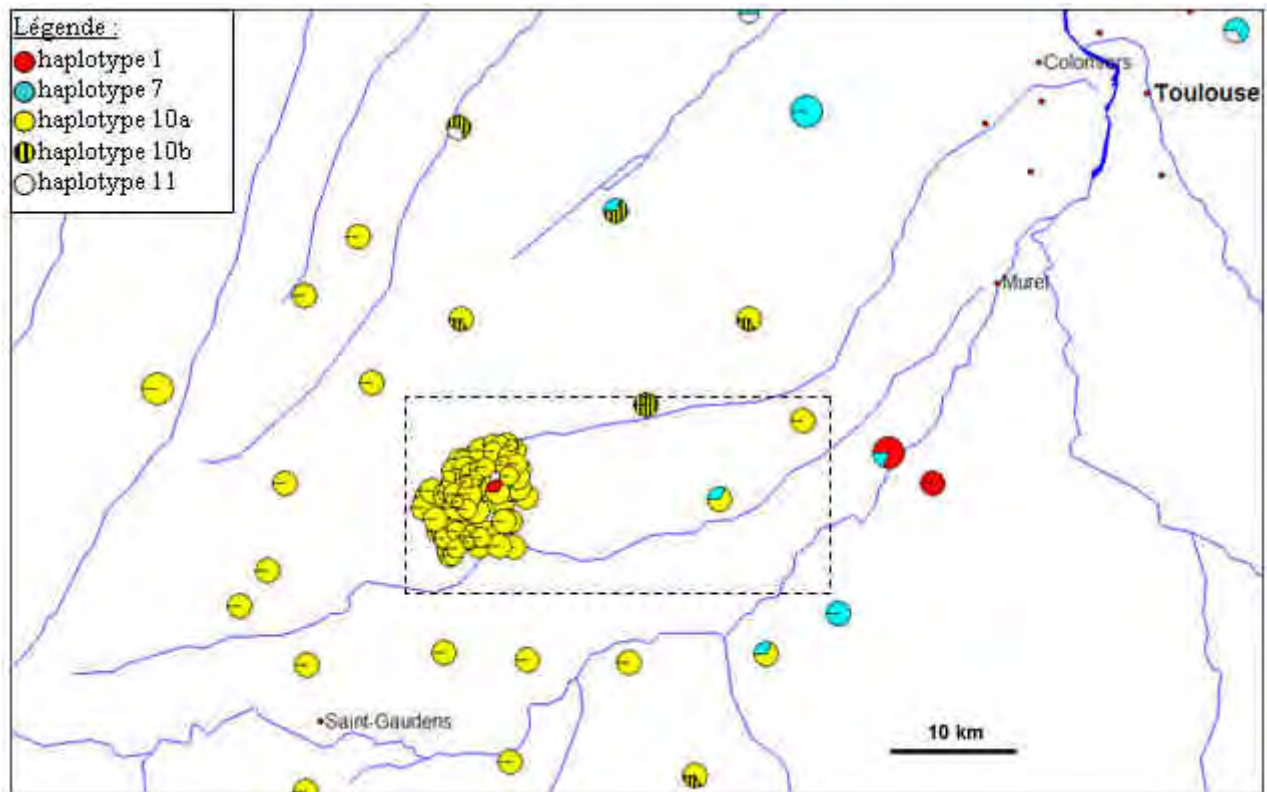


Figure 29 : Carte des haplotypes d'ADN chloroplastique de la région d'Aurignac (Haute Garonne) incluant trois espèces de chênes (non différenciées sur la carte) (Lepais *et al.*, 2006a). En pointillé : étendue de la Figure 30.

#### 5.1.1.2. Evolution récente du paysage

Le bois de Briouant est un peuplement de petite taille (env. 15 ha) localisé dans un paysage de coteaux surplombant la vallée de la Nère. L'activité agricole est constituée de polyculture et d'élevage et les forêts ne représentent qu'environ 10% de la surface terrière sous forme de petits bois fragmentés dont la plupart font moins de 20 ha. La localisation des bois est conditionnée par l'activité agricole : les peuplements forestiers sont souvent maintenus dans les zones à faible potentialité agricole (pentes abruptes de bordure de coteaux, sol caillouteux,...). En plus de cette multitude de petits bois, on trouve quelques peuplements de plus grosse taille comme les forêts de Fabas ou de Mauboussin (Figure 30). La carte de Cassini (Figure 30) montre que ce paysage a remarquablement peu évolué depuis le XVIII<sup>ème</sup> siècle. La plupart des emplacements boisés sont restés en place. La configuration des bois dans le paysage est cependant parfois légèrement modifiée : réduction de la surface des bois (exemple : Briouant), fragmentation de certaines forêts (au sud de Fabas) et création de quelques bois depuis cette époque.



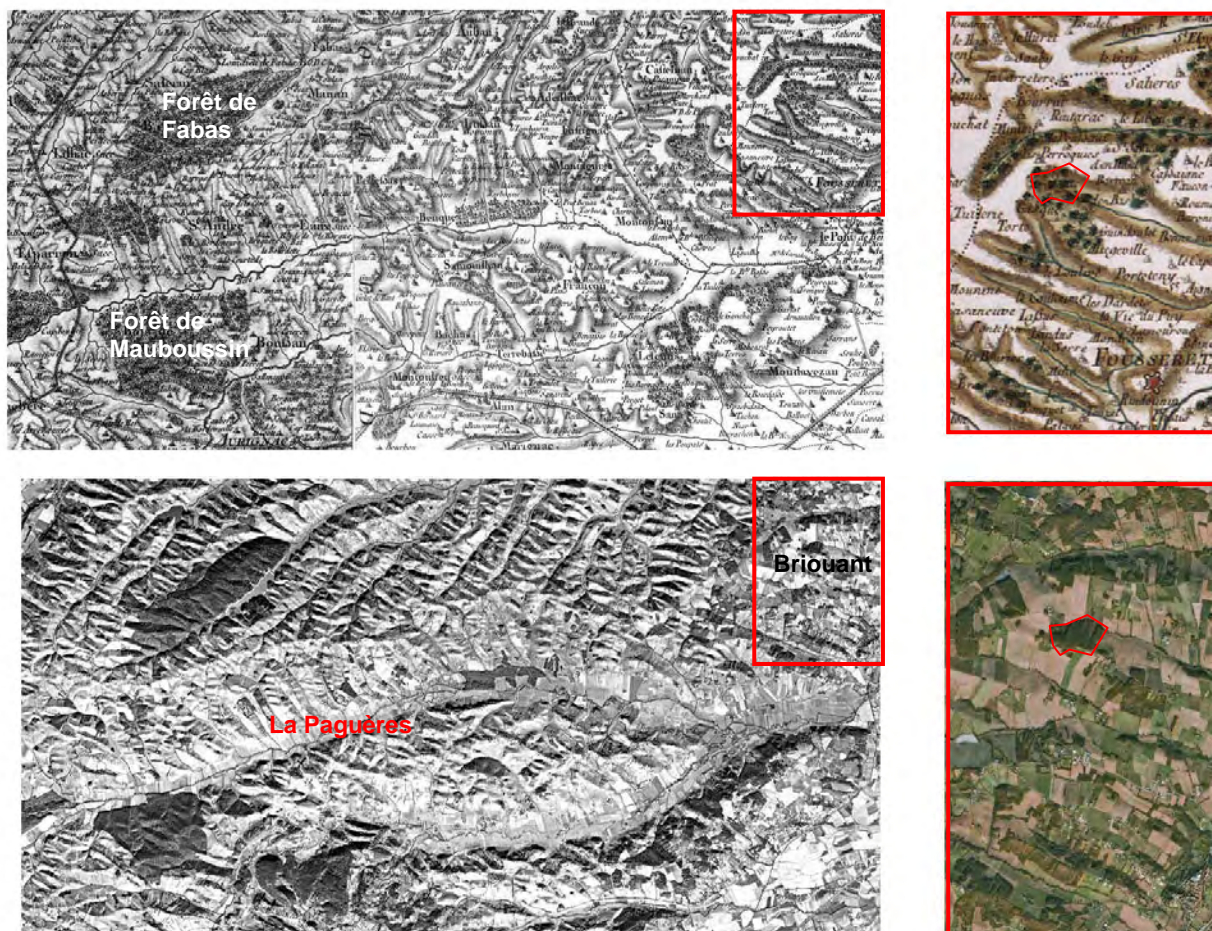


Figure 30 : Comparaison de l'étendue des surfaces boisées de la région d'Aurignac au XVIIIème siècle (en haut, carte de Cassini, n°75 et 39, 1770-1776) et au XXème siècle (photographie satellite, 1998). Le cadre rouge est centré sur les alentours du bois de Briouant (entouré en rouge dans le médaillon de droite).

### 5.1.1.3. Gestion des petites forêts privées

La plupart des bois présents dans la zone d'étude sont la propriété d'agriculteurs qui exploitent leur parcelle de forêt pour de multiples usages (bois, pâturage, chasse, cueillette, récréatif... ; Sourdril, 2007). Malgré une présence dispersée à l'échelle du paysage, la localisation des parcelles forestières est stable dans le temps car elles sont transmises de génération en génération (Guyon *et al.*, 1996). La gestion de ces bois est souvent différente d'un propriétaire à l'autre, mais en général, ils sont constitués de taillis voire de taillis sous futaies (De Warnaffe *et al.*, 2006). Ainsi lors des coupes, les souches sont laissées en place ce qui permet en un temps assez court une régénération de la forêt à partir des souches en place (arbre de cépée) mais aussi à partir de régénérations issues de graines (arbre franc-pied). Les peuplements des taillis regroupent donc plusieurs générations d'arbres et sont typiquement des milieux très hétérogènes. Par exemple, une étude incluant plusieurs bois dans le canton d'Aurignac (dont la parcelle de La Paguères) a montré que les stratégies de gestion étaient très variables d'un propriétaire à l'autre et d'une grande complexité spatio-temporelle (De Warnaffe *et al.*, 2006).

Pour illustrer cette gestion hétérogène, je présente ici des résultats obtenus par des étudiants de la Formation des Ingénieurs Forestiers lors d'une semaine de stage (FIF-ENGREF, 2006) dont l'un des but

était de faire un historique détaillé de la gestion du bois de Briouant grâce à l'analyse de photographies aériennes ainsi que d'interview de propriétaires. Le bois de Briouant est divisé en une vingtaine de parcelles qui sont la propriété de 8 personnes (Figure 31).

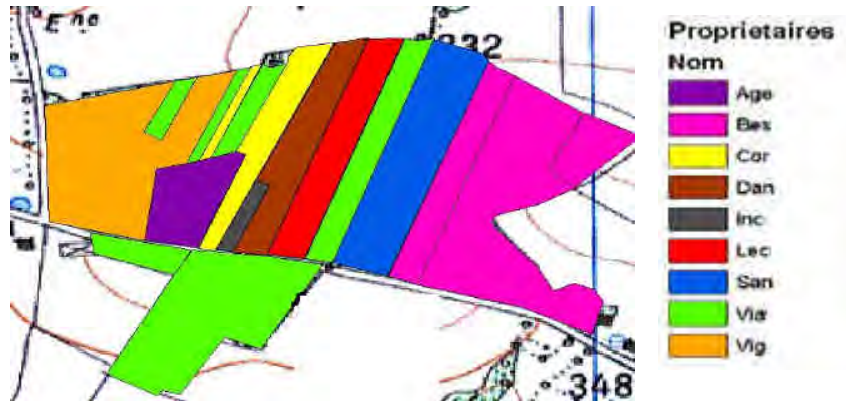


Figure 31 : Parcelles cadastrales et propriétaires du bois de Briouant.

L'analyse des photographies aériennes datant de 1942, 1953, 1971 et 1998 permet de mettre en évidence plusieurs zones de coupes (représentées en rouge, Figure 32) ainsi que des zones de régénération après coupe montrant une grande hétérogénéité de stratégie de récolte (petites zones de coupe ponctuelles, régénérations après coupe étendues et fragmentées en 1953 ou bien étendue et continue en 1971). L'intervention sur le peuplement ne se limite pas à la récolte proprement dite mais aussi à des éclaircies (peuplement mature clair de 1971) ou à des changements d'usage des parcelles (pâturage très présent dans les années 40, mise en pâturage de la zone sud plus récemment).



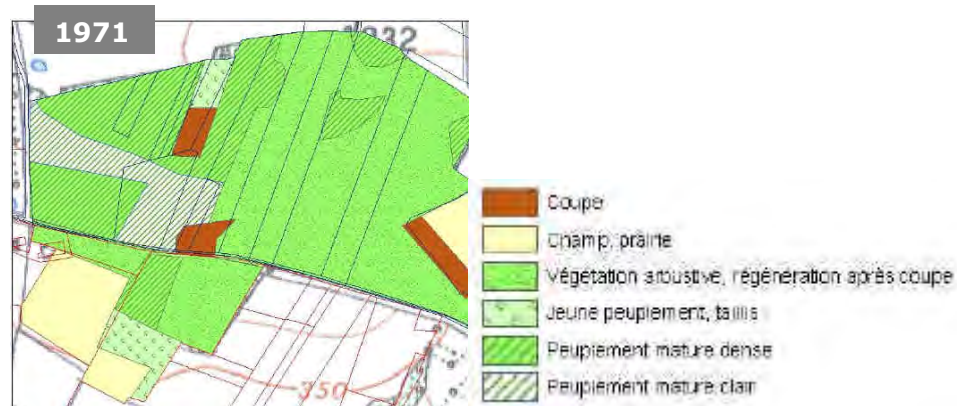


Figure 32 : Localisation des coupes et estimation de l'âge des arbres en place de 1942 à 1998 à partir de photographies aériennes (FIF-ENGREF, 2006).

La structure du bois de Briouant au printemps 2005 (Figure 33) montre l'hétérogénéité du bois tant au niveau de la composition en espèces d'arbres (densité de châtaigniers élevée par endroits), de son usage ainsi que des zones de coupes. Cependant cette structure a nettement évolué depuis le début de l'étude du bois, les zones de coupes se sont en particulier étendues depuis l'année 2005 : la zone n°1 (Figure 33) a subi plusieurs coupes successives (2006, 2007 et finalement totalement coupée en 2008 sauf pour quelques arbres laissés intacts en lisières) et la zone n°9 a également été réduite à une fine lisière longeant la route en 2006 et 2007.

L'ensemble des chênes adultes de la zone entourée en rouge (817 individus) a été échantillonné afin de mener les analyses génétiques. Les résultats des analyses d'assignation génétique ont montrés (chapitre 4) que cette parcelle contenait majoritairement du chêne pédonculé (241 individus) et du chêne tauzin (246 individus), ainsi qu'un nombre réduit de chênes pubescents (83) et de très rares chênes sessiles (5). De plus, une grande proportion des chênes adultes sont des hybrides (environ 30% soit 242 individus).

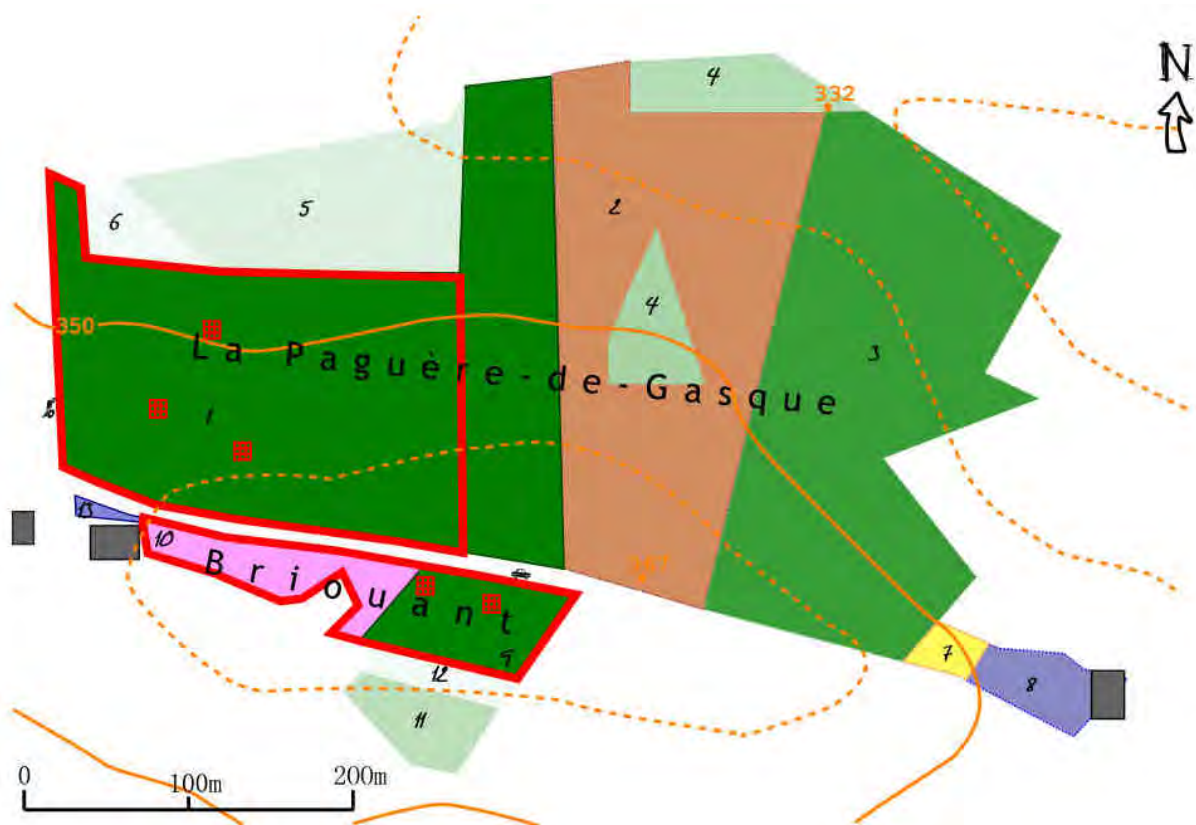


Figure 33 : Schéma structural du bois de Briouant en 2005. 1 : taillis avec de rares réserves, majoritaire : chênes (4 espèces de chênes : pédonculé > pubescent > tauzin >> sessile). 2 : taillis et quelques réserves, majoritairement châtaigniers, quelques chênes plus ou moins dispersés (surtout pédonculé et pubescent). 3 : taillis avec de rares réserves, majoritaire : chênes (surtout pédonculé et pubescent en lisière). 4 : coupes anciennes, taillis jeune impénétrable. 5 : coupe jeune, rejets de souches. 6 : coupe de l'année, souches fraîchement coupées (cernes : âges des souches les plus grosses : 95 ans environ) sans rejets. 7 : jeune plantation de feuillus (chênes rouges, acacia). 8 : jardin arboré de vieux chênes. 9 : taillis anciennement pâturé, majoritaire : chêne (3 espèces : tauzin > pubescent > pédonculé) ; forte densité de régénération. 10 : taillis clairsemé pâturé, uniquement chênes (4 espèces : tauzin > pédonculé > pubescent >> sessile). 11 : coupes jeunes, rejets de souches. 12 : coupe de l'année, souches de châtaigniers et de chênes sans rejet avec structures de cépée visible (cernes). 13 : vieux chênes de franc pied bordant l'étable. Echantillonnage exhaustif des zones entourées en rouge. Les petites grilles rouges représentent la localisation des échantillonnages de régénérations.

#### 5.1.1.4. Répartition des espèces de chênes dans la région

Le bois de Briouant est situé dans le département de la Haute Garonne (région Midi Pyrénées), zone de transition entre influences atlantique et méditerranéenne. Ce contexte biogéographique particulier est illustré par la coexistence de plusieurs espèces de chênes dans le bois dont le chêne pubescent, espèce de la série subméditerranéenne, très présent sur les coteaux calcaires thermophiles dans toute la région ainsi que le chêne tauzin, espèce caractéristique de la série atlantique, qui se trouve ici en limite est de son aire de répartition et dont la présence dans la région est rare. En effet le chêne tauzin se retrouve dans certains peuplements du piedmont pyrénéen, mais sa présence n'avait pas été décrite dans la zone des coteaux où se situe le bois de Briouant. Diverses prospections dans les populations entourant le bois de Briouant ainsi qu'à une échelle plus large (échantillonnage local et régional autour du canton d'Aurignac, Figure 29 et Lepais *et al.*, 2006a, ANNEXE 2) n'ont néanmoins pas permis d'identifier d'autres populations de chêne tauzin dans la région.

Ainsi dans la région du bois de Briouant les forêts sont principalement composées de trois espèces de chênes, dont la répartition s'explique par plusieurs facteurs (Gonzalez, 2006). Le chêne pédonculé se retrouve dans pratiquement tous les peuplements, principalement dans les zones les plus fraîches (bas de pente, bordure des cours d'eau) des bois. Le chêne sessile est une espèce que l'on retrouve principalement dans quelques massifs forestiers de grande taille (bois de Fabas et bois de Mauboussin) du fait de son caractère forestier. Dans les forêts de plus petite taille, quelques individus de cette espèce peuvent parfois se retrouver, principalement au centre du peuplement. Le chêne pubescent préfère les stations plus chaudes, comme les versants orientés vers le sud, et est particulièrement présent dans les lisières des bois.

#### 5.1.1.5. Répartition des espèces de chênes et de leurs hybrides dans la parcelle

Les espèces de chênes présentes ne sont pas distribuées au hasard dans le bois (Figure 34A). Le chêne pubescent se retrouve principalement en lisière ouest, le chêne pédonculé est bien représenté dans nord. Les quelques chênes sessiles sont principalement localisés dans la partie centrale du peuplement, en accord avec la tendance forestière de cette espèce. Le chêne tauzin est situé dans la grande partie sud-est du bois, zone présentant d'une part un sol un peu particulier, composé d'alluvions supérieures avec de gros galets siliceux (Figure 35), et d'autre part une utilisation pastorale historique. Le chêne tauzin est donc localisé dans les zones où le sol est le plus acide, en accord avec ses préférences écologiques. Enfin, même si les espèces semblent globalement être bien localisées, elles sont mélangées pied à pied en de nombreux endroits du bois, certaines espèces étant même présente en dehors de leur zone locale (par exemple, quelques chênes pédonculés et pubescent au sud, ou bien quelques chênes tauzins au nord, Figure 34A).

Contrairement à la répartition des espèces, les hybrides sont distribués de façon plus homogène (Figure 34B). Les hybrides x pédonculé ou x sessile sont localisés préférentiellement dans la zone où se situent les espèces parentales (hybrides x sessile dans la partie centrale et hybrides x pédonculé dans la zone nord), alors que les hybrides x pubescent et x tauzin se répartissent de façon homogène dans l'ensemble du bois.



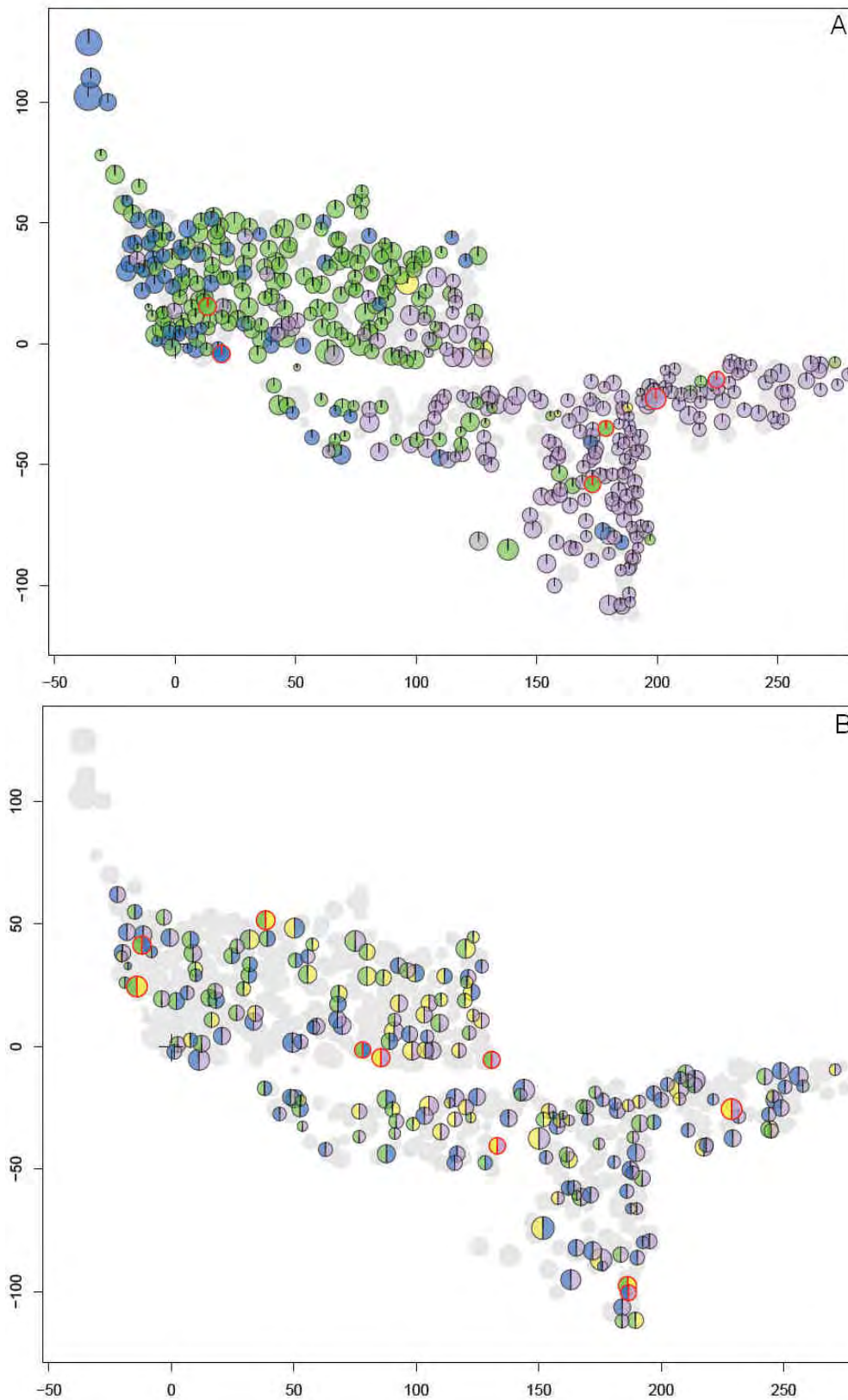


Figure 34 : Localisation des chênes assignées à une espèce pure (A) ou comme hybride (B) dans le peuplement de Briouant. Les couleurs représentent l'espèce des individus (vert : chêne pédonculé, jaune : chêne sessile, bleu : chêne pubescent et violet : chêne tauzin). Le diamètre du cercle est proportionnel au diamètre de l'arbre à hauteur de poitrine (dbh).

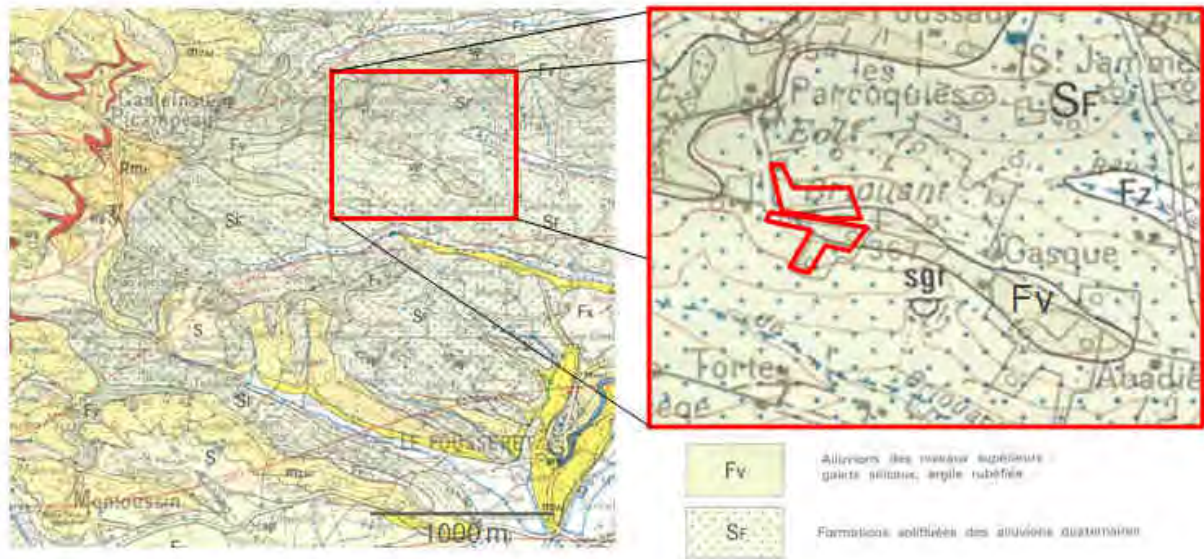


Figure 35 : Détails de la carte géologique 1/50000 Le Fousseret n°1033.

## 5.2. Morphologie foliaire

Afin de vérifier la concordance entre assignation génétique et détermination morphologique des espèces ainsi que de caractériser la morphologie foliaire des hybrides, nous avons échantillonné 10 feuilles de 336 chênes du bois. Cet échantillonnage comprend 77 chênes pédonculés, 5 chênes sessiles, 54 chênes pubescents, 67 chênes tauzins ainsi que des hybrides : 11 hybrides pédonculé x sessile, 21 pédonculé x pubescent, 29 pédonculé x tauzin, 8 sessile x pubescent, 23 sessile x tauzin et 41 pubescent x tauzin. Pour chaque feuille, une photographie est prise et un ensemble de 19 variables ont été mesurées par Florian Alberto (voir Tableau 14), certaines étant des mesures antérieurement utilisées pour l'analyse de la morphologie foliaire des chênes pédonculés et sessiles (Kremer *et al.*, 2002a) alors que d'autres ont été ajoutées afin d'essayer de discriminer les chênes pubescents et tauzins (Tableau 14).

La moyenne de chaque mesure pour les 10 feuilles de chaque arbre fournit un jeu de 19 variables pour chaque chêne. Une analyse en composantes principales (ACP) est appliquée à ces données, elle permet de synthétiser les variables morphologiques les plus discriminantes entre espèces (ici, les 2 premiers axes expliquent environ 60% de la variabilité totale). Cette analyse a été effectuée en prenant en compte seulement les chênes assignés à une espèce pure, ce qui a permis de construire un modèle reliant variables morphologiques et espèces de chênes. Ce modèle est ensuite utilisé pour projeter les individus hybrides dans l'espace vectoriel obtenu.

Une analyse qui prend en compte la moyenne de tous les chênes appartenant à la même espèce ou classe d'hybride a offert une vue plus générale de la relation morphologique entre les espèces et leurs hybrides.

Tableau 14 : Description des variables de morphologie foliaire mesurées

Variable	Description	Nature	Référence
pl	longueur du pétiole	mesure (mm)	Kremer <i>et al.</i> , 2002a
ll	longueur de la lamina	mesure (mm)	
lw	largeur du lobe	mesure (mm)	
sw	largeur du sinus	mesure (mm)	
wp	longueur de la lamina mesurée à partir du lobe le plus large	mesure (mm)	
nl	nombre de lobes	comptage	
nv	nombre de veines secondaires	comptage	
bs	forme basale de la lamina (de 1 à 9)	observation	
hr	densité de pubescence sur la partie abaxiale de la lamina (de 1 à 5)	observation	
ob	forme de la lamina	$100 \times wp / ll$	
pr	ratio de la longueur du pétiole	$100 \times pl / (ll + pl)$	
ldr	ratio de la profondeur du lobe	$100 \times (lw - sw) / lw$	
pv	pourcentage de veination	$100 \times nv / nl$	
lwr	ratio largeur sur longueur de la lamina	$100 \times lw / ll$	
sr	densité de pubescence sur la partie adaxiale de la lamina (de 1 à 5)	observation	Viscosi <i>et al.</i> , 2008a
nr	densité de pubescence au niveau de la nervure centrale (de 1 à 5)	observation	
ptr	densité de pubescence au niveau du pétiole (de 1 à 5)	observation	
permieter	périmètre	mesure (mm)	
circularity	circularité	$4\pi (\text{area} / \text{perimeter}^2)$	



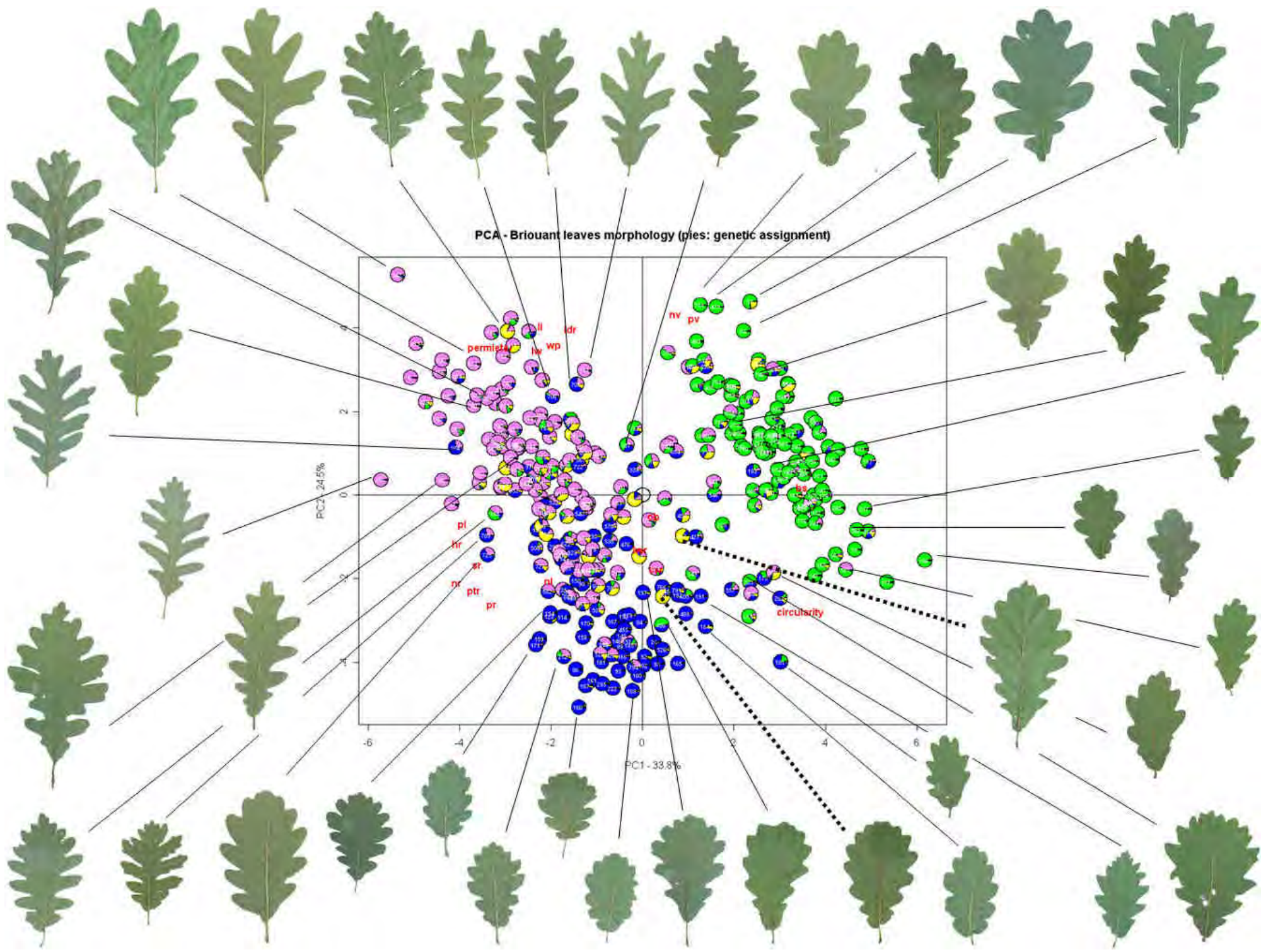




Figure 36 (page précédente) : Projection des individus et des variables dans l'espace vectoriel résultant de l'analyse en composante principale et visualisation de la morphologie foliaire pour certains des chênes en différents points de l'espace vectoriel. Les couleurs des camemberts représentent les résultats de l'assignation génétique (proportion de gènes de chaque espèce dans le génome de l'individu). Vert : chêne pédonculé, jaune : chêne sessile, bleu : chêne pubescent et violet : chêne tauzin.

Le premier axe de l'espace vectoriel (PC1) explique 33.8% de la variation pour l'analyse individuelle (Figure 36). Il différencie les chênes pédonculés (PC1 positif) des autres espèces (PC1 négatif). Cet axe est positivement corrélé à la forme basale de la lamina (bs) et partiellement à la circularité (circularity) et négativement corrélé à la longueur du pétiole (pl), à la pubescence et partiellement au périmètre. Le second axe (PC2) explique 24.5% de la variation dans l'analyse individuelle. Il différencie principalement les chênes pubescents (PC2 négatif) des chênes tauzins (PC2 positif). Les variables positivement corrélées à cet axe sont le nombre de veines secondaires (nv), la longueur et la largeur de la lamina (ll et lw), la profondeur du lobe (ldr), la longueur de la lamina à partir de sa partie la plus large (wp) ainsi que le périmètre. La largeur du sinus (sw), la circularité ainsi que le rapport longueur sur largeur de la lamina (lwr) sont négativement corrélés à l'axe 2.

Le chêne pédonculé se distingue donc nettement des deux autres espèces par un pétiole réduit, la forme typique en oreillette de la base de la lamina ainsi que par l'absence de pubescence et la présence de nervures secondaires. La distinction des chênes tauzins et pubescents est moins nette mais rendue possible grâce à un périmètre plus grand des feuilles pour le chêne tauzin lié à des sinus profonds et à des feuilles généralement plus grandes. On note également la présence de veines secondaires chez le chêne tauzin, caractère absent chez le chêne pubescent. Ce dernier présente en outre des feuilles plus petites avec davantage de lobes et une forme généralement plus circulaire (voir Figure 36 pour une comparaison de quelques formes de feuilles).

L'analyse des moyennes par espèce produit les mêmes résultats au niveau de la description des variables morphologiques (Figure 37) et fait apparaître de façon systématique que les hybrides ont, en moyenne, une morphologie intermédiaire entre les deux espèces parentales (voir Figure 37A). Cette tendance globale cache cependant une grande variabilité au sein des espèces et des hybrides (Figure 37). La variabilité morphologique intra-espèce est très importante, ce qui pourrait en partie être lié au fait que certains de ces chênes sont issus d'un individu hybride dans leur proche généalogie (3 générations précédentes ou plus). Néanmoins, les espèces de chênes sont connues pour avoir une morphologie très variable en fonction du peuplement, des conditions environnementales, de leur origine géographique (Camus, 1936) et même de la position de la feuille dans l'arbre. Certains hybrides présentent toutes les caractéristiques morphologiques typiques d'une des deux espèces parentales alors que d'autres présentent des caractères intermédiaires. Il apparaît donc que les hybrides représentent une mosaïque très diversifiée de formes.

Outre l'analyse de la morphologie à l'aide des 19 mesures spécifiques, ces feuilles ont aussi fait l'objet d'analyses morphométriques à l'aide de deux méthodes automatiques (méthodes "landmark" et "outline") mises au points spécifiquement pour les espèces du genre *Quercus* (Viscosi *et al.*, 2008a). L'analyse comparative de ces trois méthodes a fait l'objet d'une publication dans la revue *Plant Biosystem* (Viscosi *et al.*, 2008b) montrant que ces trois techniques aboutissent à des résultats semblables, bien que l'analyse des 19 variables morphologiques mesurées manuellement se soit montrée légèrement plus performante. Néanmoins, les méthodes morphométriques ont l'avantage d'être automatiques ce qui permet d'analyser un nombre important de feuilles. De plus elles capturent la forme générale des feuilles, ce qui est un avantage important comparé à la mesure d'un nombre limité de variables préalablement choisies pour distinguer les espèces.

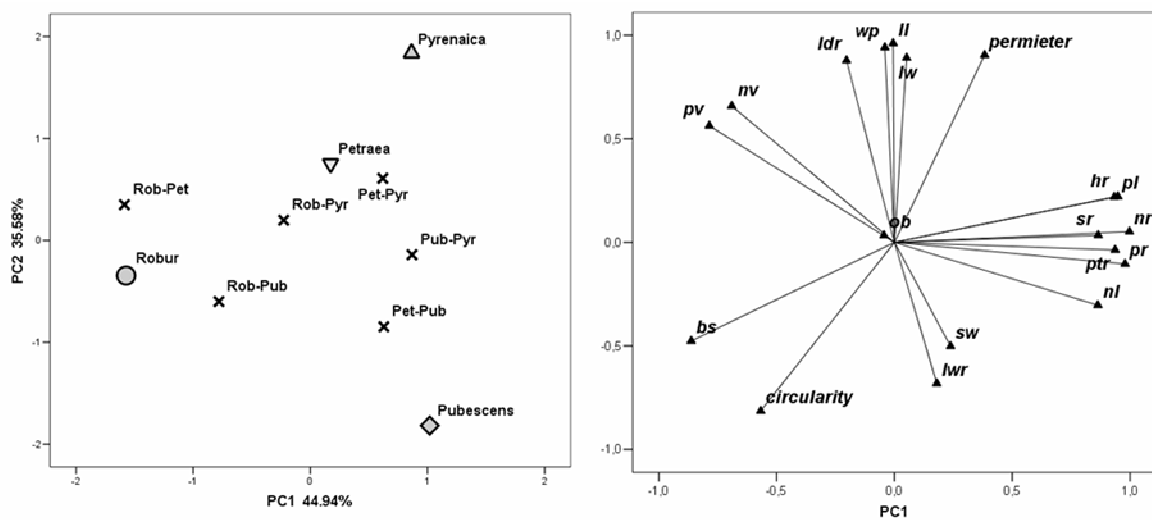


Figure 37 (d'après Viscosi *et al.*, 2008b ; Annexe C) : Ordination diagrams (PCA) computed on 19 leaf morphological variables and 10 average group (4 pure species and 6 hybrid group). The distribution of average leaf on PC1 and PC2 axes is reported on the left; the vector plot of variables on PC1 and PC2 axes is reported on the right. Filled circles (*Q. robur*); filled diamonds (*Q. pubescens*); open triangles (*Q. pyrenaica*); open upturned triangles (*Q. petraea*); cross (hybrid groups).

### 5.3. Structure génétique

L'assignation génétique présentée au chapitre 4 nous a permis de montrer la présence de quatre espèces dans le bois de Briouant et de définir au niveau individuel le statut spécifique de chaque chêne. Dans ce paragraphe, nous allons étudier l'organisation de la diversité génétique au sein des trois espèces les plus fréquentes, les seuls 5 chênes sessiles ne permettant pas d'inclure cette espèce dans les analyses.

Deux types de groupements ont été effectués. Le premier s'intéresse uniquement aux chênes assignés à une espèce (proportion du génome appartenant à une espèce supérieure ou égale à 0.90 ; groupes notés Ped, Toz ou Pub). Le deuxième a pour but de mettre en évidence d'éventuelles conséquences de l'hybridation sur l'organisation de la diversité génétique. Il comprend tous les individus assignés à une espèce ( $Q \geq 0.9$ ) et les hybrides entre cette espèce et une autre ( $0.5 < Q < 0.9$ ), ces groupes sont notés Ped+hyb, Toz+hyb et Pub+hyb.

#### 5.3.1. Méthodes

Le déficit d'hétérozygote est estimé grâce à l'indice de fixation  $F_{IS}$  (Weir & Cockerham, 1984) pour chaque locus et chaque espèce pure grâce au logiciel Genetix (Belkir *et al.*, 1996-2004). Afin de tester leur significativité, 1000 permutations des allèles à l'intérieur de chaque espèce ont été effectuées. Les taux d'hétérozygote attendus calculés sans biais,  $H_{NC}$  (Nei & Chesser, 1983), et l'hétérozygotie observé sont calculés pour les différents groupes précédemment décrits. Le nombre d'allèles estimé par une technique de raréfaction permet de comparer la diversité alléliques entre groupes d'individus n'ayant pas les mêmes effectifs d'échantillonnage (Petit *et al.*, 1998). Ainsi dans notre cas, le nombre d'allèles a été obtenu après raréfaction à 50 individus pour les différentes espèces grâce au logiciel HP-RARE (Kalinowski, 2005), permettant une comparaison non biaisée de la richesse allélique entre les espèces, incluant ou non les hybrides.

Enfin, les différenciations génétiques  $F_{ST}$  (Weir & Cockerham, 1984) pour chaque locus ainsi que la différenciation multilocus entre chaque paire d'espèces, incluant ou non les hybrides, ont été calculées grâce au logiciel Genetix (Belkir *et al.*, 1996-2004).

#### 5.3.2. Résultats et discussions

Quelle que soit l'espèce, un  $F_{IS}$  significatif et élevé est observé pour les locus QR39, QR5 et QR65 (Tableau 15) avec respectivement 0.08, 0.21 et 0.39. Ce résultat démontre un déficit d'hétérozygote plus ou moins important pour ces trois locus, ce qui est concordant avec les résultats du chapitre 2 qui faisaient apparaître la présence d'allèles nuls ou d'erreurs de génotypage relativement fréquents pour ces locus. En dehors de Qr11 chez les chênes tauzins (déficit en hétérozygote significatif) et Qr7 chez les chênes pubescents (excès d'hétérozygote significatif), tous les autres locus montrent des valeurs de  $F_{IS}$  non différentes de zéro (Tableau 15). Globalement, le  $F_{IS}$  sur l'ensemble des 10 locus est significatif pour toutes les espèces ( $F_{IS} = 0.068$ ), alors qu'il n'est pas différent de zéro lorsque les trois locus à fort déficit en hétérozygote

sont retirés ( $F_{IS} = -0.009$ ). La comparaison de la diversité et de la richesse alléliques entre espèces prendra donc en compte uniquement les 7 microsatellites qui sont à l'équilibre de Hardy-Weinberg (Tableau 16).

Tableau 15 : Déficit en hétérozygote, FIS (Weir & Cockerham, 1984), par locus et par espèce.

Fis	Qp110	Qr11	Qr112	Qr39	Qr96	Qr20	Qr5	Qr65	Qr7	Qr87	F <sub>IS</sub> (10ssr)	F <sub>IS</sub> (7ssr)
Ped	-0.044	-0.000	0.044	<b>0.049*</b>	-0.010	0.010	<b>0.159***</b>	<b>0.343***</b>	0.003	-0.015	<b>0.062***</b>	-0.002
Toz	0.007	<b>0.057*</b>	0.004	<b>0.072***</b>	-0.017	0.017	<b>0.222***</b>	<b>0.309***</b>	0.002	-0.040	<b>0.069***</b>	0.003
Pub	-0.048	-0.006	-0.015	<b>0.111**</b>	0.041	-0.016	<b>0.241***</b>	<b>0.505***</b>	<b>-0.074*</b>	-0.085	<b>0.072***</b>	-0.029
Moyenne	-0.028	0.017	0.011	<b>0.077</b>	0.005	0.004	<b>0.207</b>	<b>0.386</b>	-0.023	-0.047	<b>0.068</b>	-0.009

Tests sur  $F_{IS}$  obtenus par 1000 permutations des allèles à l'intérieur de chaque population.

\* :  $p < 5\%$ , \*\* :  $p < 1\%$ , \*\*\* :  $p < 0.1\%$ , pas d'étoile : non significatif.

La diversité génétique estimée par le niveau d'hétérozygotie et la richesse allélique apparaît légèrement plus élevée pour le chêne pubescent même si la différence avec les deux autres espèces n'est pas significative (Tableau 16). Lorsque des hybrides sont ajoutés aux différentes espèces (notation Esp+hyb), on constate que le déficit d'hétérozygote reste proche de zéro sauf pour le groupe des chênes pubescents et leurs hybrides qui montre un léger excès d'hétérozygote significatif ( $F_{IS} = -0.031$ ). Les niveaux de richesse allélique montrent les mêmes tendances (Tableau 16) : ils sont plus élevées pour le chêne pubescent (en moyenne 14.43 allèles par locus) et plus faible pour le chêne pédonculé (moyenne de 12.7 allèles par locus), bien que les différences ne soient pas significatives du fait de l'écart-type élevé entre locus. Lorsque des hybrides sont ajoutés aux différentes espèces, les richesses alléliques classent les espèces dans le même ordre mais elles sont plus élevées bien que la différence ne soit pas significative (14.07 allèles par locus pour les chênes pédonculés et leurs hybrides et 15.02 allèles par locus pour les chênes pubescents et leurs hybrides). Globalement, le niveau de diversité génétique est très élevé dans ce peuplement, conformément aux nombreuses analyses similaires précédemment réalisées chez ces espèces. L'hybridation joue un rôle dans cette diversité génétique puisque le fait d'ajouter quelques individus hybrides dans les groupements se traduit globalement par une légère augmentation des indices de diversité génétique.

Tableau 16 : Diversité génétique et richesse allélique pour 7ssr.

Esp	N	$H_{NC}$ (e. t.)	$H_O$ (e. t.)	$F_{IS}$	$A_{R-50}$ (e. t.)	Esp+hyb	N	$H_{NC}$ (e. t.)	$H_O$ (e. t.)	$F_{IS}$	$A_{R-50}$ (e. t.)
Ped	197	0.76 (0.16)	0.76 (0.17)	-0.002	12.70 (3.02)	Ped+hyb	248	0.78 (0.14)	0.78 (0.14)	0.004	14.07 (3.42)
Toz	217	0.79 (0.14)	0.79 (0.15)	0.003	13.06 (3.82)	Toz+hyb	292	0.80 (0.15)	0.80 (0.15)	0.003	13.88 (3.86)
Pub	63	0.85 (0.06)	0.88 (0.08)	-0.029	14.43 (3.93)	Pub+hyb	95	0.86 (0.07)	0.88 (0.08)	-0.031*	15.02 (3.48)

$H_{NC}$  : hétérozygotie attendue calculée sans biais Nei & Chesser, 1983,  $H_O$  : hétérozygotie observée,  $F_{IS}$  : déficit en hétérozygote (Weir & Cockerham, 1984),  $A_{R-50}$  : richesse allélique estimée par raréfaction à 50 individus (Kalinowski, 2004), e. t.: écart-type, \* :  $p < 0.05$ .

La différenciation génétique ( $F_{ST}$ ) entre espèces est significativement différente de zéro pour tous les locus et toutes les paires d'espèces (Tableau 17 et Tableau 18). Elle est la plus forte pour la paire chêne pédonculé / chêne tauzin ( $F_{ST(7SSR)} = 0.134$ ) alors que la paire chêne pubescent / chêne tauzin montre la différenciation interspécifique la plus faible ( $F_{ST(7SSR)} = 0.075$ ). Les niveaux de différenciation génétique interspécifique est particulièrement hétérogènes selon les locus (Tableau 17). Deux locus (Qr11 et Qr112) montrent une différenciation interspécifique élevée (moyenne 0.141 et 0.152 respectivement). Le locus Qr96 montre un fort niveau global de différenciation interspécifique ( $F_{ST} = 0.268$ ) en grande partie lié à la forte différenciation pour les paires d'espèces pédonculé / pubescent ( $F_{ST} = 0.363$ ) et pédonculé / tauzin ( $F_{ST} = 0.308$ ) alors que la différenciation est remarquablement plus faible pour la paire pubescent / tauzin ( $F_{ST} = 0.081$ ). La structuration de la diversité génétique à ce locus concorde remarquablement avec le fait qu'il a été localisé dans un QTL impliqué dans la longueur du pétiole (Saintagne *et al.*, 2004) : il différencie principalement les espèces ayant des pétioles de longueurs différentes (le chêne pédonculé a un pétiole très court, alors que les deux autres espèces ont des pétioles plus long). Globalement, les trois locus présentant des déficits en hétérozygotes élevés montrent aussi un assez faible niveau de différenciation interspécifique. L'estimation multilocus excluant ces trois locus est globalement plus élevé ( $F_{ST} = 0.121$ ) que celle incluant les 10 locus ( $F_{ST} = 0.093$ ). Enfin, la différenciation génétique interspécifique pour 7 locus est plus faible lorsque les hybrides sont inclus dans l'analyse (Tableau 18).

Tableau 17 : Différenciation génétique ( $F_{ST}$ ) par locus et par paire d'espèces.

$F_{ST}$	Qp110	Qr11	Qr112	Qr39	Qr96	Qr20	Qr5	Qr65	Qr7	Qr87	$F_{ST}$ (10ssr)	$F_{ST}$ (7ssr)
Ped-Toz	0.081	<b>0.149</b>	<b>0.169</b>	0.019	<b>0.308</b>	0.071	0.066	0.024	0.068	0.092	0.103	0.134
Ped-Pub	0.033	<b>0.113</b>	<b>0.136</b>	0.013	<b>0.363</b>	0.064	0.050	0.023	0.046	<b>0.152</b>	0.097	0.128
Toz-Pub	0.046	<b>0.141</b>	<b>0.118</b>	0.023	0.081	0.004	0.032	0.015	0.069	0.072	0.058	0.075
Global	0.066	<b>0.141</b>	<b>0.152</b>	0.018	<b>0.268</b>	0.057	0.056	0.022	0.064	0.099	0.093	0.121

Toutes les valeurs de  $F_{ST}$  sont significatives au moins au seuil de 5% après 1000 permutations globales des individus.

Tableau 18 :  $F_{ST}$  multilocus pour 7 microsatellites par paire d'espèces avec leurs hybrides (tableau droite) ou sans (tableau gauche).

$F_{ST}$	Ped	Toz	$F_{ST}$	Ped+hyb	Toz+hyb
Toz	0.134		Toz+hyb	0.111	
Pub	0.128	0.075	Pub+hyb	0.100	0.049

Les résultats concernant le déficit en hétérozygote, la diversité génétique ainsi que la différenciation interspécifique montrent que l'hybridation intervient dans la structuration de la diversité génétique chez ces espèces. Les espèces de chênes présentes à Briouant sont en condition panmictique (pas de déficit d'hétérozygote) que les hybrides soient inclus ou non dans l'analyse. En théorie, l'hybridation devrait créer un excès d'hétérozygote ( $F_{IS}$  négatif), ce qui est le cas seulement pour le chêne pubescent et ses hybrides, groupe pour lequel le nombre d'hybrides est élevé par rapport au nombre d'individus d'espèces pure (Tableau 16). Pour les chênes pédonculés et tauzins, la proportion d'hybrides est plus faible ce qui pourrait expliquer la faible différence entre le  $F_{IS}$  avec ou sans hybrides.

La diversité génétique, estimée par le taux d'hétérozygotes et la richesse allélique, augmente lorsque les hybrides sont ajoutés dans l'analyse. Cette tendance systématique n'est pas significative car là encore les groupes réalisés avec ou sans hybrides se différencient uniquement par l'ajout de quelques hybrides dans l'analyse.

Enfin, conformément avec les résultats précédents, la différenciation génétique interspécifique est diminuée lorsque les hybrides sont inclus dans l'analyse, ce qui démontre le caractère intermédiaire de leurs compositions alléliques.

## 5.4. Structure génétique spatiale fine

Le système de reproduction peut influencer la formation de structures génétiques locales : une distance de dispersion limitée des graines peut par exemple conduire au regroupement d'individus apparentés autour des individus maternels. D'autres facteurs comme la distance de dispersion du pollen, la reproduction entre individus phénotypiquement similaires (*associative mating*, reproduction préférentielle entre individus ayant une même période de floraison par exemple) peuvent aussi contribuer à la formation d'une structure génétique fine, c'est-à-dire à un apparentement plus élevé des individus séparés par une faible distance. Plusieurs études ont montré que les chênes présentaient une structuration spatiale significative mais faible pour de faibles distances, indiquant un apparentement légèrement plus élevé des individus proches (Streiff *et al.*, 1998 ; Valbuena-Carabana *et al.*, 2007).

D'autre part, la densité en individus capables de se reproduire (c'est-à-dire, la densité efficace,  $D_e$ ) peut aussi influencer la structure génétique spatiale car un plus petit nombre d'individus reproducteurs va conduire à augmenter l'apparentement. Les modèles théoriques d'isolement par la distance prédisent que l'apparentement diminue linéairement avec le logarithme de la distance à un taux proportionnel à  $1/4\pi D_e \sigma^2$  (avec  $D_e$  : densité efficace et  $\sigma^2$  : la distance de dispersion (distance parent-enfant) moyenne au carré). En théorie, l'hybridation peut avoir deux conséquences opposées sur la structuration génétique spatiale des individus (Valbuena-Carabana *et al.*, 2007). La reproduction préférentielle entre individus de la même espèce ou entre hybrides peut réduire la densité efficace et donc augmenter la structuration génétique, ou bien, si les hybrides se reproduisent librement avec les espèces pures, l'augmentation de la densité efficace va au contraire réduire la structuration génétique spatiale. Une étude portant sur le couple d'espèces chêne sessile / chêne tauzin dans trois populations différentes a montré que l'hybridation avait un impact sur la structuration génétique spatiale des individus (Valbuena-Carabana *et al.*, 2007). D'une part, la structuration génétique spatiale est plus faible lorsque les hybrides sont inclus dans l'analyse, d'autre part, cette structuration spatiale a une composante interspécifique non négligeable, c'est-à-dire que des paires d'individus d'espèces différentes sont significativement plus apparentées lorsqu'elles sont séparées par de faibles distances.

Ce paragraphe a deux objectifs : (1) comparer la structuration génétique spatiale des trois espèces de chênes du bois de Briouant avec ou sans leurs hybrides et (2) tester l'existence d'une composante interspécifique de la structuration génétique spatiale.

### 5.4.1. Méthodes

Cette analyse a été motivée par le développement récent d'un indice, la statistique  $S_p$ , rendant possible la comparaison de la structure génétique entre population ou espèces indépendamment des caractéristiques spatiales de l'échantillonnage ou du nombre d'individus étudiés (Vekemans & Hardy, 2004). Cette statistique est calculée à partir des paramètres de la droite de régression de l'apparentement entre individus en fonction du logarithme de la distance qui les sépare. Pour chaque espèce (avec ou sans ses hybrides), le coefficient d'apparentement  $F_{ij}$  entre individus

(Loiselle *et al.*, 1995) est calculé pour 15 classes de distances grâce au logiciel Spagedi (Hardy & Vekemans, 2002). Le graphique représentant le coefficient d'apparement en fonction du logarithme de la distance de chaque classe est tracé. La partie linéaire de la courbe est utilisée pour calculer la pente ( $b$ ) qui avec la valeur du coefficient d'apparement de la classe de distance la plus petite ( $F_1$ ) sert à calculer la statistique  $S_p$  selon la formule (Vekemans & Hardy, 2004) :

$$S_p = \frac{-b}{1 - F_1}$$

On calcule  $S_p$  pour chaque espèce avec ou sans leurs hybrides.

Enfin, afin de mettre en évidence une éventuelle composante interspécifique de la structure génétique spatiale, l'apparement est calculé entre paires d'individus d'espèces différentes pour 15 classes de distances.

#### 5.4.2. Résultats et discussions

Que ce soit pour les espèces pures (Figure 38 A, B et C) ou les espèces et leurs hybrides (Figure 38 D, E et F), l'apparement diminue de façon linéaire avec le logarithme de la distance. L'apparement est significatif pour les premières classes de distances et cette structure spatiale s'étend jusqu'à 25 mètres pour le chêne pédonculé et pubescent alors qu'elle reste significative jusqu'à 40 mètres pour le chêne tauzin (Figure 38 et Figure 40). Le coefficient d'apparement pour la première classe de distance (Tableau 19) est plus élevé pour le chêne pubescent ( $F_1 = 0.038$ ) alors qu'il est plus faible pour le chêne tauzin ( $F_1 = 0.014$ ) et pédonculé ( $F_1 = 0.07$ ). Lorsque les hybrides sont inclus dans l'analyse, les valeurs d'apparements sont légèrement plus faibles (Tableau 19). L'indice  $S_p$  montre que la structuration génétique spatiale est plus élevée chez le chêne pubescent comparé avec le chêne pédonculé et tauzin (Figure 39). De plus, lorsque les hybrides sont intégrés aux analyses, la structuration spatiale est nettement plus faible (Figure 39). Ce résultat montre que l'hybridation tend à diminuer la structuration génétique spatiale, ce qui suggère que les espèces et leurs hybrides constituent une seule et même unité reproductive.

Tableau 19 : Paramètres de structure spatiale des espèces

Esp	$b$	$F_1$	$S_p$ (e.t.)	Esp+hyb	$b$	$F_1$	$S_p$ (e. t.)
Ped	-0.0058	0.0073	0.0058 (0.0005)	Ped+hyb	-0.0044	0.0069	0.0044 (0.0004)
Toz	-0.0083	0.0147	0.0084 (0.0005)	Toz+hyb	-0.007	0.0135	0.0071 (0.0003)
Pub	-0.0231	0.0382	0.0240 (0.0012)	Pub+hyb	-0.0146	0.0222	0.0149 (0.0008)

$b$  : pente de la régression linéaire de l'apparement selon le logarithme de la distance (voir figure 8),  $F_1$  : valeur de l'apparement pour la première classe de distance,  $S_p$  (e.t.) : statistique  $S_p$  et écart type.



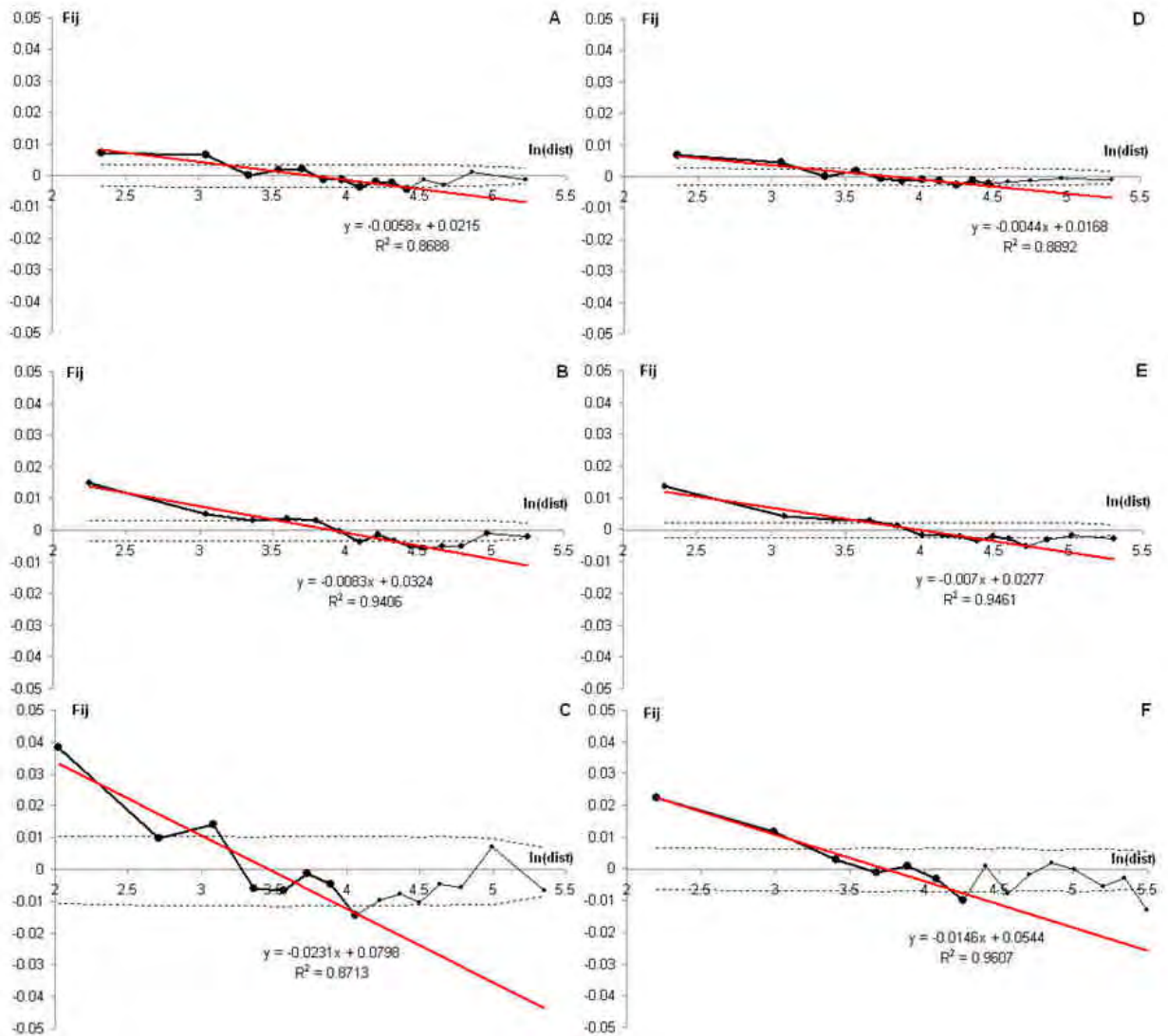


Figure 38 : Coefficients d'apparentements  $F_{ij}$  calculés pour 15 classes de distances (exprimées en logarithme) pour les chênes pédonculés (A, et leurs hybrides D), les chênes tauzins (B, et leurs hybrides E) et pubescents (C et leur hybrides F). Les traits pointillés représentent l'intervalle de confiance à 95% du coefficient d'apparentement sous hypothèse nulle d'absence de structure spatiale obtenu par 1000 permutations des individus. La partie épaisse de la courbe représente la partie linéaire utilisée pour la pente ( $b$ ) de la régression linéaire (droite rouge) permettant d'estimer la statistique SP.

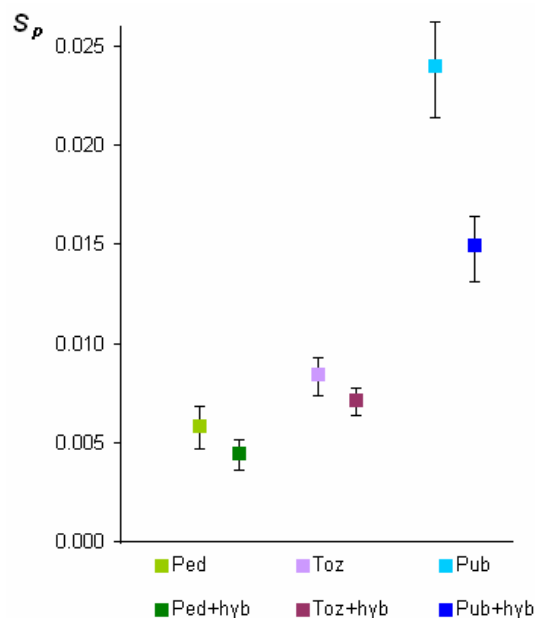


Figure 39 : Valeur et écart-type de la statistique SP pour les différentes espèces et leurs hybrides.

L'analyse d'apparement inter-espèce (Figure 40) indique que les individus d'espèces différentes ont des coefficients d'apparement négatifs lorsque les deux espèces comparées sont génétiquement éloignées (cas du couple pédonculé / tauzin, Figure 40 A et D ; et du couple pédonculé / pubescent, Figure 40 B et E). Cet apparement est significatif pour les premières classes de distance dans le cas des comparaisons pubescent / tauzin, que les hybrides soient inclus ou non (Figure 40 C et F), mais elle reste significative sur une plus grande distance lorsque les hybrides sont pris en compte. Pour ce qui est de la paire pédonculé / tauzin, une structuration génétique spatiale interspécifique significative apparaît seulement lorsque les hybrides sont inclus (Figure 40 D). Les résultats de analyse de la paire d'espèce pédonculé / pubescent (Figure 40 B et F) sont difficiles à interpréter car il semble que l'apparement soit significatif seulement pour des classes de distance intermédiaire, que les hybrides soient intégrés à l'analyse ou non, indiquant que les individus appartenant à ces deux espèces sont plus apparementés lorsqu'ils sont séparés par une distance d'au moins 30 mètres. De plus, la comparaison interspécifique incluant seulement les individus d'espèce pure est significativement négative pour la première classe de distance indiquant que les individus de ces deux espèces sont moins apparementés que la moyenne pour de très faibles distances.

Mis à part le cas de la paire d'espèces pédonculé / pubescent, une composante interspécifique à l'apparement entre individus à été trouvée à faible distance. Ces résultats corroborent ceux obtenus au chapitre 4 dans lesquels nous avons estimé la proportion d'hybride attendue sous hypothèse de panmixie en fonction des effectifs des espèces en présence (chapitre 4). En effet, la paire pubescent / tauzin présentant un excès d'hybrides montre aussi un apparement interspécifique relativement élevé alors que la paire d'espèces pédonculé / tauzin qui produit moins d'hybrides qu'attendu sous panmixie possède une composante interspécifique de la structure génétique spatiale seulement lorsque les hybrides sont pris en compte dans l'analyse. Enfin, les résultats sont moins clairs pour le couple d'espèce pédonculé / pubescent puisque nous

avons montré que le nombre d'hybrides formés correspondait à une reproduction panmictique entre espèce alors qu'aucune composante interspécifique nette à la structure génétique spatiale à faible distance n'a été trouvée.

Néanmoins, le fait de trouver un apparemment significatif entre individus d'espèces différentes démontre des événements d'hybridation relativement récents. La dispersion principalement locale et à faible distance des glands chez les chênes crée une structuration familiale du peuplement, mais de plus, nos résultats montrent que dans le bois de Briouant la structuration spatiale familiale a une composante interspécifique, démontrant l'importance des flux de gènes interspécifiques dans l'évolution locale des espèces de chênes.

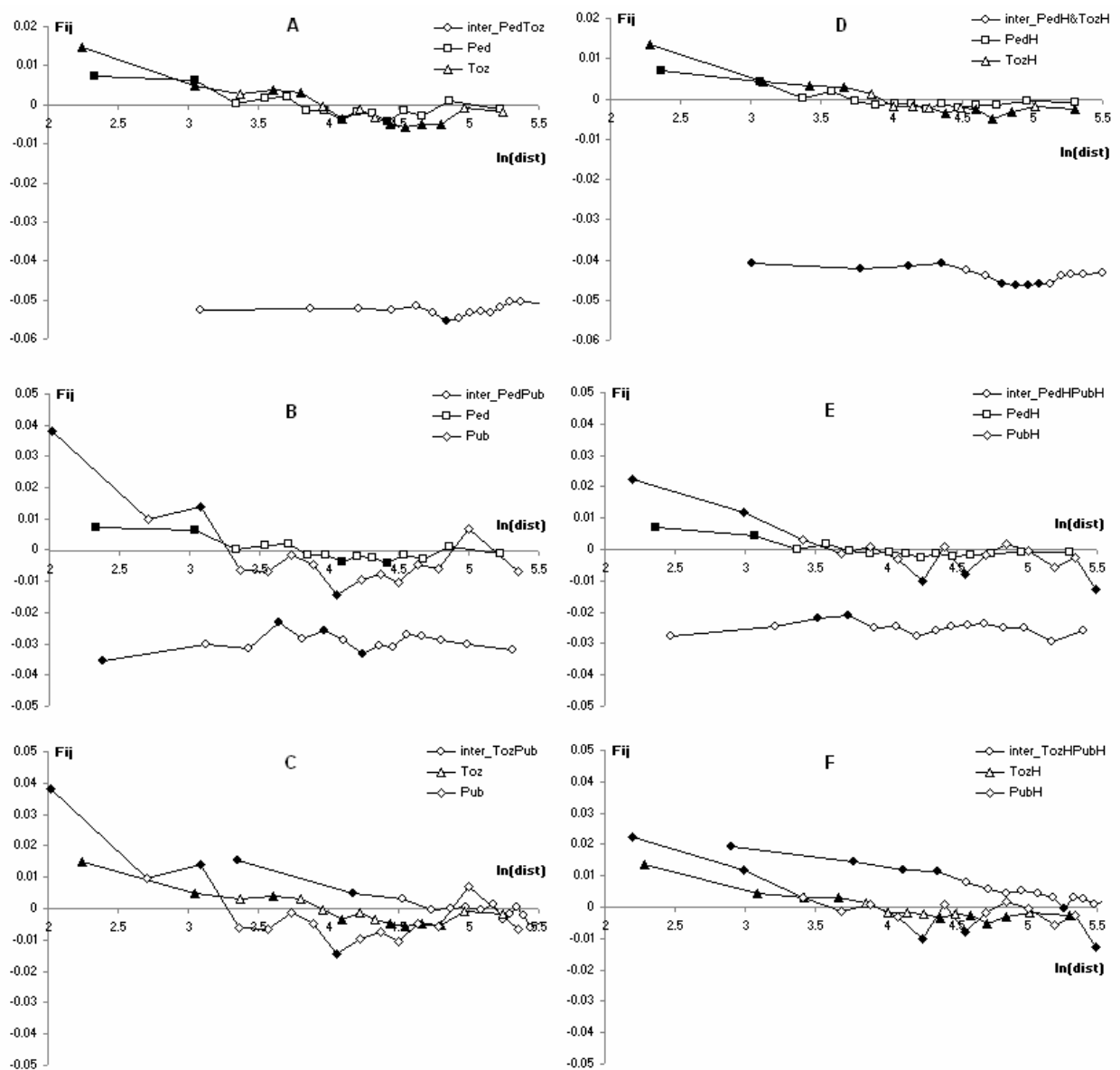


Figure 40 : Coefficients d'apparements  $F_{ij}$  intra- et inter-espèce (représentés par les ronds) calculés pour 15 classes de distances (exprimées en logarithme) pour les chênes pédonculés (carré) et tauzins (triangle) (sans, A, et avec les hybrides D), les chênes pédonculés et pubescents (losange) (sans, B, et avec leurs hybrides E) et les chênes pubescents et tauzins (sans, C, et avec les hybrides F). Les valeurs significatives sont représentées par des symboles pleins (noir).

## 5.5. Conclusion

Le bois de Briouant est situé dans une zone de contact entre espèces de chênes provenant de refuges différents à l'issue de la dernière glaciation. Malgré la fragmentation importante de la forêt et la gestion intensive des bois en taillis, l'étude du polymorphisme de l'ADN chloroplastique a permis de montrer que toutes les espèces de chênes de cette région partagent le même haplotype en provenance de refuges ibériques. Cet haplotype atypique pour le chêne pubescent pourrait s'expliquer par des événements d'hybridations relativement anciens. D'autre part, la gestion en taillis induit une forte perturbation du milieu, un facteur qui pourrait augmenter le taux d'hybridation entre espèces (Anderson, 1948). Cette gestion pourrait expliquer le fort taux d'hybrides détecté à Briouant comparé aux autres peuplements analysés au chapitre 4.

L'étude de caractères morphologiques a permis de conforter l'assignation génétique : chaque cluster génétique identifié correspond à une unité morphologique distincte. D'autre part nous avons pu montrer que les hybrides ont en moyenne une morphologie intermédiaire entre leurs deux espèces parentales. Ces résultats ont été confirmés grâce à trois méthodes d'analyses morphologiques distinctes.

L'analyse de la diversité génétique réalisée sur les espèces et leurs hybrides a permis de montrer une légère augmentation du niveau de diversité lorsque les hybrides sont inclus dans l'analyse. Les espèces, avec ou sans leurs hybrides, ne présentent pas de définit en hétérozygote ce qui indique une reproduction panmictique entre individus purs et hybrides. Une structuration génétique spatiale fine a été mise en évidence pour les trois espèces analysées. Cette structuration est plus ou moins prononcée en fonction des espèces mais elle est plus faible lorsque les hybrides sont inclus ce qui montre qu'ils participent à la reproduction avec leurs espèces parentales. Une composante interspécifique à l'apparementement entre individus proches a de plus été détectée indiquant que des individus d'espèces différentes peuvent avoir une origine familiale commune et que l'hybridation plus ou moins récente conduit à modifier l'organisation spatiale de la structure génétique du peuplement.



# **PARTIE C – SYSTÈME DE REPRODUCTION**

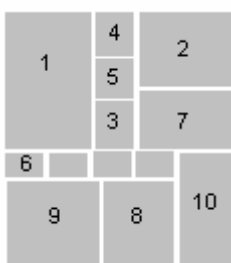


# Chapitre 6- Flux de gènes interspécifiques en conditions contrôlées





Page précédente : Croisements contrôlés chez les chênes.



1 : Isolement de branches d'un chêne pubescent adulte à l'aide d'une nacelle, 2 : isolement de jeunes arbres entiers, 3 : Guy Roussel effectuant l'injection de pollen, 4 et 5 : glands immatures de chênes pédonculés et sessiles, 6 : plaques de glands, 7 : germination de glands, 8 : développement de la tige, 9 : premières feuilles, 10 : première saison de croissance.

## 6.1. Introduction

Les espèces constituent des groupes de populations qui sont isolés reproductivement de tous les autres groupes similaires ("Biological Species Concept"; Mayr, 1963). Selon cette définition "standard", l'isolement reproducteur est considéré comme une barrière permettant de délimiter deux espèces. Cet isolement peut avoir de nombreuses causes et intervenir à différents stades de vie. L'addition de plusieurs barrières partielles peut permettre un isolement reproducteur total.

Cette définition de l'espèce permet d'envisager l'hybridation sous deux angles. D'une part, le croisement entre deux espèces, même s'il a lieu, peut ne pas permettre de produire un hybride viable. Dans ce cas, l'isolement entre espèces est total, les espèces sont clairement délimitées. D'autre part, si deux espèces produisent une descendance fertile, elles sont alors classées comme deux sous-espèces. Ce concept d'espèce a motivé de nombreuses études sur les processus de spéciations dans lesquelles une attention particulière a été portée à l'hybridation pour comprendre les facteurs responsables de l'isolement reproducteur entre espèces. Les expériences d'hybridations artificielles ont permis d'identifier plusieurs types de barrières reproductives agissant avant ou après la fécondation. Les mécanismes intervenant pour empêcher la fécondation sont qualifiés de prézygotiques, ceux qui réduisent la viabilité ou la fertilité des hybrides ou de leurs descendants sont qualifiés de postzygotiques (voir chapitre 1). Ces études ont également permis d'élaborer des théories et parfois de vérifier les modèles de spéciation et d'évolution des barrières reproductives lors de la divergence entre espèces. Enfin, outre le fait d'être un simple test du statut d'espèce, l'hybridation artificielle permet d'éliminer certaines des barrières reproductives, en particulier les barrières prézygotiques liées à l'isolement éco-géographique ou temporel. L'isolement génétique prézygotique ainsi que les barrières postzygotiques peuvent alors être quantifiées en conditions contrôlées.

Les principaux résultats d'hybridation artificielle chez les chênes blancs européens concernent les chênes pédonculés et sessiles (Steinhoff, 1993 ; Steinhoff, 1998 ; Kleinschmit & Kleinschmit, 2000). Ces expériences indiquent que les deux espèces produisent des hybrides viables. Cependant elles mettent également en évidence une asymétrie, le chêne pédonculé peut être fécondé par du pollen sessile, alors que le pollen pédonculé féconde le chêne sessile avec une très faible efficacité. En population naturelle, cette asymétrie, si elle persiste dans la descendance, va conduire à l'introgression directionnelle de gènes sessiles dans un fond génétique pédonculé, restaurant au fur et à mesure des générations un génome sessile.

Les chênes pédonculé et sessile sont souvent accompagnés des chênes pubescent et tauzin qui appartiennent à la même section *Quercus* et dont plusieurs hybrides ont été décrits en Europe (chapitre 1). Ainsi nous avons voulu inclure ces quatre espèces dans une expérience d'hybridation artificielle. Le premier but était de quantifier l'isolement reproducteur entre ces espèces. Est-il possible d'obtenir des hybrides F1 pour toutes les combinaisons testées? Le deuxième objectif était de préciser, grâce à diverses observations des stades de vie (nombre de glands produits,

germination, survie), l'importance relative de barrières reproductives prézygotiques (compatibilité génétique et préférence au pollen conspécifique) et postzygotiques (germination et survie) dans l'isolement reproducteur entre espèces.

## 6.2. Matériels et méthodes

### 6.2.1. Croisements contrôlés chez les chênes

La pollinisation artificielle sur chênes a débuté à Pierroton en 1987 et la technique a été continuellement améliorée lors de ces 10 années d'expérience. La technique utilisée ainsi que ses évolutions sont décrites en détail dans un document accessible en ligne (Roussel, 2002; <http://www.pierroton.inra.fr/biogeco/genetique/documents/pollinisation.pdf>).

Les anthères sont récoltées sur les chênes avant déhiscence du pollen et sont mises à sécher dans des poches-tamis. Une fois le pollen mur, il est tamisé et stocké au congélateur. La viabilité du pollen est estimée grâce à un test à l'acétate de fluorescéine (Heslop-Harrison & Heslop-Harrison, 1970). Les grains de pollen viable contiennent une estérase qui libère la fluorescéine de son groupement acétate ce qui la rend fluorescente. Au microscope sous éclairage UV, un grain de pollen viable fluoresce alors qu'un grain non viable reste sombre. La viabilité du pollen est alors estimée par le rapport entre le nombre de grains de pollen viable et le nombre de grains de pollen total. Lorsqu'un mélange de pollen de plusieurs chênes est utilisé dans un croisement, le taux de viabilité sert à ajuster la quantité de pollen de chaque arbre.

Les arbres mères sont empochés au début de la période de débourrement afin de les isoler du pollen extérieur. Dès le début de la période de réceptivité des fleurs femelles, des injections de pollen sont réalisées grâce à un injecteur à air comprimé (Roussel, 1999). L'injection de pollen est répétée 3 à 4 fois pendant 10 jours environ afin de maximiser les chances de pollinisation. Entre fin septembre et début octobre, les glands sont récoltés, comptés et pesés puis conservés en chambre froide avant d'être mis à germer.

### 6.2.2. Chênes utilisés dans les croisements

#### 6.2.2.1. Arbres mères

Les croisements ont été réalisés sur deux types d'arbres : des greffes de chênes pédonculés et sessiles conservés en pots à l'Unité Expérimentale de Pierroton et des chênes tauzins et pubescents adultes localisés respectivement dans un parc à Cestas (Gironde) et en bordure d'une vigne à Branne (Gironde).

Les greffes sont empochées en entier et plusieurs copies (greffes) d'un même arbre mère sont utilisées comme répétition. Dans le cas des arbres adultes, ce sont des branches qui sont isolées par des poches et plusieurs poches constituent les répétitions.

Les conditions de croisement sont assez différentes entre ces deux types d'arbre mère et vont influencer la quantité de glands récoltés. Les greffes (chênes pédonculés et sessiles) sont irriguées, traitées contre divers parasites des glands et leur proximité au laboratoire ainsi que leur

petite taille permet des injections de pollen plus fréquentes. A l’opposée, les arbres adultes (chênes tauzins et pubescents) subissent les conditions environnementales parfois défavorables (sécheresse, parasites...) et leur accessibilité (distance, utilisation d’une nacelle pour les injections de pollen) contraint à réduire le nombre d’injection de pollen. L’identification des arbres mères utilisés est présentée dans le Tableau 20.

Les croisements ont également été réalisés sur des chênes hybrides obtenus par croisements contrôlés en 1991 et 1992. Les jeunes arbres fructifères ont été échantillonnés pour analyse génétique afin de confirmer leur caractère hybride (génotypes conformes aux deux parents connus a priori). Cependant, les résultats ne sont pas présentés car peu d'arbres ont été utilisés et le nombre de glands produit est faible du fait de la jeunesse des individus.

**Tableau 20** : Arbres mères utilisés dans les croisements contrôlés

Identifiant	Espèce	Origine	Statut
3P	pédonculé	Pierroton	greffe
11P	pédonculé	Pierroton	greffe
A3	pédonculé	Arcachon	greffe
A4	pédonculé	Arcachon	greffe
QS11	sessile	Orléans	greffe
QS27	sessile	Orléans	greffe
QS30	sessile	Orléans	greffe
QS32	sessile	Orléans	greffe
Bra2	pubescent	Branne	adulte
Bra55	pubescent	Branne	adulte
Bra56	pubescent	Branne	adulte
T9	tauzin	Cestas	adulte
T29	tauzin	Cestas	adulte
T30	tauzin	Cestas	adulte

#### 6.2.2.2. Arbres pères

Nous avons utilisé du pollen de 5 chênes par espèces soit 20 chênes au total. Nous avons utilisé des stocks de pollen de chêne pédonculé et sessile préexistant alors que pour les chênes tauzins et pubescent, qui ne sont pas couramment utilisés dans les croisements contrôlés, nous avons récolté du pollen en 2005. Du pollen a également été récolté sur certains chênes hybrides. Le Tableau 21 présente les chênes pères utilisés lors des croisements.

**Tableau 21** : Origine du pollen utilisé dans les croisements contrôlés, viabilité du pollen et proportion dans les mélanges en 2005 et 2006. (n. d. : non déterminé, le mélange a été fait à partir de pollen frais sans mesure de viabilité).

Mélange pollen	Individu	Espèce	Origine	Récolte	Viabilité 2005	Proportion 2005	Viabilité 2006	Proportion 2006
Pe	3P	pédonculé	Pierroton	2003	0.60	0.15	0.20	0.30
	11P	pédonculé	Pierroton	2004	0.65	0.14	0.30	0.20
	27P	pédonculé	Pierroton	2003	0.50	0.19	0.70	0.07
	A4	pédonculé	Arcachon	2004	0.25	0.37	0.20	0.28
	A6	pédonculé	Arcachon	2003	0.60	0.15	0.40	0.15
Se	QS11	sessile	Orléans	1999	0.35	0.13	0.20	0.32
	QS27	sessile	Orléans	1999	0.07	0.57	0.30	0.21
	QS30	sessile	Orléans	2004	0.45	0.10	0.50	0.13
	QS31	sessile	Orléans	1996	0.35	0.13	0.30	0.21
	QS32	sessile	Orléans	2002	0.60	0.07	0.50	0.13
Pu	Bra55	pubescent	Branne	2005	0.50	0.26	0.50	0.10
	Bra58	pubescent	Branne	2005	0.60	0.23	0.20	0.27
	PU1	pubescent	Lesparre	2005	0.80	0.17	0.20	0.27
	PU2	pubescent	Lesparre	2005	0.80	0.17	0.60	0.09
	PU7	pubescent	Lesparre	2005	0.80	0.17	0.20	0.27
Ta	T12	tauzin	Cestas	2005	n. d.	0.20	0.60	0.15
	T16	tauzin	Cestas	2005	n. d.	0.20	0.30	0.28
	T25	tauzin	Cestas	2005	n. d.	0.20	0.30	0.28
	T26	tauzin	Cestas	2005	n. d.	0.20	0.70	0.12
	T27	tauzin	Cestas	2005	n. d.	0.20	0.50	0.17

### 6.2.3. Plan de croisement

#### 6.2.3.1. Mélanges polliniques

Le pollen de 5 chênes d'une même espèce ont été mélangés afin de polliniser chaque arbre mère avec plusieurs pères. Ce mélange permet de limiter les effets éventuels d'incompatibilités individuelles. Quatre lots de pollen monospécifiques (un lot par espèce) ont été constitués (Tableau 21) : Pe (pollen des 5 chênes pédonculés), Se (pollen des 5 chênes sessiles), Pu (pollen des 5 chênes pubescents) et Ta (pollen des 5 chênes tauzins). Afin de corriger les différences individuelles de viabilité pollinique, la quantité du pollen de chaque individu a été pondérée par son taux de viabilité (Tableau 21). Enfin, un mélange de pollen des quatre espèces de chênes (multispécifique, noté Me) a été réalisé en associant une quantité égale des quatre mélanges précédents ( $Me = Pe + Se + Pu + Ta$ ).

#### 6.2.3.2. Plan de croisement

Les croisements ont été organisés de telle façon que chaque arbre mère reçoive chaque mélange de pollen au moins une fois. Les croisements ont été réalisés sur deux années par Guy Roussel, certains croisements peu fructueux la première année ont été répétés la seconde année. Au total, 2963 glands ont été obtenus (Tableau 22).

Tableau 22 : Nombre de glands obtenus par année (nombre de répétitions).

Arbre mère	Année	Mélanges de pollen				
		Pe	Se	Pu	Ta	Me
3P	2005	92 (1)	73 (2)			
	2006		9 (1)	14 (1)	20 (1)	19 (1)
11P	2005	40 (1)	430 (2)			
	2006			54 (1)	1 (2)	814 (1)
A4	2005	140 (1)	209 (3)			
	2006			0 (1)	0 (2)	40 (1)
A3	2005					
	2006					318 (4)
QS11	2005	0 (1)	14 (1)			
	2006			51 (1)	0 (1)	
QS30	2005	1 (3)	1 (2)			
	2006			0 (2)	0 (2)	43 (1)
QS32	2005	23 (3)	27 (2)			
	2006	0 (1)		29 (2)	0 (2)	0 (1)
QS27	2005					
	2006					10 (3)
Bra2	2005					
	2006	0 (12)	0 (12)	1 (12)	4 (12)	1 (12)
Bra55	2005					
	2006	0 (13)	1 (12)	4 (13)	4 (12)	0 (12)
Bra56	2005					
	2006	1 (12)	0 (12)	0 (12)	0 (12)	0 (11)
T9	2005	5 (10)	12 (9)	3 (11)	17 (20)	
	2006			3 (12)	1 (12)	7 (12)
T29	2005	14 (11)	35 (15)	30 (11)	77 (16)	
	2006					30 (12)
T30	2005	41 (13)	21 (13)	34 (15)	9 (14)	
	2006	5 (12)				12 (12)

### 6.2.3.3. Suivi des descendants

L'ensemble des glands obtenus (à l'exception des croisements A3xMe et 11PxMe, très fructueux pour lesquels 240 glands ont été sélectionnés au hasard) ont été traités de la même façon. Chaque gland est identifié, sectionné à la base afin de prélever un morceau de tissu pour des extractions d'ADN ultérieures, collé sur une plaque en plastique pouvant contenir 60 glands. Une fois remplies, les plaques sont photographiées et mises dans de la vermiculite humide. Toutes les semaines, les plaques sont inspectées et les glands ayant germé sont plantés dans des pots de terre de 4L numérotés individuellement et placés sous serre. La hauteur des plantules est mesurée régulièrement.

## 6.2.4. Analyses génétiques

### 6.2.4.1. Extraction d'ADN

Un protocole d'extraction d'ADN pour méga-gamétophyte de pin (Bousquet *et al.*, 1990) a été adapté pour extraire en plaques de 96 puits, ce qui a permis d'obtenir l'ADN de 192 glands par

jour (François Hubert, contrat de main d'œuvre occasionnelle, 2006). Les tissus de réserve contenant beaucoup de glyco-protéines et de sucres, l'extraction comporte plusieurs étapes utilisant de la protéinase K, du béta-mercapto-éthanol, du chlorophorm-isoamylalcool, de l'isopropanol puis plusieurs lavages à l'éthanol. Cette technique permet d'obtenir de l'ADN suffisamment propre pour les réactions d'amplification PCR à partir de tissu collecté sur des glands de chênes.

#### 6.2.4.2. Amplification de microsatellites

Cinq microsatellites (QpZAG110, QrZAG11, QrZAG112, QrZAG39, QrZAG96) sont amplifiés grâce au protocole décrit précédemment (chapitre 2 et ANNEXE A, Lepais *et al.*, 2006b). Le polymorphisme pour ces microsatellites est tel que chaque parent possède un génotype unique ce qui permet de déterminer les parents de chaque gland sans ambiguïté.

#### 6.2.4.3. Identification des parents

La première étape consiste à vérifier que les descendants possèdent un des deux allèles de la mère pour chaque marqueur, ce qui permet de vérifier d'éventuelles erreurs d'identification. Les pères sont identifiés grâce aux allèles paternels aux différents locus. Si des allèles absents chez les pères sont détectés, le croisement est attribué à un pollen d'origine extérieure (pollution).

#### 6.2.5. Analyses statistiques

L'isolement reproducteur total entre espèces peut être lié à des barrières prézygotiques (isolement géographique et écologique, compatibilité génétique, compétition pollinique) et postzygotique (germination, survie des plantules, fertilité). Les valeurs d'isolement reproducteur (*RI*) de chaque composante de l'isolement spécifient l'importance de chaque barrière (Coyne & Orr, 1997).

Les résultats des croisements contrôlés nous permettent de quantifier l'isolement reproducteur (*RI*) de quatre barrières reproductives entre espèces :

- **Barrières prézygotiques :**

- Compatibilité génétique : pour chaque arbre mère, les proportions de glands obtenus par croisements intraspécifiques et interspécifiques pour les croisements monospécifiques permettent d'estimer l'isolement reproducteur lié à l'incompatibilité génétique ( $RI_{comp\text{genet}}$ ). Cette composante est calculée de la façon suivante :

$$RI_{\text{genetcomp}} = 1 - \frac{\text{Contribution Paternelle Interspécifique}(\%)}{\text{Contribution Paternelle Intraspécifique}(\%)}$$

Ce paramètre prend une valeur de 1 lorsque l'incompatibilité génétique est totale (aucun gland hybride produit) et de 0 lorsque il n'y a pas d'incompatibilité génétique entre espèce (autant de glands produits par croisement interspécifique et intraspécifique). Il peut éventuellement être négatif lorsque davantage de glands sont produits par des croisements interspécifiques. Dans le cas de croisement

interspécifique, le nombre de glands obtenus peut être réduit à causes d'interaction génétiques incompatibles entre le pollen et le pistil. Pour arriver à féconder l'ovule, le grain de pollen doit en effet réussir plusieurs étapes de contrôles impliquant plusieurs mécanismes de reconnaissance (Johnson & Preuss, 2002 ; Swanson *et al.*, 2004) : capture du pollen, liaison pollen stigmate, hydratation du pollen ; germination, adhésion, élongation, guidage et réception du tube pollinique. Chacune de ces étapes nécessite l'interaction entre plusieurs gènes identifiés chez des espèces modèles (Mayfield *et al.*, 2001 ; Hancock *et al.*, 2003 ; Swanson *et al.*, 2004 ; Clark *et al.*, 2006). L'incompatibilité génétique entre pollen et stigmate se traduit par une absence de germination du pollen ou une destruction du tube pollinique et finalement la réduction du nombre de glands produits. Il faut cependant noter que la production de glands peut être aussi réduite à cause d'avortements post-fécondations (postzygotiques) qui, n'ayant pas été quantifiés expérimentalement, sont inclus ici dans l'isolement génétique.

- Compétition pollinique : la proportion de glands hybrides rapportée à la proportion de glands non hybrides pour chaque croisement multispécifique (Me) est utilisée pour calculer l'isolement reproducteur lié à la compétition pollinique ( $RI_{compoll}$ ) :

$$RI_{pollcomp} = 1 - \frac{\text{Contribution Paternelle Interspécifique}(\%)}{\text{Contribution Paternelle Intrasécifiques}(\%)}$$

La valeur de l'isolement reproducteur en condition de compétition pollinique vaut 1 lorsqu'aucun gland hybride n'est produit et 0 lorsque autant de glands hybrides et non hybrides sont produits. Ce paramètre peut avoir une valeur négative lorsque le nombre de glands hybrides produits dépasse celui des non hybrides. Cette mesure de l'isolement inclut l'isolement lié à l'incompatibilité (détaillée ci-dessus) ainsi que l'effet de la compétition pollinique. La présence de pollen de différentes espèces simultanément sur le stigmate peut avoir des effets contrastés. La présence de pollen conspécifique peut réduire l'efficacité des mécanismes d'isolements génétiques (on parle de *mentor effect*) en déclenchant les conditions nécessaires à la croissance normale du pollen hétérospécifique conduisant à l'augmentation du nombre d'hybrides produits. Au contraire, la compétition peut permettre au pistil de favoriser la réussite du pollen conspécifique (on parle de *conspécific pollen precedence*). Dans ce cas, la quantité d'hybrides formés va être réduite (par exemple croissance du tube pollinique conspécifique plus rapide).



- **Barrières postzygotiques** : du fait du long temps de génération des chênes, la viabilité des hybrides ne peut être estimée que pour des stades de vie précoces. Nous estimons cette viabilité grâce à deux caractéristiques : la germination et la survie à l'issue de la première année de croissance. La viabilité de chaque type d'hybride est ainsi estimée par rapport aux individus non hybrides pour ces deux paramètres :

$$\begin{aligned} \circ \quad RI_{germ} &= 1 - \frac{TauxGerminationHybrides}{TauxGerminationIntraspecifices} \\ \circ \quad RI_{survie} &= 1 - \frac{TauxSurvieHybrides}{TauxSurvieIntraspecifices} \end{aligned}$$

Ces deux paramètres sont estimés en rassemblant tous les types de croisements. Les taux de germination et de survie sont d'abord calculés par arbre mère, puis leur moyenne par espèce mère est utilisée pour estimer l'isolement reproducteur.

L'ensemble de ces indices d'isolement reproducteur ( $RI$ ) permettent d'estimer l'isolement reproducteur total ( $T$ ) et la contribution relative ( $RC$ ) de chaque barrière grâce à une feuille de calcul Excel disponible en ligne (Ramsey *et al.*, 2003).

Pour chaque stade de vie, la contribution absolue ( $AC$ ) de chaque composante de l'isolement reproducteur peut être estimée de la façon suivante (Ramsey *et al.*, 2003) :

$$\begin{aligned} AC_1 &= RI_1 \\ AC_2 &= RI_2 (1 - AC_1) \\ AC_3 &= RI_3 [1 - (AC_1 + AC_2)] \end{aligned}$$

pour trois stades de vie, et plus généralement, pour  $n$  stades de vie :

$$AC_n = RI_n \left(1 - \sum_{i=1}^{n-1} AC_i\right)$$

Une barrière reproductive donnée élimine une partie du flux de gènes qui n'a pas été déjà éliminé par une barrière reproductive intervenant précédemment. Donc, pour  $m$  barrières reproductives, l'isolement reproducteur total ( $T$ ) :

$$T = \sum_{i=1}^m AC_i$$

La contribution relative de chaque barrière dans l'isolement reproducteur total est calculée par :

$$RC_n = \frac{AC_n}{T}$$

## 6.3. Résultats

### 6.3.1. Identification des pères

Les résultats d'identification des pères au niveau individuel sont reportés en annexe (ANNEXE A). Un total de 2292 individus a été analysé dont 1724 (75%) ont un père identifié. Les 568 autres individus (25%) correspondent à 374 individus issus d'un événement de reproduction avec du

pollen extérieur (pollution) et 194 individus n'ayant pas pu être génotypé (échec lors de l'extraction d'ADN ou amplification partielle des microsatellites). L'identification de pollen extérieur à l'expérience indique que les poches déployées autour des arbres mères ne sont pas totalement imperméables au pollen. Les analyses suivantes utilisent les individus dont le père a été identifié et est conforme au pollen injecté. Les croisements effectués sur chênes pubescent ne sont pas intégrés, il n'ont en effet donné qu'un nombre très limité de glands (ANNEXE A).

### 6.3.2. Compatibilité génétique

La comparaison du nombre de glands obtenus par croisement pour des pollinisations conspécifiques et hétérospécifiques permet d'estimer l'impact de la compatibilité génétique intrinsèque comme barrière à l'hybridation (voir chapitre 1). Cette barrière reproductive englobe plusieurs phénomènes dépendant directement de l'interaction entre les génotypes maternel et paternel (Swanson *et al.*, 2004).

Tableau 23 : Nombre de glands, dont la paternité a été déterminée par analyse génétique, obtenus par croisements monospécifiques (nombre de croisements) et contributions paternelles. La contribution paternelle moyenne par espèce mère est indiquée en gras, les contributions paternelles pour les croisements intraspécifiques sont soulignées.

Espèce mère	Arbres mères	Nbn. glands / espèces pollen (N. crois.)				Contributions paternelles			
		Ped	Ses	Pub	Toz	Ped	Ses	Pub	Toz
<b>Ped</b>	11P	4 (1)	32 (2)	54 (1)	0 (2)	<u>5.4%</u>	21.6%	73.0%	0.0%
	A4	89 (1)	183 (3)	0 (1)	0 (2)	<u>59.3%</u>	40.7%	0.0%	0.0%
	3P	64 (1)	54 (3)	13 (1)	18 (1)	<u>56.6%</u>	15.9%	11.5%	15.9%
						<b>moyenne 40.5%</b>	<b>26.1%</b>	<b>28.2%</b>	<b>5.3%</b>
<b>Ses</b>	QS32	0 (4)	22 (2)	26 (2)	0 (2)	0.0%	<u>45.8%</u>	54.2%	0.0%
	QS11	0 (1)	13 (1)	49 (1)	0 (1)	0.0%	<u>21.0%</u>	79.0%	0.0%
						<b>moyenne 0.0%</b>	<b>33.4%</b>	<b>66.6%</b>	<b>0.0%</b>
<b>Toz</b>	T9	4 (10)	10 (9)	5 (23)	1 (32)	22.7%	63.1%	12.4%	<u>1.8%</u>
	T29	12 (11)	11 (15)	8 (11)	59 (16)	17.5%	11.8%	11.7%	<u>59.1%</u>
	T30	30 (25)	19 (13)	21 (15)	2 (14)	28.5%	34.8%	33.3%	<u>3.4%</u>
						<b>moyenne 22.9%</b>	<b>36.6%</b>	<b>19.1%</b>	<b>21.4%</b>

Ped : pédonculé, Ses : sessile, Pub : pubescent, Toz : tauzin.

Quand tous les croisements monospécifiques sont considérés, les arbres mères pédonculé produisent 40% de glands issus de pollen pédonculé (Tableau 23 et Figure 41A). Cette espèce semble capable de produire des glands hybrides avec toutes les espèces de pollen. Cependant, le nombre moyen de glands par croisement diminue avec du pollen hétérospécifique : il est plus faible avec du pollen de chêne sessile et pubescent et nettement réduit avec du pollen de chêne tauzin.

Les arbres mères sessiles produisent un nombre de glands plus faible que les mères pédonculées lorsque du pollen conspécifique est injecté (Tableau 23). Cette espèce semble incapable de produire les glands hybrides avec du pollen de chêne pédonculé et tauzin. En revanche, le pollen

de chêne pubescent semble particulièrement efficace car la proportion de glands issus de pollen pubescent (66.6% des contributions, Tableau 23 et Figure 42A) est supérieure à celle des glands issus de pollen conspécifique (33.3% des contributions).

Enfin, les croisements sur les arbres mères tauzin produisent approximativement la même proportion de glands issus des différents pollen (Tableau 23 et Figure 43A). Cependant le pollen sessile est légèrement plus efficace (36.6% des contributions) que le pollen conspécifique (21.4% des contributions).

En dehors des arbres mères sessiles qui ne sont pas fécondés par du pollen de chêne pédonculé et pubescent, toutes les autres combinaisons de croisements ont donné des glands hybrides. L'incompatibilité génétique représente donc une barrière reproductive totale pour seulement deux types de croisements (sessile x pédonculé et sessile x tauzin) alors qu'elle restreint partiellement les croisements interspécifiques chez les arbres mères pédonculés et tauzins. L'incompatibilité génétique n'intervient pas dans les croisements sessile x pubescent et tauzin x sessile puisqu'ils produisent en moyenne plus de glands que les croisements intraspécifiques.

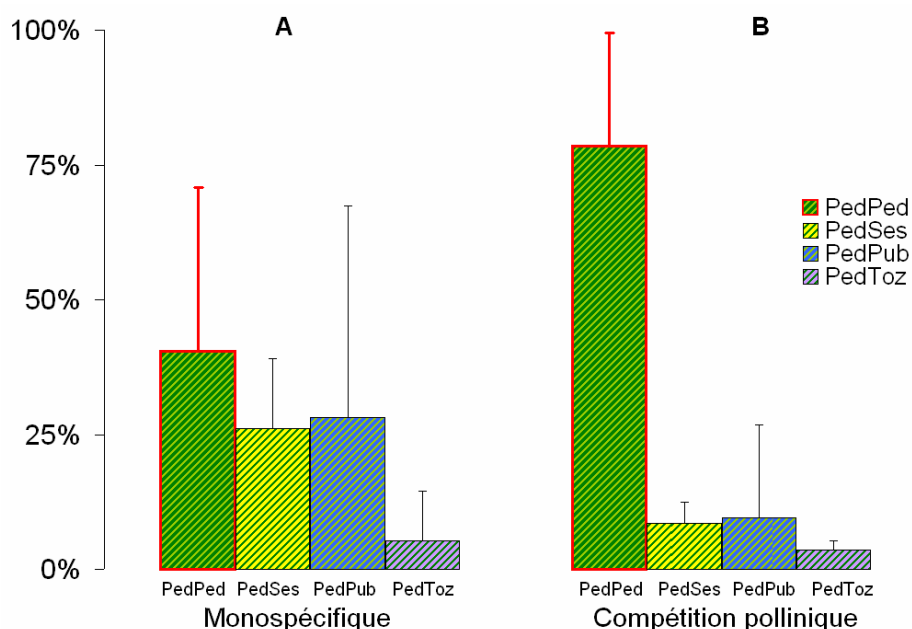


Figure 41 : Contributions paternelles (%) pour différentes espèces de pollen lors des croisements monospécifiques (A) et multispécifiques (B) réalisés sur les arbres mères pédonculés. La bordure rouge indique les croisements intraspécifiques. L'identifiant en abscisse correspond à l'espèce de la mère (Ped : pédonculé) suivi de l'espèce du pollen (Ses : sessile, Pub : pubescent et Toz : tauzin).

### 6.3.3. Compétition pollinique

L'application simultanée et en proportion égale de pollen de quatre espèces aux différents arbres mères permet d'estimer l'impact de la préférence donnée au pollen conspécifique comme barrière reproductive prézygotique, en étudiant la sélection du pollen dans le cas où plusieurs types de pollen se retrouvent simultanément à la surface du même stigmate. Cette compétition peut réduire les croisements interspécifiques et agir comme barrière reproductive. Le pollen hétérosppécifique croît moins vite dans certains cas que le pollen conspécifique, ce dernier est alors plus efficace

pour la fécondation (Carney *et al.*, 1996). À l'inverse, la présence de pollen conspécifique sur le stigmate peut inhiber les mécanismes de destruction de pollen hétérosppécifique ou fournir les conditions favorables à la croissance des tubes polliniques. Dans ce cas, le taux de fécondation intersppécifique peut être augmenté car les mécanismes d'isolements reproducteurs sont inhibés par la présence de pollen conspécifique (Gaget *et al.*, 1989; Knox *et al.*, 1972; Pandey, 1977). Ce phénomène, nommé *mentor effect*, peut aussi augmenter le taux d'autofécondation chez les espèces normalement autoincompatible.

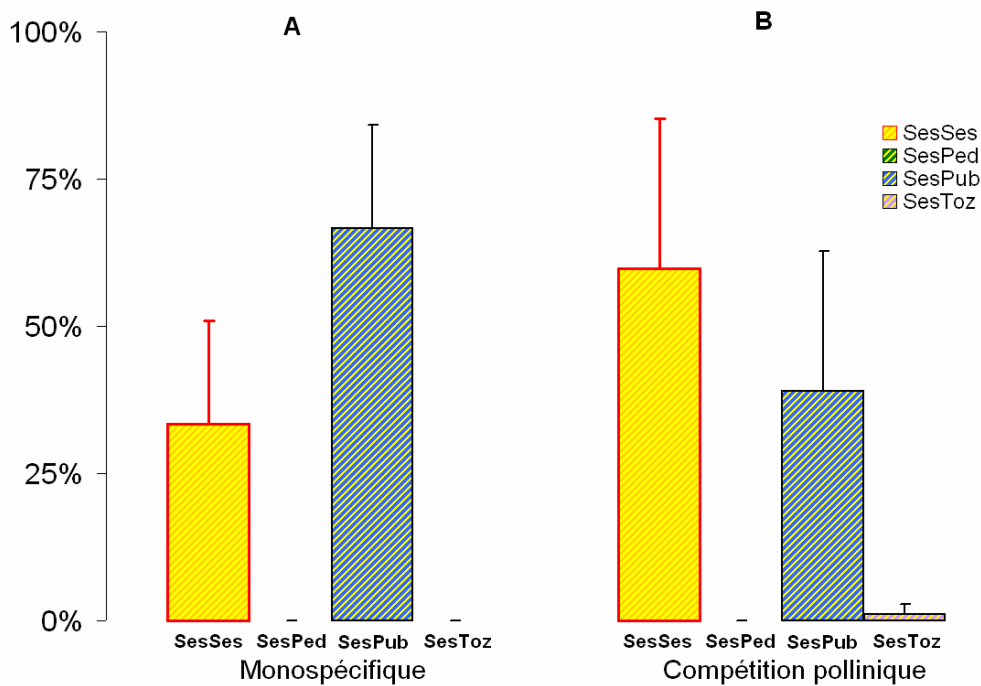


Figure 42 : Contributions paternelles (%) pour différentes espèces de pollen lors des croisements monospécifiques (A) et multispécifiques (B) réalisés sur les arbres mères sessiles. La bordure rouge indique les croisements intraspécifiques. L'identifiant en abscisse correspond à l'espèce de la mère (Ses : sessile) suivi de l'espèce du pollen (Ped : pédonculé, Pub : pubescent et Toz : tauzin).

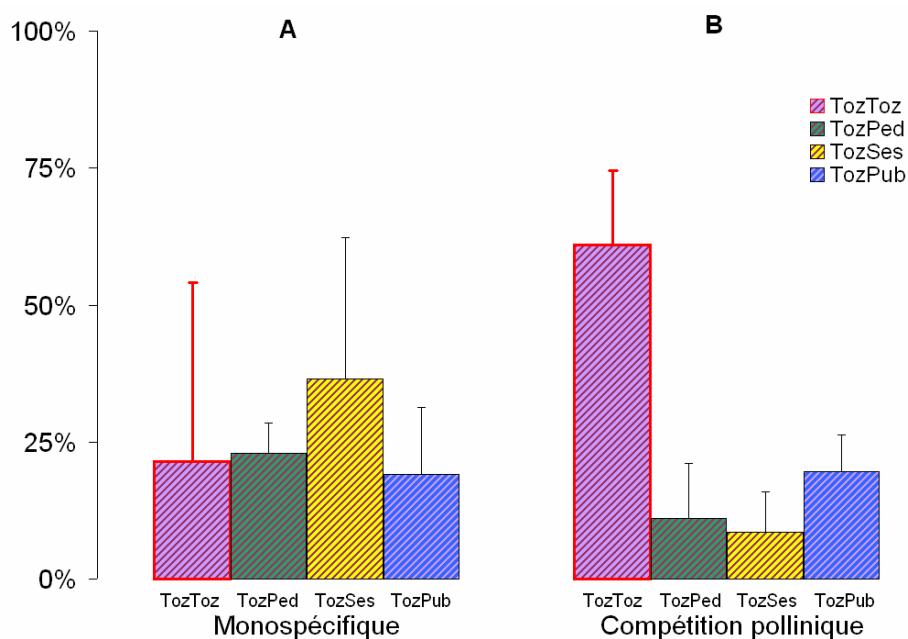


Figure 43 : Contributions paternelles (%) pour différentes espèces de pollen lors des croisements monospécifiques (A) et multispécifiques (B) réalisés sur les arbres mères tauzins. La bordure rouge indique les croisements intraspécifiques. L'identifiant en abscisse correspond à l'espèce de la mère (Toz : tauzin) suivi de l'espèce du pollen (Ped : pédonculé, Ses : sessile et Pub : pubescent).

Tableau 24 : Nombre de glands, dont la paternité a été déterminée par analyse génétique, obtenus par croisements multispécifiques (mélange de pollen Me) et contributions paternelles. La contribution paternelle moyenne par espèce mère est indiquée en gras, les contributions paternelles pour les croisements intraspécifiques sont soulignées.

Espèce mère	Arbre mère	N	Nbn. glands / espèces pollen				Contributions paternelles			
			Ped	Ses	Pub	Toz	Ped	Ses	Pub	Toz
<b>Ped</b>	11P	1	101	3	3	4	<u>91.0%</u>	2.7%	2.7%	3.6%
	A4	1	34	4	0	1	<u>87.2%</u>	10.3%	0.0%	2.6%
	3P	1	8	2	6	1	<u>47.1%</u>	11.8%	35.3%	5.9%
	A3	4	154	16	0	4	<u>88.5%</u>	9.2%	0.0%	2.3%
			<b>moyenne</b>				<b>78.4%</b>	<b>8.5%</b>	<b>9.5%</b>	<b>3.6%</b>
<b>Ses</b>	QS30	1	0	18	24	1	0.0%	<u>41.9%</u>	55.8%	2.3%
	QS27	3	0	7	2	0	0.0%	<u>77.8%</u>	22.2%	0.0%
				<b>moyenne</b>				<b>0.0%</b>	<b>59.8%</b>	<b>39.0%</b>
<b>Toz</b>	T9	12	1	0	1	3	20.0%	0.0%	20.0%	<u>60.0%</u>
	T29	12	3	3	6	11	13.0%	13.0%	26.1%	<u>47.8%</u>
	T30	12	0	1	1	6	0.0%	12.5%	12.5%	<u>75.0%</u>
				<b>moyenne</b>				<b>11.0%</b>	<b>8.5%</b>	<b>19.5%</b>

N : nombre de répétitions du croisement. Ped : pédonculé, Ses : sessile, Pub : pubescent, Toz : tauzin.

Dans nos conditions de compétition pollinique, la proportion d'hybrides formée est globalement inférieure comparée aux croisements sans compétition pollinique (comparer A et B Figure 41, Figure 42 et Figure 43). La compétition pollinique chez les chênes représente donc bien une barrière reproductive, il ne semble pas y avoir de "mentor effect" lié à la présence simultanée de

pollen de plusieurs espèces. On retrouve cependant des croisements interspécifiques, indiquant que la préférence au pollen conspécifique est une barrière reproductive partielle qui limite, mais n'empêche pas, la formation d'hybrides.

Les arbres mères pédonculés produisent un nombre de glands non hybrides comparable aux expériences précédentes, alors que le nombre d'hybrides formé est réduit à quelques individus par croisements (Tableau 24). Le taux d'hybridation global pour les arbres mères pédonculés est de 13% (44 descendants hybrides pour 297 descendants pédonculés).

La proportion de glands hybrides formés sur les chênes sessiles est plus faible en condition de compétition pollinique par rapport aux croisements sans compétition (Figure 42B), mais reste importante (39%, Tableau 24). Les arbres mères sessiles rejettent le pollen de chêne pédonculé et tauzin (bien qu'un descendant sessile x tauzin ait été identifié dans ces croisements, contrairement aux croisements précédents). Le taux d'hybridation sur les mères sessiles est de 52% (27 descendants hybrides pour 25 descendants sessiles).

Le nombre de glands par croisements formés en compétition pollinique sur les chênes tauzins est globalement divisé par deux pour tous les types de descendance (Tableau 24). Ce résultat n'est peut-être pas seulement lié à la compétition pollinique puisque ces croisements ont été réalisés en 2006 alors que les premiers datent de l'année 2005. Les arbres mères tauzins utilisées étant des adultes localisés dans un parc, les conditions climatiques et plus généralement environnementales peuvent modifier la productivité de ces arbres d'une année à l'autre. Néanmoins, la proportion de descendants hybrides est inférieure au nombre de descendants tauzins (Figure 43B). La préférence au pollen conspécifique semble intervenir modérément puisque il y a proportionnellement moins d'hybrides formés en condition de compétition pollinique comparé aux croisements sans compétition. Le taux d'hybridation pour ces arbres mères tauzins est de 44% (16 descendants hybrides pour 20 descendants tauzins).

#### 6.3.4. Taux de germination et de survie

L'ensemble des descendants est pris en compte afin de comparer le taux de germination et de survie des hybrides par rapport aux non hybrides.

Globalement le taux de germination est similaire pour les glands issus des arbres mères (greffes) pédonculés et sessiles alors qu'il est plus faible pour les glands issus des arbres mères (arbre adulte de parc) tauzins (Figure 44). Les conditions physiologiques des arbres mères pourraient expliquer cette différence puisque les greffes sont irriguées et fertilisées alors que les chênes tauzins dépendent des conditions du milieu (pas d'irrigation ni de fertilisation). La même tendance émerge pour la survie bien que la différence soit moins nette (Figure 45).

Les glands hybrides issus des mères pédonculés ont un taux de germination comparable aux glands non hybrides (environ 85%, Figure 44), sauf pour les glands pédonculé x tauzin qui présentent un taux de germination inférieur (environ 70%). Le taux de survie suit la même tendance (environ 80% pour les plantules pédonculés et hybrides pédonculé x sessile ; environ 60% de survie pour les plantules hybrides pédonculé x tauzin) bien que les hybrides pédonculé x

pubescent présentent une survie légèrement supérieure (environ 95%) mais non significative (Figure 45).

Les glands hybrides sessile x pubescent présentent un meilleur taux de germination (environ 90%) que les glands sessiles (environ 80%). La même tendance, non significative, se retrouve pour le taux de survie (Figure 45).

Les glands hybrides issus de croisements sur les arbres mères tauzins présentent des taux de germination inférieurs aux glands issus de croisement intraspécifique (environ 20% contre 50%, Figure 45). Cependant, le taux de survie montre des résultats inverses (Figure 45) : les plantules hybrides ont de meilleurs taux de survie comparés aux plantules tauzins (environ 40%). Cette tendance à une meilleure survie des hybrides n'est significative que pour les hybrides tauzin x sessile.

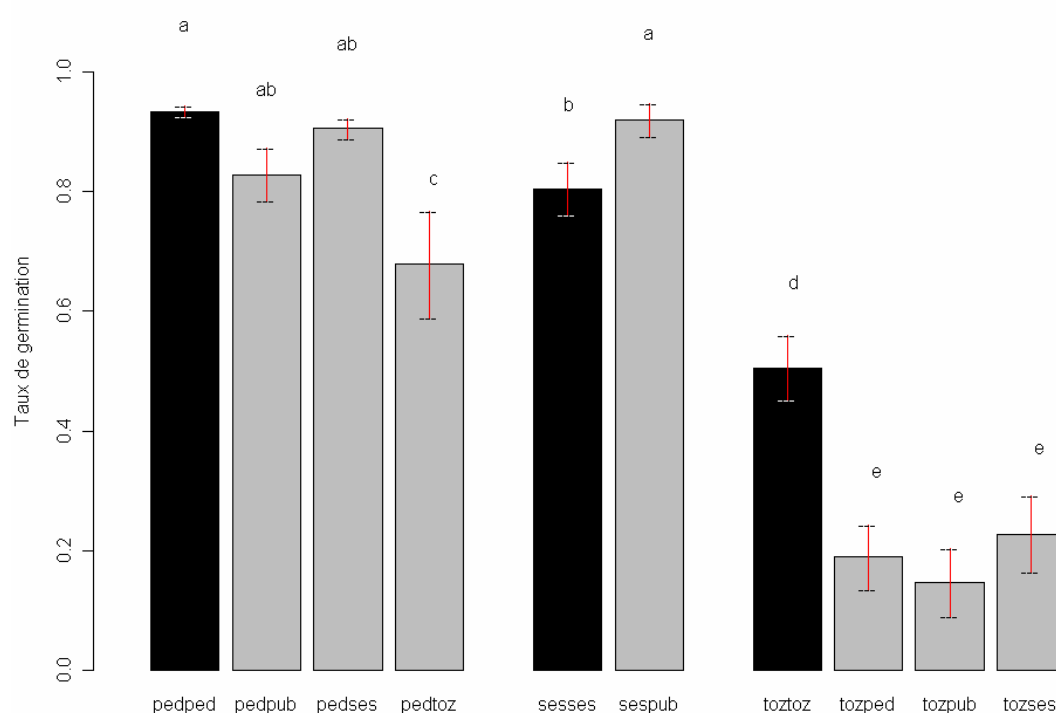


Figure 44 : Taux de germination par type de croisement. Le type de croisement est indiqué en abscisse, la première espèce est l'espèce de la mère, la seconde espèce est celle du père. Les histogrammes noirs indiquent les croisements intraspécifiques, les gris les croisements interspécifiques. Des lettres différentes indiquent une différence significative au seuil 5% obtenue par un test T de Student (comparaison multiple par Least Significant Differences).

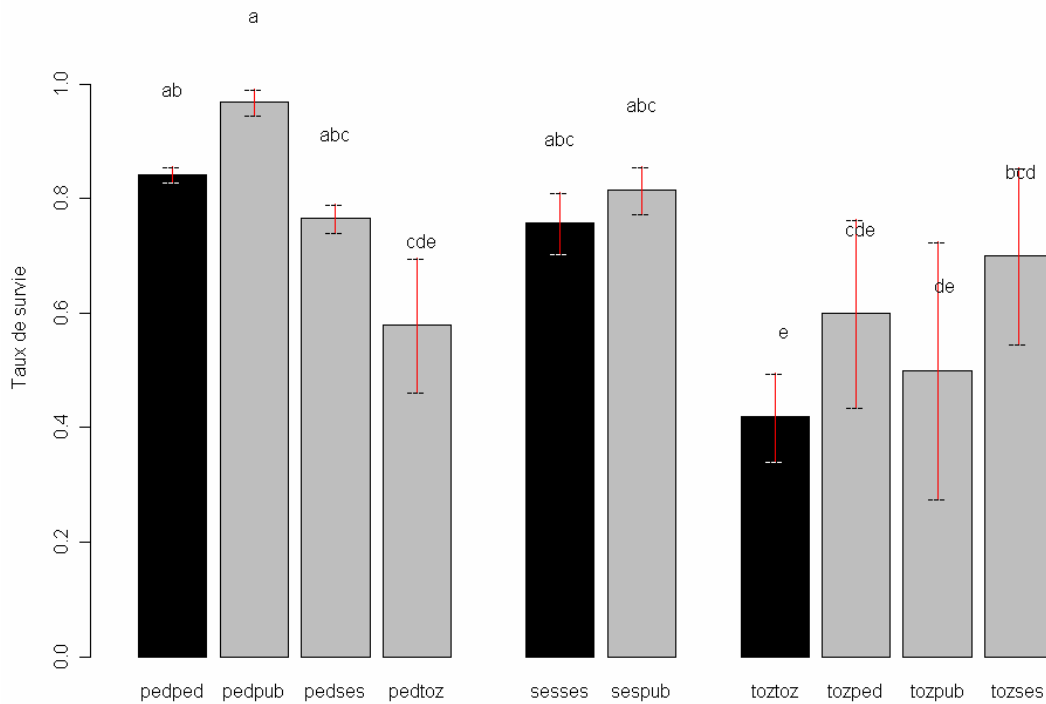


Figure 45 : Taux de survie par type de croisement. Le type de croisement est indiqué en abscisse, la première espèce est l'espèce de la mère, la seconde espèce est celle du père. Les histogrammes noirs indiquent les croisements intraspécifiques, les gris les croisements interspécifiques. Des lettres différentes indiquent une différence significative au seuil 5% obtenue par un test T de Student (comparaison multiple par Least Significant Differences).

### 6.3.5. Isolement reproducteur

Les résultats présentés précédemment nous ont permis de calculer pour chaque type de croisement et chaque barrière reproductrice potentielle les indices d'isolement reproducteur ( $RI$ ) correspondants en contrastant les résultats des croisements interspécifiques à ceux des croisements intraspécifiques. La contribution relative ( $RC$ ) de chaque barrière reproductrice ainsi que l'isolement reproducteur total ( $T$ ) sont ensuite estimés (Tableau 25).

En moyenne, la contribution relative des barrières prézygotiques est plus forte ( $RC_{genetcomp} = 0.21$  et  $RC_{pollcomp} = 0.55$ ) que celle des barrières postzygotiques ( $RC_{germ} = 0.03$  et  $RC_{survie} = -0.03$ ). La compatibilité génétique entre espèces ainsi que la compétition pollinique jouent un rôle important pour réduire les flux de gènes interspécifiques. Les barrières postzygotiques ici prises en comptes jouent un rôle faible voire nul dans l'isolement reproducteur : la germination des hybrides est parfois inférieure mais leur survie est souvent meilleure ( $RC_{survie}$  négatif). On note cependant de grandes différences en fonction de l'espèce des arbres mères, en particulier pour les mères sessiles. En effet, nos résultats montrent un isolement reproducteur asymétrique pour cette espèce. Les ovules de chênes sessiles présentent une incompatibilité génétique totale avec le pollen des chênes pédonculé et taurin ( $T_{Ses \leftarrow Ped} = T_{Ses \leftarrow Toz} = 1$ ). A l'inverse, le pollen de chênes sessile est capable de féconder des chênes pédonculés et taurins ( $T_{Ped \leftarrow Ses} = 0.93$  et  $T_{Toz \leftarrow Ses} = 0.59$ ). Cette caractéristique de flux de pollen asymétrique peut créer en conditions naturelles une hybridation directionnelle, qui si elle persiste au fil des générations d'hybrides, peut conduire au retour d'individus sessiles par rétrocroisement successifs et unidirectionnels. Ce résultat est en



accord avec les croisements contrôlés précédemment réalisés entre chênes pédonculé et sessile montrant que le pollen de chêne pédonculé réussit très peu à féconder les chênes sessiles alors que le pollen sessile produit une descendance nombreuse sur des chênes pédonculés (Kleinschmit & Kleinschmit, 2000; Steinhoff, 1993). Nos résultats indiquent que cette asymétrie existe aussi entre les chênes sessiles et tauzins.

De plus, le pollen de chêne pubescent est remarquablement efficace sur les arbres mères sessiles puisque deux contributions relatives ont des valeurs négatives, démontrant un avantage aux hybrides sessile x pubescent comparativement aux croisements intraspécifique ( $T_{Ses \times Pub} = -0.31$ ). Dans ce cas, bien que nous ne disposions pas de résultats de croisements sur des arbres mères pubescents, l'hybridation risque d'être asymétrique conduisant potentiellement à une introgression directionnelle du chêne pubescent d'autant plus rapide que l'hybridation entre ces deux espèces semble favorisée.

Pour les autres croisements, les barrières prézygotiques réduisent partiellement les flux de gènes interspécifiques. Les arbres mères pédonculés présentent un isolement reproducteur total assez fort ( $T$  de 0.93 à 0.99) alors que l'isolement est nettement plus faible chez les arbres mères tauzin ( $T$  de 0.59 à 0.85). La compétition pollinique contribue plus fortement à l'isolement reproducteur que la compatibilité génétique (sauf pour les cas d'incompatibilités génétique totales discuté précédemment ainsi que les croisements pédonculé x tauzin, Tableau 25).

Tableau 25 : Estimation de l'isolement reproducteur de quatre barrières reproductives ( $R$ ) ainsi que leur contribution relative ( $RC$ ) dans l'isolement reproducteur total ( $T$ ) entre paires d'espèces.

Stade	Composante individuelle ( $R$ )				Contribution relative ( $RC$ )				Total ( $T$ )
	Prézygotique		Postzygotique		Prézygotique		Postzygotique		
Barrière	Génétique	Compétition	Germination	Survie	Génétique	Compétition	Germination	Survie	
Ped/Ses	0.36	0.89	0.04	0.03	0.36	0.57	0	0	0.93
Ped/Pub	0.30	0.88	0.27	-0.15	0.30	0.61	0.02	-0.01	0.93
Ped/Toz	0.87	0.95	0.05	0.29	0.87	0.13	0	0	0.99
Ses/Ped	1	1			1	0	0	0	1
Ses/Pub	-0.99	0.35	0.05	-0.06	-0.99	0.69	0.07	-0.07	-0.31
Ses/Toz	1	0.98			1.00	0	0	0	1
Toz/Ped	-0.07	0.82	0.34	-0.43	-0.07	0.88	0.07	-0.05	0.82
Toz/Ses	-0.71	0.86	0.02	-0.77	-0.71	1.47	0	-0.18	0.59
Toz/Pub	0.11	0.68	0.42	0.11	0.11	0.61	0.12	0.02	0.85
Moyenne	0.21	0.82	0.17	-0.14	0.21	0.55	0.03	-0.03	0.77

## 6.4. Conclusion

Globalement, les différents croisements contrôlés nous ont permis d'obtenir des plantules pour toutes les combinaisons de croisements à l'exception des hybrides sessile x pédonculé qui n'ont jamais pu être obtenus artificiellement. Le taux d'hybridation observé, en condition de compétition pollinique, varie de 22% à 52% en fonction de l'espèce des arbres mères. La nature et la force des barrières reproductives sont également variables en fonction de l'espèce de l'arbre

mère (Tableau 24). Ces résultats présentés au niveau des espèces cachent une assez grande variabilité de comportement entre arbres mères (Tableau 23) et leur généralisation semble difficile au regard du nombre limité d'individus et d'origines géographiques utilisés dans ces croisements. Sachant que les barrières reproductives ainsi que la réussite des croisements interspécifiques peuvent être liés aux conditions environnementales (par exemple la quantité de calcium dans le sol influence la réussite de croisements interspécifiques entre deux espèces de plantes herbacées, *Phlox drummondii* et *Phlox cuspidata* ; Ruane & Donohue, 2007, 2008), une meilleure interprétation du rôle de chacune de ces barrières pourrait être établie en réalisant ce type de croisements directement dans le milieu naturel. Néanmoins, les contraintes techniques de la méthode de croisements chez les chênes rendant difficile une telle approche, il serait utile d'effectuer de tels croisements sur des greffes de chênes de plusieurs espèces provenant de la même population. Il serait également intéressant d'inclure davantage de provenances dans l'origine des individus afin de tester l'effet de l'origine géographique sur leur système de reproduction. En particulier, les résultats négatif pour les croisements sessile x pédonculé nécessiteraient d'être confirmés en utilisant d'autres provenance de chênes sessiles. Ce résultat obtenu plusieurs fois de façon indépendante dans des croisements contrôlés apparaît en effet contradictoire avec des analyses de flux de gènes en population naturelle (voir chapitre 7 et les résultats OAKFLOW) qui montrent dans certaines population une dynamique d'hybridation dans le sens inverse. D'autre part, l'analyse de l'hybridation entre chênes sessiles et tauzins dans plusieurs populations en Espagne montre une hybridation bidirectionnelle, sans asymétrie apparente (Valbuena-Carabana *et al.*, 2005; Valbuena-Carabana *et al.*, 2007). Les différences observées entre hybridation artificielle et naturelle montrent que les facteurs intervenant dans l'isolement reproducteurs sont complexes, rendant difficile la généralisation des résultats obtenus en condition contrôlées.

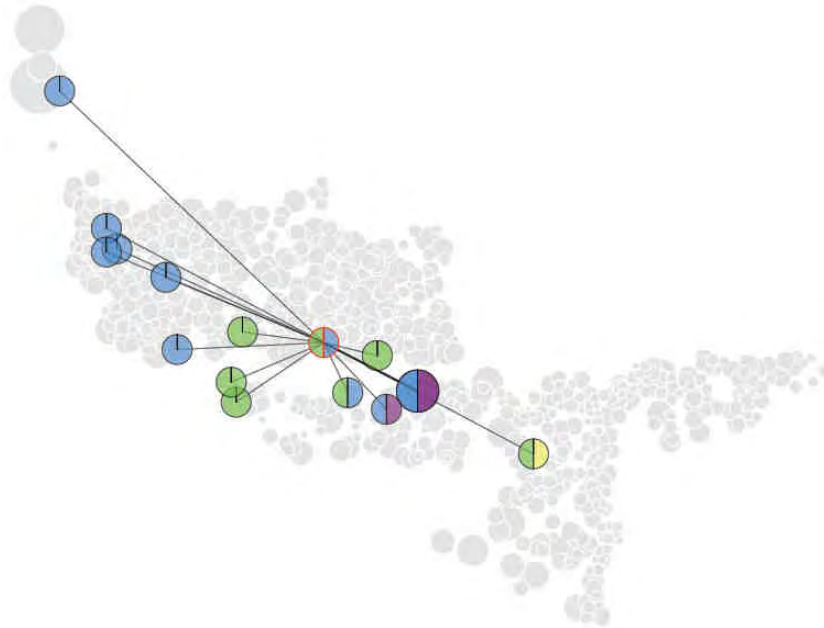
Cependant, nos expériences d'hybridations artificielles, vérifiées par marqueurs génétiques, nous ont permis de quantifier le rôle de certaines barrières reproductives. Nous avons pu montrer que la compatibilité génétique et la préférence au pollen conspécifique limite les flux de gènes entre espèces, parfois de façon asymétrique. Les chênes sont des espèces à stigmates secs ce qui nécessite de nombreuses interactions génétiques entre pollen et pistil aux diverses étapes précédant la fécondation. Il paraît donc important de réaliser des expériences de pollinisation in vitro suivies d'observations au microscope afin d'identifier les étapes clés de l'incompatibilité génétique pour les combinaisons incompatibles (sessile x pédonculé et sessile x tauzin). Bien qu'assez difficile à mettre en place chez des arbres, l'identification de ces étapes clés permettrait d'identifier des gènes candidats potentiellement impliqués dans l'isolement reproducteurs entre espèces, certains de ces gènes étant connus chez plusieurs espèces modèles (Swanson *et al.*, 2004). De plus, des techniques de marquages différentiels des pollens de plusieurs espèces pourraient être utilisées pour mieux comprendre l'origine cellulaire de la préférence au pollen conspécifique, barrière qui semble particulièrement efficace à isoler les espèces de chêne. Une

partie des objectifs de la thèse de Pierre Abadie (thèse 2007-2010 encadrée par Pauline Garnier-Géré, UMR BIOGECO) est de mettre en place de telles expériences et devrait apporter des précisions concernant la nature des différentes barrières prézygotiques.

Les barrières reproductives prises en compte dans nos analyses ne suffisent pas à isoler totalement les espèces. Ainsi, l'application stricte du concept biologique de l'espèce dans cette situation d'interfertilité devrait nous amener à définir une seule espèce composées de sous-espèces (Kleinschmit & Kleinschmit, 2000). Cependant, en milieu naturel, d'autres barrières à la reproduction peuvent intervenir comme par exemple l'isolement géographique et écologique des espèces ou la phénologie de la floraison. Ces barrières pré-reproductives interviennent avant les barrières testées ici et jouent probablement un rôle plus important dans l'isolement reproducteur des espèces (Ramsey *et al.*, 2003). De plus, la viabilité des hybrides, barrière postzygotique dont nous avons estimé un faible impact dans l'isolement reproducteur, dépend fortement des conditions de l'environnement (barrière écologique extrinsèque) qui pourraient jouer un rôle plus important en population naturelle.

Enfin, une barrière reproductrice difficilement quantifiable du fait du long temps de génération des chênes est la fertilité des hybrides. Chez certaines espèces animales et végétales la stérilité des hybrides est la source majeure d'isolement reproducteur entre espèces. Nous avons réalisé quelques croisements en utilisant des chênes hybrides (pédonculé x sessile, sessile x pubescent et tauzin x pubescent) comme arbre mère et père. Ces croisements ont produit quelques glands viables montrant que les gamètes mâles et femelles de ces hybrides de première génération peuvent produire une descendance, observation corroborée par des croisements réalisés sur un chêne hybride sessile x pédonculé (Ollrik & Kjaer, 2007).

# Chapitre 7- Flux de gènes interspécifiques en population naturelle



Page précédente : Recherche de paternité dans une descendance récoltée sur un chêne hybride (entouré en rouge).

## 7.1. Introduction

Nous avons vu au chapitre 4 que le taux d'hybridation chez les chênes blancs européens était variable entre populations mais surtout particulièrement élevé dans certaines populations, en particulier dans la parcelle de Briouant (chapitre 5) qui comprend quatre espèces en mélange. Afin d'expliquer ce taux d'hybridation élevé, de comprendre la dynamique d'hybridation et le mécanisme de maintien des espèces malgré des flux de gènes interspécifique importants, nous allons nous focaliser sur l'étude du système reproducteur de quelques chênes de cette parcelle, incluant des chênes appartenant aux quatre espèces et aux différentes classes d'hybrides.

Les chênes étant des espèces dont le pollen est dispersé par le vent (pollinisation anémophile), la période de floraison pourrait constituer une barrière pré-reproductive en empêchant les croisements interspécifiques. Si la floraison des espèces a lieu à des périodes différentes (floraison asynchrone), les possibilités d'hybridation pourraient être fortement réduites. La période de floraison pourrait être un facteur particulièrement pertinent dans notre cas puisque Briouant contient des chênes tauzins connus pour avoir un début de période végétative tardif par rapport aux autres espèces et donc probablement une floraison retardée. Nous avons donc observé les périodes de floraison des espèces et des hybrides durant deux années consécutives pour estimer le rôle de ce paramètre dans l'isolement reproducteur des espèces.

Nous avons ensuite voulu identifier a posteriori les événements de reproduction ayant eu lieu lors d'une saison de reproduction. Disposant des génotypes de l'ensemble des chênes adultes de la parcelle, nous avons analysé les descendants (glands récoltés sur plusieurs arbres mères et plantules) à l'aide de deux méthodes complémentaires. Une méthode directe par analyse de paternité (ou de parenté pour les plantules) nous a permis d'assigner pour une partie des descendants un père (pour les glands) ou un couple de parents (pour les plantules) compatibles. La détermination de l'espèce par assignation génétique des couples de parents permet alors de compter les différents types de reproduction ayant eu lieu lors de la saison de reproduction précédant l'échantillonnage. Une seconde approche, dite indirecte, a été réalisée afin de prendre en compte l'ensemble des descendants, et pas seulement ceux dont les deux parents ont pu être identifiés dans la parcelle. Des modèles de reproduction corrélées, utilisant seulement les génotypes des descendants, ont permis de caractériser d'une part la structure du pool pollinique reçu par les différents arbres mères, et d'autre part de comparer la nature du pollen accepté par ces différentes mères. Enfin, l'assignation génétique appliquée aux génotypes des descendants permet de compléter ces analyses.

## 7.2. Matériels et méthodes

### 7.2.1. Parcelle étudiée

La parcelle étudiée dans ce chapitre correspond au bois de Briouant dont la composition en espèces et hybrides a été analysée dans un chapitre précédent (voir chapitre 4). Tous les arbres adultes de

cette parcelle ont été analysés pour 10 locus microsatellites et cartographiés. La répartition des différentes espèces et des hybrides dans le peuplement est illustrée dans la Figure 46. D'après l'assignation génétique effectuée dans le chapitre 4, ce bois comprend 241 chênes pédonculés, 246 chênes tauzins et 83 chênes pubescents. L'espèce sessile est représentée par seulement 5 individus. Globalement, ce bois comprend une forte proportion d'hybrides puisque environ 70% des arbres adultes peuvent être assignés à l'une des espèces alors que les 30% restant ne sont assignés à aucune espèce et leur assignation génétique est compatible avec un statut hybride (voir chapitre 3 et 4).

### 7.2.2. Observations de la phénologie du débourrement

La floraison des chênes est particulièrement difficile à observer. Les fleurs sont d'une part localisées à la cime des arbres, difficilement observable. D'autre part, bien que les fleurs mâles, regroupées sous forme de chatons, soient relativement visibles, les fleurs femelles ont une taille très réduite (environ 1 mm) rendant pratiquement impossible l'observation précise des stades de développements floraux de ces espèces en population naturelle. Nous avons donc mesuré non pas les stades de développements floraux mais les stades de développement foliaires, plus faciles à observer et corrélés à la floraison. Le développement foliaire se compose de 5 stades (stade 0 : dormance des bourgeons, stade 1 : élongation des bourgeons, stade 2 : ouverture des bourgeons, stade 3 : début de l'expansion des feuilles, stade 4 : au moins une feuille est sortie du bourgeon et stade 5 : élongation des entre-nœuds). L'observation de l'arbre entier permet d'apprécier le stade atteint par la majorité des bourgeons et le pourcentage de bourgeons ayant atteint ce stade. A chaque date de visite et pour chaque arbre, la notation comprend donc le stade atteint et le pourcentage de bourgeon concerné par ce stade (par exemple stade 3 à 50%). Nous avons vérifié sur des greffes de chênes pédonculés et sessiles de petite taille (facilement observable) la relation entre le développement foliaire et floral. En général, la floraison suit la fin du développement des feuilles (la floraison commence lorsque le chêne a atteint le stade 5 à 50% environ). Chez le chêne, espèce monoïque, on considère généralement que la floraison mâle (libération du pollen) précède de quelques heures la période de réceptivité des ovules (protandrie), limitant les possibilités d'autofécondation. Cependant, l'observation de 36 chênes greffés indique que c'est le cas pour 11 d'entre eux alors que pour 15 individus la fleur femelle paraît réceptive avant la libération du pollen et dans 6 cas, la floraison est synchrone. Bien que les observations de terrain confirment qu'en général le pollen est mature et libéré à la fin du stade 5, il est apparu que les chatons mâles des chênes tauzins sortent du bourgeon avant les feuilles, indiquant que la floraison mâle arrive en avance par rapport au développement foliaire chez cette espèce. La relation entre développement foliaire et floral est donc difficile à évaluer précisément car elle est très variable entre individus et entre espèces.

Nous avons observé et noté la phénologie du développement foliaire dans le bois de Briouant pendant deux printemps consécutifs (2006 et 2007) pour cinq dates identiques (jours de l'année : 94, 101, 113, 122 et 130 ; début avril à mi-mai). L'observation a été effectuée pour 65 chênes l'année

2006 (14 pédonculés, 3 sessiles, 10 pubescents, 14 tauzins et 24 hybrides) et 106 chênes l'année 2007 (12 pédonculés, 2 sessiles, 9 pubescents, 19 tauzins et 64 hybrides).

Les observations ont ensuite été transformées en variables facilement analysables. Chaque notation (stade - pourcentage) a été transformée en valeur numérique (stade = unité et pourcentage = dixième ; par exemple, stade 3-50% = 3.5). Ensuite, nous avons effectué une régression linéaire entre les jours d'observation et la valeur de développement pour chaque arbre et chaque année, pour calculer le nombre de jours nécessaire à un chêne pour atteindre le stade de développement foliaire 5-50% qui correspond à la libération du pollen. Cette valeur permet de comparer la précocité de débourrement entre individus et entre années.

### 7.2.3. Echantillonnages

#### 7.2.3.1. Echantillonnage de glands

Nous avons voulu caractériser le régime de reproduction des chênes composant ce peuplement en nous focalisant d'une part sur des arbres assignés à une espèce, et d'autre part sur des chênes hybrides. Nous nous sommes concentrés sur ces deux types de chênes pour répondre à deux questions essentielles permettant de mieux comprendre le fort pourcentage d'hybrides observés à Briouant. L'analyse de descendance de chênes assignés à une espèce va permettre d'estimer directement à un stade précoce (étape post-fertilisation et pré-dispersion) le taux d'hybridation entre espèces (formation d'hybrides de première génération). Enfin, l'analyse de descendance des chênes hybrides va permettre de préciser si les hybrides sont fertiles et d'identifier le statut spécifique du pollen fertilisant ces chênes.

Nous avons donc sélectionné une trentaine d'arbres mères à échantillonner. Cependant, tous les chênes sélectionnés ne présentaient pas de glands, ce qui a limité l'échantillonnage final à 16 arbres mères (entourés en rouge sur la Figure 46) incluant 6 chênes d'une espèce (3 chênes pédonculés, 2 chênes tauzins, un chêne pubescent) et 10 chênes hybrides (3 hybrides pédonculé x sessile, 2 hybrides pédonculé x pubescent, un hybride pédonculé x tauzin, 3 hybrides sessile x tauzin et un hybride pubescent x tauzin). Il a été particulièrement difficile de trouver des chênes pubescents portant des glands l'année d'échantillonnage (2006), ce qui limite le nombre de descendance de cette espèce à un seul individu. Les caractéristiques des chênes dont la descendance a été échantillonnée sont synthétisées dans le Tableau 26. Le nombre de glands échantillonné par arbre mère va de 23 à 76 (Tableau 26) pour un total de 727 glands. A l'issue de l'échantillonnage, le poids total de la descendance est mesuré, permettant d'estimer un poids moyen de gland par descendance. Un petit morceau de chaque gland dans sa partie basale a été prélevé afin de disposer de matériel végétal qui servira pour les extractions d'ADN. Ce prélèvement épargne l'embryon et le gland conserve donc sa capacité de germination. Le matériel ainsi prélevé est placé dans des tubes identifiés qui sont stockés dans un congélateur à -80°C en attendant d'être traités.



### 7.2.3.2. *Echantillonnage de plantules*

L'estimation du taux d'hybridation basée sur l'analyse de graines ne tient pas compte de l'éventuel isolement reproducteur postzygotique puisque les graines sont analysées avant installation de la plantule. Divers facteurs postzygotiques peuvent intervenir entre l'étape de maturation de la graine et son établissement effectif. Bien que l'analyse de la germination à posteriori des glands récoltés en environnement contrôlé puisse nous apporter quelques résultats sur d'éventuelles non viabilité des graines, la sélection qui s'exerce en forêt au cours des processus de post-dispersion n'a rien de comparable. Il est donc essentiel de s'intéresser à l'étape du cycle de vie suivant celle du stade graine, c'est-à-dire, l'étape de vie sous forme de plantule. En effet, contrairement aux glands récoltés sur un chêne, les plantules ont subi et subissent une sélection forte par divers facteurs à la fois génétiques et environnementaux.

La parcelle de Briouant comprend, outre les chênes adultes dominants, une strate de jeunes plants de chênes plus ou moins dense dont la hauteur ne dépasse pas 20 cm. Il n'existe pas de strate intermédiaire entre adultes et jeune plantule indiquant que les jeunes plants sont dominés par les adultes et subissent l'herbivorie. Ainsi, un suivi de mortalité et de croissance des plantules échantillonnées un an après la collecte montre que les plantules ont une croissance très faible lors d'une année de végétation. Le taux de mortalité est faible, les plantules peuvent ainsi probablement survivre dans cette strate pendant plusieurs années. De plus le diamètre des tiges des plantules est très variable et plusieurs nœuds de croissance peuvent souvent être observés ce qui confirme que les plantules trouvées dans la parcelle stagnent plusieurs années dans la strate herbacée.

Nous avons échantillonné les plantules dans plusieurs placettes. Chaque placette correspond à une surface de 1 mètre carré dont nous avons prélevé un morceau de feuille pour chaque plantule et repéré leur position grâce à un cadre grillagé fiché dans des tubes enfoncés dans le sol. Ce système permet d'effectuer un suivi de croissance des plantules en repositionnant le cadre de façon identique. Lors de l'échantillonnage, chaque plantule est localisée dans le cadre, mesurée pour la hauteur et le diamètre de la tige à 5 cm du sol (données non présentées). Après chaque prélèvement, une bague plastique colorée est positionnée au pied de la plantule afin de repérer les individus échantillonnés.

Cinq placettes ont ainsi été échantillonnées (voir localisation Figure 46), les distances entre les placettes et les arbres adultes les plus proches sont mesurées (Tableau 27). Le nombre de plantules varie de 19 à 81 pour les placettes 3 et 1 respectivement (Tableau 27), un total de 208 plantules a été prélevé. Le matériel végétal est conservé dans un tube numéroté puis stocké dans un congélateur à -80°C jusqu'à l'extraction d'ADN.

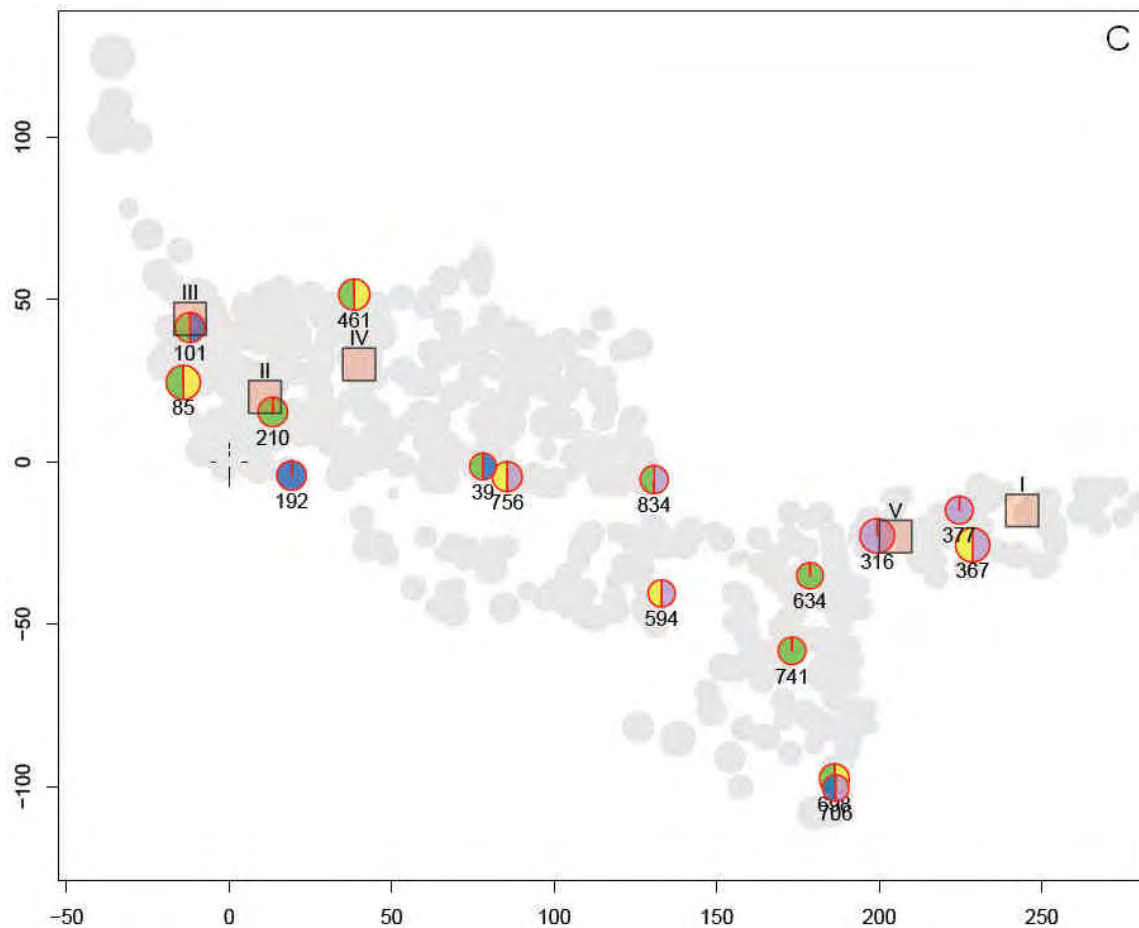


Figure 46 : Localisation des arbres mères échantillonnés pour leurs glands ainsi que des placettes (carrés I à V) échantillonnées pour des plantules. Les couleurs des symboles représentent l'espèce, ou les deux espèces composant majoritairement le génome des individus hybrides (vert : pédonculé, jaune : sessile, bleu : pubescent et violet : tauzin). Enfin, la taille des cercles est proportionnelle au diamètre du tronc à hauteur de poitrine.

#### 7.2.4. Suivi de germination des glands

Les glands récoltés sur 11 arbres mères ont fait l'objet d'un suivi de germination et de croissance (Tableau 26). Chaque gland a été disposé à la surface d'un cigare contenant du terreau, chaque cigare a été placé sur une plaque de polystyrène appropriée contenant 7 x 11 cigares. La position de chaque gland a été repérée sur un plan et les limites entre les descendance ont été signalées dans les plaques par une étiquette plastique. Les cigares sont copieusement arrosés une première fois puis un arrosage manuel est pratiqué tous les deux jours pendant trois semaines, date à laquelle les premières mesures de hauteur des plantules sont réalisées. Pendant les 4 mois suivants, 7 autres mesures de hauteur ont été réalisées à raison d'une mesure toutes les deux semaines environ. Ces mesures de hauteur permettent de déterminer si un gland donné a germé ou non, et finalement de calculer un pourcentage de germination par descendance. De plus, les hauteurs des plantules à la fin de la campagne de mesure sont comparées entre descendance, afin de mettre en évidence d'éventuels effets du poids du gland, de l'espèce de l'arbre mère ou du pourcentage de germination sur la croissance des plantules.

Tableau 26 : Effectifs des glands échantillonnés par arbre mère et des familles de demi-frère retrouvées parmi les plantules (voir partie Résultats – Analyse de parenté). L'espèce de chaque arbre mère, le nombre de glands récolté ou plantules retrouvées ainsi que le nombre de descendants génotypés pour au moins 5 microsattellites sont indiqués. Le poids moyen des glands ainsi que le pourcentage de germination pour 11 familles suivies sont indiquées dans les deux dernières colonnes (nd : non suivi pour la germination).

	Code descendance	Mère	Espèce	Descendants	N desc.	N desc. (5ssr min)	Poids moyen des glands (g)	Pourcentage de germination
Espèces	2	210	Ped	Glands	23	16	1.86	nd
	13	634	Ped	Glands	38	33	1.70	nd
	9	741	Ped	Glands	48	46	1.72	37.5%
	4	192	Pub	Glands	76	69	0.83	72.4%
	16	377	Toz	Glands	28	17	1.28	nd
	1	316	Toz	Glands	31	18	1.61	48.1%
	20	316	Toz	Plantules	21	21		
	17	388	Toz	Plantules	38	38		
	18	443	Ped	Plantules	17	17		
	19	211	Ped	Plantules	10	10		
	Hybrides	8	85	Ped/Ses	Glands	48	42	2.40
15		461	Ped/Ses	Glands	30	29	0.64	nd
6		698	Ped/Ses	Glands	56	52	2.22	32.1%
12		39	Ped/Pub	Glands	62	57	2.08	66.1%
10		101	Ped/Pub	Glands	37	28	0.52	nd
11		834	Ped/Toz	Glands	59	50	2.94	61.0%
3		367	Ses/Toz	Glands	31	28	1.75	72.7%
14		594	Ses/Toz	Glands	53	43	1.31	nd
5		756	Ses/Toz	Glands	57	50	2.53	69.1%
	7	706	Pub/Toz	Glands	48	48	1.91	79.2%
Total		20 desc.			813	712		11 desc.

### 7.2.5. Génotypage des glands et des plantules

L'extraction d'ADN des morceaux de feuille de plantules est effectué avec un protocole standard Lepais *et al.*, 2006b alors que l'extraction d'ADN à partir des morceaux de glands a nécessité une mise au point supplémentaire. L'ADN de chaque morceau de gland est extrait grâce à un protocole basé sur du CTAB mais spécialement adapté par François Hubert (MOO, 2007) pour extraire l'ADN des tissus de réserves riches en divers composés pouvant interférer lors de l'amplification PCR. Brièvement, le protocole reste globalement le même (Lepais *et al.*, 2006b) mais le tampon d'extraction a été additionné de protéinase K ( $0.5\text{mg.mL}^{-1}$ ) et de béta-mercaptoéthanol (0.2%). A l'étape de précipitation des protéines et des lipides, le dichlorométhane est remplacé par un mélange chloroforme : isoamylalcool (24 : 1). Enfin, une étape de précipitation supplémentaire de l'ADN à l'alcool est réalisée afin de purifier davantage l'ADN extrait.

L'ADN ainsi extrait est dilué au 1/400 et sert de matrice pour l'amplification des 10 microsattellites suivant le protocole habituel (Lepais *et al.*, 2006b).

### 7.2.6. Assignment génétique des glands et des plantules

Le génotype des 10 microsatellites des glands et des plantules analysés sont utilisés pour une analyse d'assignment génétique par le logiciel Structure version 2.1 (Pritchard *et al.*, 2000) afin de caractériser la composition globale en espèce de chaque descendance. L'ensemble des populations analysées dans le chapitre 4 a été utilisé comme base de référence pour l'assignment. Les génotypes des descendants ont été rajoutés à ces données et l'assignment a été menée en utilisant l'option "POPFLAG" de Structure. Les descendances ayant une forte structuration génétique familiale (surreprésentation des allèles maternels car apparemment élevé), il est important que leurs génotypes ne soient pas pris en compte pour l'estimation initiale des fréquences alléliques des 4 clusters. Nous avons donc spécifié au logiciel de prendre en compte l'ensemble des individus adultes pour l'estimation de fréquences alléliques ("POPFLAG = 1") alors que les descendances sont exclues de l'estimation de ces fréquences alléliques ("POPFLAG = 0"). Les modèles "Admixture" ainsi que "Correlated Allele Frequency among Population" sont utilisés avec 100000 itérations pour le temps de chauffe (Burnin Period) et 500000 répétitions MCMC après le temps de chauffe et un nombre de groupe  $K$  fixé à 4 (4 espèces présentes dans le jeu de données). L'analyse est répétée 3 fois afin de s'assurer de la stabilité des résultats. Chaque descendant est ainsi caractérisé pour la proportion de son génome appartenant à chacune des 4 espèces. Certains descendants analysés, en particulier les glands, sont des hybrides de seconde génération car issus de mères hybrides. Vu la faible précision d'assignment pour des hybrides au-delà de la première génération (voir chapitre 3), nous n'assignons pas les descendants au niveau individuel à une espèce ou une classe d'hybride, mais nous analysons les résultats d'assignment par descendance.

Les distances génétiques entre descendances (groupées en familles) sont calculées grâce au logiciel *Populations* (Langella, 1999) en utilisant la distance génétique  $D_{SA}$  calculée selon la méthode du nombre d'allèles partagés (Jin & Chakraborty, 1994). Entre deux individus, cette distance génétique est égale à :

$$D_{SA_i} = 1 - P_{SA_i}, \text{ et } P_{SA_i} = \frac{\sum_j^r S}{2r}$$

où  $S$  représente le nombre d'allèles partagés entre deux individus et  $r$  le nombre de locus.

Entre populations, cette distance s'exprime par la formule :

$$D_{SA_B} = 1 - \frac{2\bar{P}_{SA_B}}{\bar{P}_{SA_X} + \bar{P}_{SA_Y}}$$

où la proportion moyenne d'allèles partagés entre deux populations ( $\bar{P}_{SA_B}$ ) et au sein de chaque population ( $\bar{P}_{SA_X}$  et  $\bar{P}_{SA_Y}$ ) est calculée pour chaque combinaison d'individus échantillonnés.

La matrice de distances (inter descendances) est représentée à l'aide d'un arbre non enraciné selon la méthode de *neighbourhood joining* grâce au package APE (Paradis *et al.*, 2004) du logiciel R.

### 7.2.7. Analyse de paternité

Sur les 727 glands génotypés, 626 (soit 86%) ont un génotype complet pour au moins 5 microsatellites et sont utilisés pour les analyses suivantes (Tableau 26). Trois locus présentant des taux d'erreur de génotypage supérieur à 5% (Qr39, Qr5 et Qr65) sont exclus pour l'analyse de paternité (voir chapitre 2). Une analyse de paternité a été utilisée grâce au génotype des glands et des chênes adultes (7ssr) pour identifier la provenance du pollen ayant contribué à la formation de chaque gland parmi les 817 chênes du bois de Briouant. La paternité est identifiée par une méthode de maximum de vraisemblance implémentée dans le logiciel FaMoz (Gerber *et al.*, 2003). Tout d'abord, la probabilité d'exclusion de paternité connaissant la mère ( $EP_p$ ) est calculée à partir des fréquences alléliques des chênes adultes pour les 7 locus cumulés (Tableau 29). L'identification du père le plus probable est réalisée grâce à un LOD score basé sur les fréquences alléliques des chênes adultes. Afin de déterminer un seuil de confiance permettant d'optimiser l'identification de père potentiels, nous avons comparé la distribution des LOD scores des pères les plus probables pour 100000 glands simulés à partir des génotypes des 817 adultes et de la distribution des LOD scores des pères les plus probables de 100 000 glands simulés à partir de génotypes recréés à partir des fréquences alléliques des chênes adultes. L'intersection des deux distributions permet de définir un seuil de LOD score qui minimise les erreurs de type I (rejeter un père potentiel alors que c'est un vrai père) et de type II (accepter à tort un père alors que ce n'en est pas un). Une fois ce seuil de LOD score déterminé, les erreurs de type I et II sont calculées et les paternités sont identifiées pour les pères potentiels qui ont un LOD score supérieur au seuil déterminé. Pour éviter une exclusion de pères potentiels liée à une erreur de génotypage (estimées à 0.7% pour les 7ssr), un paramètre d'erreur fixé à 0.000001 est introduit dans les simulations ainsi que dans les calculs de LOD score, comme recommandé dans le manuel de FaMoz.

Les résultats de cette analyse de paternité sont synthétisés par descendance. Le nombre de glands dont le père n'est pas retrouvé dans la parcelle ainsi que l'espèce (ou le type hybride) des pères identifiés sont reportés par descendance dans un tableau synthétique (Tableau 30). Enfin, les coordonnées de chaque chêne sont utilisées pour représenter les événements de pollinisation ayant eu lieu l'année de la récolte sur une carte de la parcelle en faisant ressortir l'espèce ou le type hybride des chênes impliqués dans ces événements de reproduction.

### 7.2.8. Analyse de parenté

Parmi les 10 microsatellites analysés pour les 206 plantules, les 3 locus présentant des taux d'erreur supérieur à 5% ne sont pas utilisés, l'analyse de parenté, tout comme l'analyse de paternité, est donc conduite avec 7 microsatellites. Connaissant le génotype des 817 adultes et des plantules pour ces marqueurs, l'analyse de parenté a pour objectif d'assigner à chaque plantule soit (1) un couple de parents, soit (2) un seul parent si aucun couple compatible n'est retrouvé, indiquant que le second parent n'est pas présent dans l'échantillonnage des adultes (soit parce qu'il est à l'extérieur de la parcelle, soit parce qu'il a été coupé avant échantillonnage ou alors il présente une erreur de

génotypage qui le rend incompatible avec le descendant), ou bien aucun parent si aucun adulte n'est compatible. Le fait de ne pas retrouver un des parents ou même les deux parents peut avoir de multiples causes. Il est souvent estimé qu'un parent non retrouvé correspond à un événement de migration de pollen (un seul parent retrouvé) ou de graine (aucun parent retrouvé). Néanmoins, d'autres causes peuvent générer le même résultat comme par exemple la disparition d'un parent (coupe ou mortalité) ou les erreurs de génotypage (excluant à tort un individu comme parent). Nous allons dans ce chapitre nous focaliser sur les plantules dont au moins un parent est assigné de façon probabiliste, avec un taux d'erreur de type I et II connu, grâce à la recherche de parenté par méthode de maximum de vraisemblance implémenté dans FaMoz (Gerber *et al.*, 2003).

La probabilité d'exclusion d'un parent seul ( $EP_{SP}$ ) et d'un couple de parents ( $E_{PP}$ ) est calculée à partir des fréquences alléliques des adultes pour les 7 microsatellites cumulés (Tableau 29). L'assignation d'un parent ou d'un couple de parents est réalisée avec les mêmes paramètres que l'analyse de paternité.

Les résultats globaux de cette analyse de parenté sont tout d'abord présentés par placette (Tableau 31). Parmi les plantules dont un parent ou un couple est assigné, certains arbres à proximité immédiate des placettes interviennent comme parent plusieurs fois (Figure 46), permettant de définir sans ambiguïté 4 familles de demi-frère composés de 10 plantules ou plus (voir Tableau 26). Ces 4 familles sont rajoutées aux 16 familles constituées par les glands, permettant une comparaison de la composition génétique des graines et des plantules. En particulier, le chêne 316, échantillonné pour des glands et inclus dans l'analyse de paternité, intervient comme parent pour 20 plantules. Sachant que ces plantules ont été récoltées à proximité immédiate de cet arbre (50 cm), nous pouvons considérer que l'arbre 316 est la mère de ces plantules. Ainsi une comparaison directe du stade pré- et post-dispersion est possible pour les descendants de cet arbre.

### 7.2.9. Identification directe des événements de reproduction

Lorsque deux parents sont retrouvés, leur assignation génétique étant connue, la définition précise du type d'événement de reproduction ayant généré le descendant est possible. Nous pouvons comparer à ce stade les événements reproducteurs estimés au stade graine et au stade plantules. Néanmoins cette comparaison ne peut être que générale car le nombre d'arbre mère étudié est faible et d'éventuels effets individuels peuvent être décisifs. Cependant, nous disposons pour la mère 316 d'un échantillonnage de glands et de plantules ce qui permet une comparaison précise pour cette mère.

Les événements de reproduction ayant eu lieu au stade graine et plantule sont comptabilisés par arbre mère, ainsi que par type de mère (espèce pure ou hybride). Les événements de reproduction observés sont additionnés, puis, pour tenir compte du fait de la surreprésentation des mères hybrides dans l'échantillonnage, les événements reproducteurs sont estimés pour une composition réelle de 70% de chênes appartenant à une espèce et 30% de chênes hybrides (proportion réelle

d'hybrides à Briouant, voir chapitre 4) afin de refléter plus fidèlement les événements de reproduction ayant lieu dans la parcelle.

#### 7.2.10. Corrélation de paternité intra- et inter-familles

La caractérisation des événements reproducteurs par analyse directe (paternité ou parenté) est restreinte aux descendants dont les deux parents sont connus. Chez les espèces fortement allogames, et d'autant plus chez les plantes dont le pollen peut être dispersé sur de longues distances, les pères de la plupart des descendants seront localisés à l'extérieur de la parcelle étudiée et donc une grande proportion de pères ne peut pas être identifiée. De plus, la parcelle étant gérée par plusieurs propriétaires et conduite en cépée, des zones de coupe sont fréquentes et de petite taille, faisant disparaître au moins temporairement certains individus (seulement présents sous forme de souche) qui n'ont donc pas été échantillonnés.

Afin de tirer parti de l'ensemble des données génétiques des descendants (et non seulement des données génétiques des descendants dont les deux parents ont été identifiés), des méthodes d'analyses de système de reproduction dites indirectes existent et ont pour objectif de comparer les nuages polliniques reçus par différentes mères.

Le programme PolDisp (Robledo-Arnuncio *et al.*, 2007) a été utilisé afin de comparer les pools polliniques ayant généré les descendants analysés sur chaque arbre mère. PolDisp rassemble deux programmes indépendants, qui implémentent deux modèles de corrélation de paternité, KinDist (Robledo-Arnuncio *et al.*, 2006 ; Hardy *et al.*, 2004) et TwoGener (Austerlitz & Smouse, 2001 ; Austerlitz & Smouse, 2002). KinDist a été utilisé pour estimer les corrélations de paternité intra famille, correspondant à la proportion de graine de la même mère qui partagent le même père. TwoGener a été utilisé afin de calculer les différenciations de nuage pollinique entre les mères échantillonnées,  $\Phi_{FT}$ . La différenciation  $\Phi_{FT}$  des nuages polliniques inter famille est équivalente à l'inverse de la corrélation de paternité inter famille, mais la différenciation a été préférée car elle correspond à un indice de différence et non de similarité, permettant de traiter les différenciations entre paires de nuage pollinique comme une matrice de distance. La matrice des différenciations polliniques a été utilisée pour représenter un arbre *neighborhood joining* grâce au package APE (Paradis *et al.*, 2004) du logiciel R, permettant de visualiser les mères recevant le même type de pollen ou des types de pollen différents.

### 7.3. Résultats

#### 7.3.1. Phénologie des espèces et des hybrides

En 2006, les chênes mettent entre 116 et 138 jours pour atteindre le stade 5-50% alors que ce même stade est atteint entre 120 et 132 jours l'année 2007. Ainsi, la période de débourrement a été nettement plus réduite l'année 2007 puisque l'écart est de 22 jours pour l'année 2006 contre 12 jours seulement pour 2007 (Figure 47 et Figure 48). Ce résultat s'explique par une période particulièrement chaude pour la saison début avril 2007, ce qui a précipité la reprise de végétation

de l'ensemble des espèces incluant les arbres les plus tardifs. La comparaison des deux années montre également que l'ordre de débournement est largement conservé d'une année à l'autre (Figure 47).

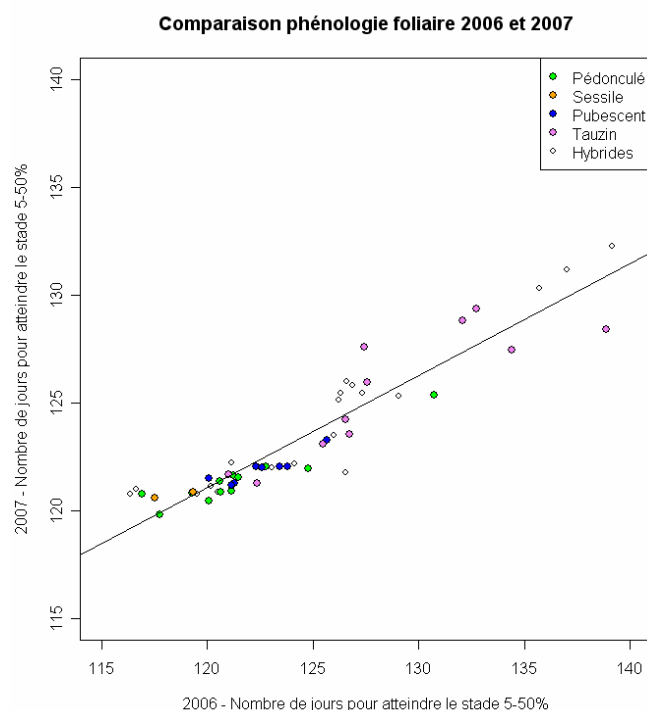


Figure 47 : Nombre de jours pour atteindre le stade 5-50% comparé pour deux années consécutives (2006 en abscisse et 2007 en ordonné).

La comparaison du nombre de jours nécessaires pour atteindre le stade 5-50% entre espèces montre que le chêne tauzin est l'espèce la plus tardive (Figure 48) puisqu'elle met environ une semaine de plus pour atteindre le même stade que les autres espèces. En moyenne, les chênes sessiles sont les plus précoces, suivis des chênes pédonculés puis des chênes pubescents et enfin des chênes tauzins. Cependant on constate aussi une grande variabilité intra spécifique pour toutes les espèces et en particulier le chêne tauzin (Figure 48). Il est ainsi possible de trouver des chênes des quatre espèces au même stade de débournement, sauf au tout début et à la fin de la période de débournement. Le décalage de phénologie ne paraît donc pas être un facteur général pouvant restreindre les croisements interspécifiques.



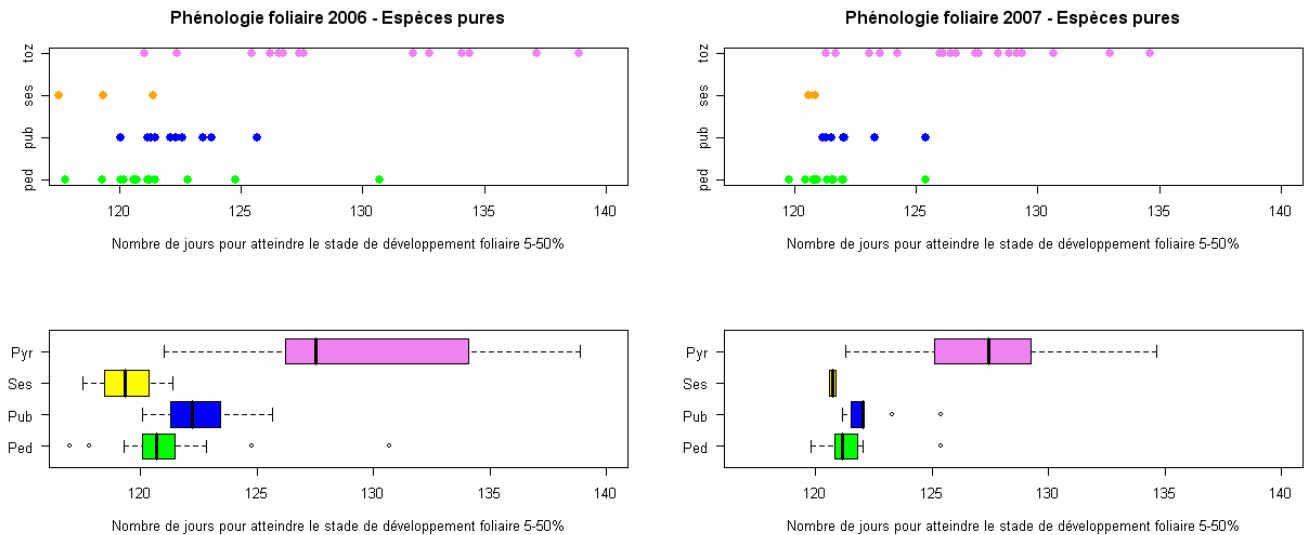


Figure 48 : Distribution du nombre de jours pour atteindre le stade 5-50% par espèce pour l'année 2006 (à gauche) et l'année 2007 (à droite).

L'analyse des chênes hybrides conforte les résultats obtenus précédemment. Ainsi les hybrides ayant comme espèce parentale le chêne tauzin sont en moyenne plus tardifs (Figure 49). De plus, l'ordre de débourrement de ces hybrides x tauzin est identique à celui des espèces parentales puisque s'exprime du plus précoce au plus tardif : hybrides sessiles x tauzin, puis hybrides pédonculé x tauzin et enfin hybrides pubescent x tauzin. Cette tendance remarquable se répète pour les deux années de mesures (Figure 49).

Pour vérifier cette tendance, nous avons analysé la relation entre pourcentage du génome contenant des gènes tauzin par rapport au nombre de jours pour atteindre le stade 5-50% (Figure 50). Une corrélation significative est trouvée lorsque tous les chênes sont pris en compte (Pearson's product-moment correlation, année 2006 :  $t = 7.27$ ,  $df = 63$ ,  $p\text{-value} < 0.001$ ,  $r = 0.68$  ; année 2007 :  $t = 5.93$ ,  $df = 104$ ,  $p\text{-value} < 0.001$ ,  $r = 0.50$ ) mais aussi lorsque seuls les arbres ayant plus de 10% de leur génome provenant du chêne tauzin (ce qui revient à conserver uniquement les chênes tauzins ou leur hybrides) sont considérés (Pearson's product-moment correlation, année 2006 :  $t = 2.62$ ,  $df = 33$ ,  $p\text{-value} = 0.013$ ,  $r = 0.41$  ; année 2007 :  $t = 2.76$ ,  $df = 78$ ,  $p\text{-value} = 0.007$ ,  $r = 0.30$ ). Ce résultat indique que la période de débourrement a un déterminisme génétique important puisque globalement, plus un individu a une proportion élevée de gènes tauzins, plus il est tardif.

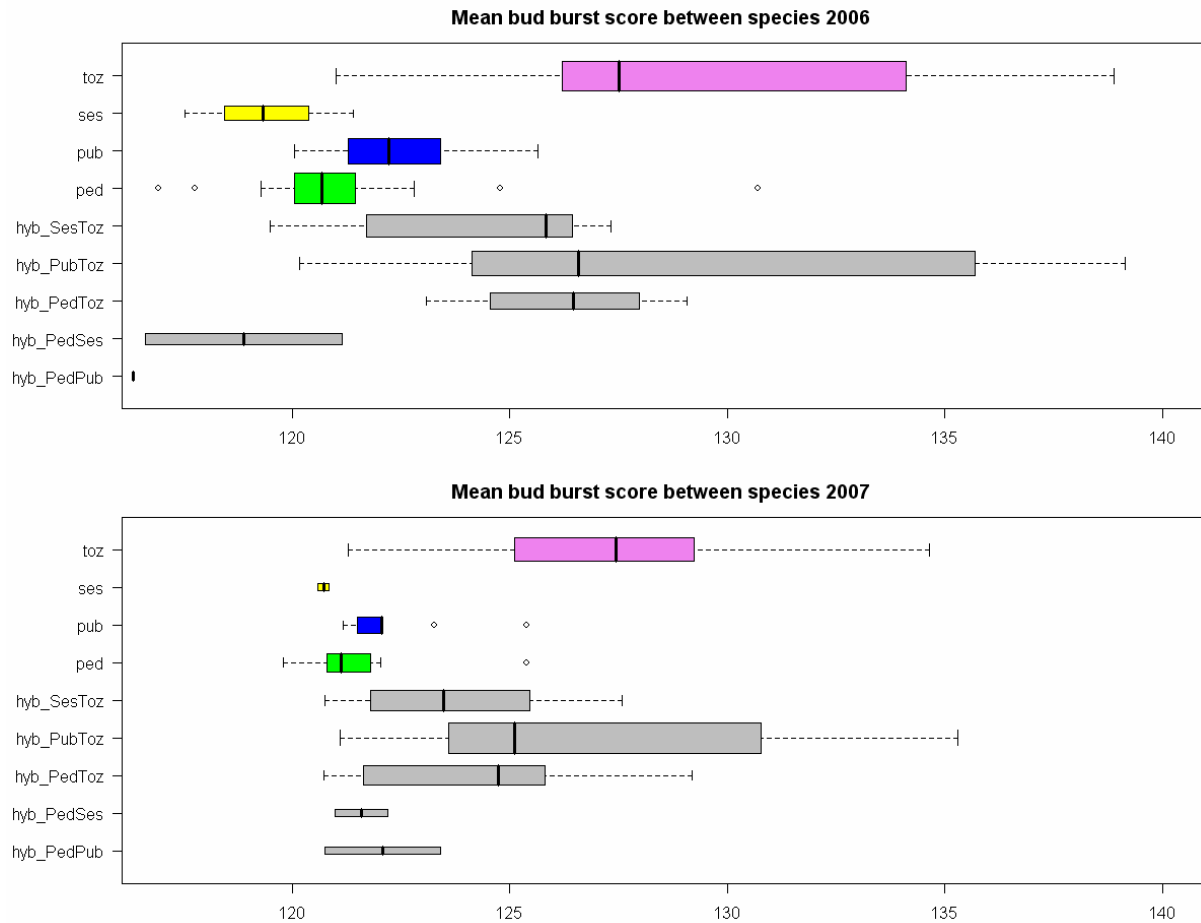


Figure 49 : Distribution du nombre de jours pour atteindre le stade 5-50% par espèce et type d'hybride pour l'année 2006 (en haut) et l'année 2007 (en bas).

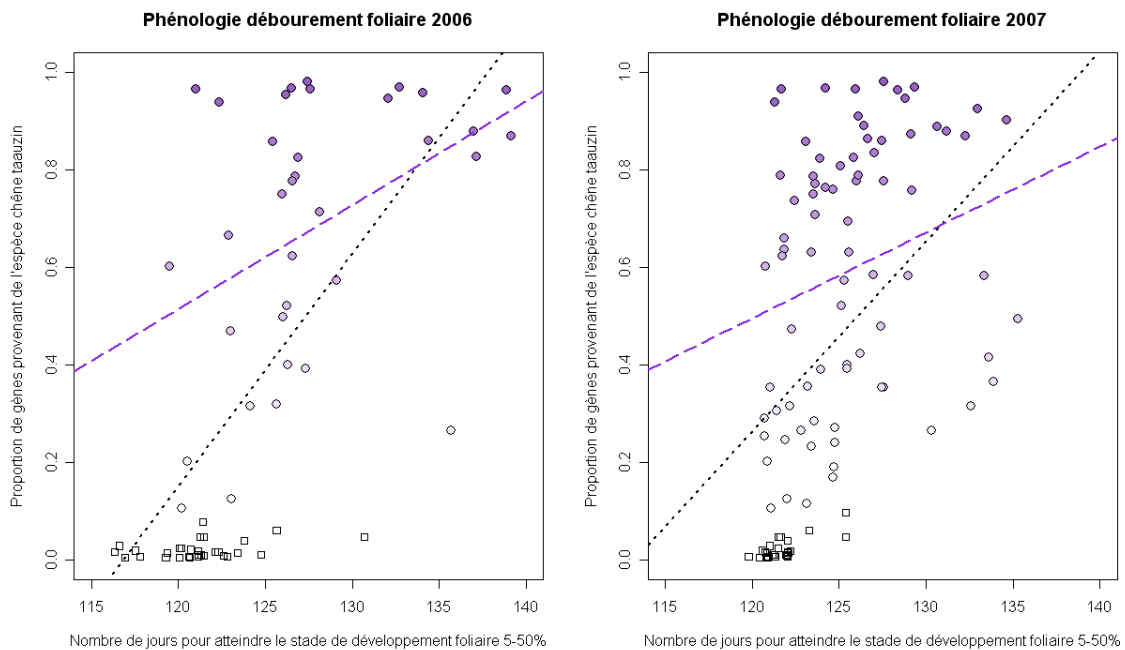


Figure 50 : Corrélation entre le nombre de jours pour atteindre le stade 5-50% et la proportion de gènes tauzins dans le génome des individus pour l'année 2006 (à gauche) et 2007 (à droite). La droite noire représente la régression incluant l'ensemble des individus, la droite violette représente la régression incluant les individus ayant au moins 10% de leur génome composé de gènes tauzins.

Le décalage global de la phénologie entre espèces permet de conclure que la nature du pollen libéré varie pendant la saison de floraison. Il sera essentiellement composé de pollen sessile et pédonculé au début de la saison de floraison puis va s'enrichir de pollen pubescent et finalement laisser place à une proportion de plus en plus élevée de pollen tauzin à la fin de la saison. Le pollen circulant dans la parcelle sera donc composé d'une proportion variable de pollen des quatre espèces (ainsi que des hybrides) au cours du temps. Parallèlement, la même tendance peut être attendue concernant les fleurs femelles, bien que peu de données soient disponibles sur la durée de réceptivité des stigmates (qui pourrait s'étendre sur une plus longue période que celle de la libération du pollen). Excepté pour les arbres ayant des périodes de débournements extrêmes (très précoces ou très tardifs), les différences de phénologie ne semblent pas suffisantes pour empêcher l'hybridation. En outre, les individus présentant une phénologie atypique pour leur espèce vont avoir un plus grand risque d'être fécondé par une autre espèce. C'est notamment le cas pour les chênes pédonculés ou pubescents tardif qui vont être exposé à une proportion importante de pollen de chêne tauzin et inversement, les chênes tazuins précoces vont avoir plus d'opportunité pour s'hybrider avec les espèces pédonculé, tauzin ou même sessile.

Les hybrides ayant une phénologie clairement intermédiaire pourraient représenter un pont temporel reliant les périodes de floraison des espèces parentales. Enfin, le manque de précision quant à la période de floraison ainsi que divers paramètres de la floraison elle-même (variabilité intra-arbre de la floraison, viabilité du pollen, période de réceptivité de la fleur femelle) ne permet pas de conclure précisément sur les conséquences des décalages de phénologie sur les flux de gènes qu'ils soient inter- ou intra-spécifiques. Il apparaît en effet qu'il y a autant de décalage de phénologie à l'intérieur d'une même espèce qu'entre espèces différentes.

### 7.3.2. Caractéristiques des descendance étudiées

#### 7.3.2.1. *Glands*

Nous avons échantillonné un total de 727 glands sur 16 arbres mères (Tableau 26) dont la localisation est précisée dans la Figure 46. Le poids moyen des glands par famille est compris entre 0.52 g pour la mère hybride pédonculé x pubescent 101 à 2.94 g pour l'arbre mère hybride pédonculé x tauzin 834 (Tableau 26). Bien que les glands échantillonnées sur les mères hybrides ont un poids moyen plus élevé que les glands échantillonnés sur des arbres assignés à une espèce (Figure 51), la différence n'est pas significative ( $F = 0.89$ ,  $df = 1$ ,  $p = 0.36$ ). Le suivi de germination des glands montre que les glands récoltés sur des mères hybrides ont un taux de germination légèrement supérieur à ceux issus des mères assignées à une espèces (Figure 51), bien que là aussi, la différence ne soit pas significative ( $F = 0.46$ ,  $df = 1$ ,  $p = 0.52$ ).

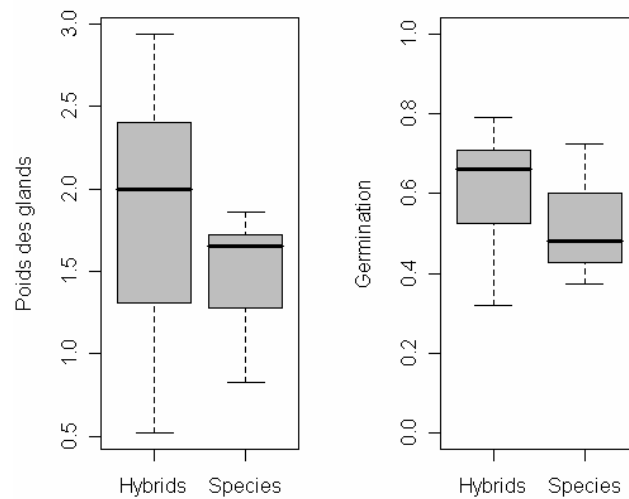


Figure 51 : Répartition des poids moyens des glands et des pourcentages de germination pour les glands échantillonnés sur des chênes assignés à une espèce pure ou hybrides. Les différences entre ces deux groupes ne sont pas significatives (pour le poids des glands :  $F = 0.89$ ,  $df = 1$ ,  $p = 0.36$  et pour le taux de germination :  $F = 0.46$ ,  $df = 1$ ,  $p = 0.52$ ).

### 7.3.2.2. Plantules

Nous avons échantillonné un total de 208 plantules réparties sur 5 placettes avec une densité variable entre placettes de 19 plantules / m<sup>2</sup> dans la placette III à 81 plantules / m<sup>2</sup> pour la placette I (Tableau 27). Les diamètres des tiges des plantules sont significativement différents entre placettes ( $F = 5.45$ ,  $df = 4$ ,  $p < 0.001$ ), en particulier un test de comparaison multiple de moyenne (test de Tukey) montre que les placettes I et V ont des plantules de diamètres significativement plus élevés que les placettes II, III et IV. Les tailles des plantules en 2004 sont significativement différentes entre placettes ( $F = 3.25$ ,  $df = 4$ ,  $p = 0.014$ ), différence générale liée seulement à une différence significative entre les placettes I et V (test de comparaison multiple de Tukey :  $p\text{-adj} = 0.022$ ). La différence dans la taille des plantules est toujours significative la seconde année ( $F = 3.69$ ,  $df = 4$ ,  $p = 0.006$ ), là encore liée à une différence significative entre les placettes I et V (test de comparaison multiple de Tukey :  $p\text{-adj} = 0.009$ ). Le taux de croissance des plantules entre les deux années semble légèrement varier entre placette ( $F = 3.03$ ,  $df = 4$ ,  $p = 0.019$ ) mais les comparaisons par placettes deux à deux ne montrent aucune différence significative. Enfin, le taux de mortalité est compris entre 2% pour la placette V à 21% pour la placette III, mais cette différence n'est pas significative à cause des faibles effectifs dans certaines placettes.

Tableau 27 : Description des placettes d'un mètre carré échantillonnées pour les plantules.

Placette	N	Diamètre moyen 2004 (mm)	Hauteur moyenne 2004 (mm)	Hauteur moyenne 2005 (mm)	Croissance moyenne (mm)	Taux de mortalité	Chênes adultes à proximité immédiate	Espèce	Dist. (cm)
I	81	1.76	121	133	12	0.11	<b>388</b>	Toz	130
							387	hyb_PedToz	283
							405	Toz	398
II	24	1.24	128	159	31	0.17	212	hyb_PubToz	200
							<b>211</b>	Ped	260
							207	Ped	300
							210	Ped	300
III	19	1.39	166	212	46	0.21	103	hyb_PubToz	215
							101	hyb_PedPub	85
IV	27	1.33	144	183	39	0.11	444	Toz	250
							<b>443</b>	Ped	290
							446	Ped	305
							442	Ped	392
V	57	1.73	158	171	13	0.02	<b>316</b>	Toz	50
							315	hyb_PubToz	200
							317	Toz	300
							334	Toz	355
							337	Toz	366
Total	208								

### 7.3.3. Assignation génétique des glands

Les données génétiques des descendants sont rajoutées aux données génétiques des chênes adultes analysées au chapitre 4 afin d'assigner les descendants sur la base de leurs seules caractéristiques génétiques. Les résultats des assignations sont moyennés par descendance et mis en relation avec les assignations génétiques des arbres mères (Tableau 28). Quelques descendances ont une caractéristique génétique moyenne identique à l'arbre mère (Tableau 28). C'est le cas des mères 741 et 211 (pédonculé), 192 (pubescent), 388 (tauzin) et 101 (hybride pédonculé x pubescent). Lorsque la descendance et la mère d'une espèce ont les mêmes compositions génétiques, cela témoigne de croisements intra-spécifiques. Cependant, la plupart des descendances ont une caractéristique génétique moyenne différente de leur arbre mère (indiqué en rouge et vert dans le Tableau 28). C'est le cas pour des mères assignées à une espèce ce qui suggère des événements d'hybridation. Par exemple, la descendance de la mère 634 (pédonculée) se compose d'une proportion moindre de gènes pédonculés et supérieure de gènes pubescents et tauzins, indiquant quelques hybridations entre ce chêne pédonculé et des chênes contenant des gènes tauzins et pubescents. De la même façon, la mère 377 (tauzin) a une descendance qui se compose d'une proportion non négligeable de gènes pubescents et pédonculés. Un autre chêne tauzin analysé pour des glands et des plantules (316) montre que pour ces deux stades de vie (pré- et post-dispersion), la descendance semble contenir une forte proportion de gènes sessiles et pubescents. Remarquons que

les hybrides produits par cette mère ne sont pas restreints au seul stade gland puisque les plantules analysées ont une composition génétique très proche des glands. Enfin, les hybrides (mis à part le chêne 101) ont des descendance dont la composition génétique moyenne diffère systématiquement de l'arbre mère mettant en évidence des phénomènes d'introgression.

Les résultats des assignations génétiques individuelles sont présentés Figure 52, les distances inter-descendance représentées sous forme d'arbre permettant de visualiser les relations génétiques entre les différentes familles. Ces distances sont toutes comprises entre 0.2 et 0.8 (Figure 52A) excepté les descendance 1 et 20 qui ont une distance génétique très faible. La distance génétique utilisée est basée sur le nombre d'allèles partagés ( $D_{SA}$ , Jin & Chakraborty, 1994) expliquant la forte proximité des familles 1 et 20 (respectivement glands et plantules) issues de la même mère (316). Les différentes parties de l'arbre concordent globalement avec les résultats de l'assignation génétique. La partie droite regroupe les descendance présentant essentiellement des gènes pédonculés, la branche la plus à gauche rassemble la famille 4 récoltée sur un chêne pubescent (192) ainsi que deux autres familles contenant des hybrides pédonculé x pubescent. Enfin, la partie supérieure inclus une famille de plantules appartenant à une mère tauzin (388) ainsi que plusieurs familles composées d'une proportion plus ou moins importante de gènes tautins.

Tableau 28 : Résultats de l'assignation des arbres mères et moyenne pour les descendances. Une étoile (\*) indique une famille de plantule.

Mères et Descendances	Espèce Mère	N	Cluster1 (pédunculé)	Cluster2 (sessile)	Cluster3 (pubescent)	Cluster4 (tauzin)
Mère 210	Ped		0.937	0.0469	0.009	0.007
2		16	<b>0.822</b>	<b>0.108</b>	0.049	0.022
Mère 634	Ped		0.906	0.026	0.023	0.046
13		33	<b>0.589</b>	0.064	<b>0.163</b>	<b>0.184</b>
Mère 741	Ped		0.974	0.015	0.007	0.004
9		46	0.889	0.056	0.032	0.023
Mère 192	Pub		0.006	0.008	0.978	0.008
4		69	0.029	0.036	0.892	0.043
Mère 377	Toz		0.020	0.006	0.008	0.966
16		17	<b>0.133</b>	0.045	<b>0.174</b>	<b>0.647</b>
Mère 316	Toz		0.006	0.073	0.010	0.910
1		18	0.042	<b>0.204</b>	<b>0.386</b>	<b>0.368</b>
*20		21	<i>0.085</i>	<b>0.236</b>	<b>0.260</b>	<b>0.419</b>
Mère 388	Toz		0.005	0.007	0.005	0.983
*17		38	<i>0.029</i>	<i>0.031</i>	<i>0.037</i>	<i>0.903</i>
Mère 443	Ped		0.925	0.046	0.014	0.015
*18		17	<b>0.800</b>	<b>0.119</b>	<i>0.041</i>	<i>0.040</i>
Mère 211	Ped		0.980	0.008	0.006	0.006
*19		10	<i>0.912</i>	<i>0.026</i>	<i>0.026</i>	<i>0.036</i>
Mère 85	Ped/Ses		0.573	0.407	0.009	0.010
8		42	<b>0.623</b>	<b>0.281</b>	0.056	0.040
Mère 461	Ped/Ses		0.761	0.217	0.009	0.013
15		29	<b>0.810</b>	<b>0.048</b>	0.073	0.068
Mère 698	Ped/Ses		0.541	0.403	0.028	0.029
6		52	<b>0.675</b>	<b>0.224</b>	0.045	0.056
Mère 39	Ped/Pub		0.490	0.034	0.416	0.060
12		57	<b>0.290</b>	0.051	<b>0.585</b>	0.074
Mère 101	Ped/Pub		0.230	0.025	0.713	0.032
10		28	0.219	0.075	0.653	0.054
Mère 834	Ped/Toz		0.438	0.027	0.036	0.499
11		50	0.398	0.106	<b>0.238</b>	<b>0.258</b>
Mère 367	Ses/Toz		0.015	0.703	0.064	0.218
3		28	0.058	<b>0.307</b>	<b>0.481</b>	0.154
Mère 594	Ses/Toz		0.008	0.136	0.012	0.844
14		43	<b>0.153</b>	0.198	<b>0.396</b>	<b>0.252</b>
Mère 756	Ses/Toz		0.005	0.572	0.031	0.392
5		50	0.051	<b>0.276</b>	<b>0.296</b>	0.377
Mère 706	Pub/Toz		0.037	0.019	0.680	0.265
7		48	0.061	0.045	<b>0.474</b>	<b>0.421</b>

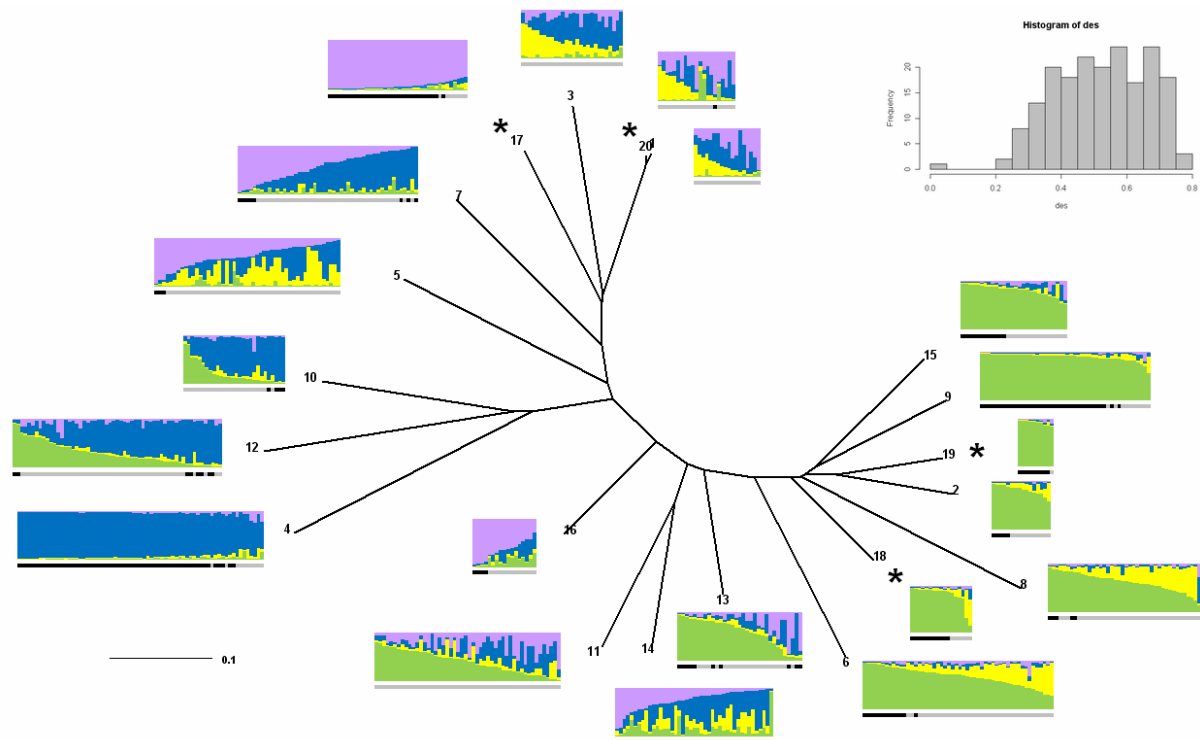


Figure 52 : Distribution des distances génétiques inter-descendants (DSA, Jin & Chakraborty, 1994), résultats de l'assignation génétique représentés par famille et arbre non enraciné calculé à partir des distances génétiques par méthode de *neighborhood joining*. Les feuilles de l'arbre sont identifiées par les numéros des familles (voir tableau 1), une étoile (\*) indique que la famille correspond à des plantules, les autres familles correspondant à des glands. Les branches de l'arbre sont proportionnelles à la distance génétique entre deux familles (échelle apposée en bas à gauche du graphique : distance génétique = 0.1). Les résultats de l'assignation génétique sont présentés au niveau individuel : chaque famille est représentée par un graphique dont chaque barre verticale représente l'assignation génétique d'un descendant de la famille. Les différentes couleurs indiquent la proportion du génome de l'individu composé de gènes pédonculés (en vert), sessiles (en jaune), pubescents (en bleu) ou tanzins (en violet). La barre horizontale sous chaque graphique en barre représente les individus assignés à une espèce (noir) ou hybride (gris).

#### 7.3.4. Analyse de paternité

Connaissant le génotype de chaque gland échantillonné ainsi que celui de la mère et des adultes présents dans la parcelle, nous avons voulu chercher à identifier le père de chaque gland parmi les chênes adultes de la parcelle. Les fréquences alléliques des pères potentiels (adultes) ont dans un premier temps été utilisées pour calculer la probabilité d'exclusion de paternité (probabilité d'exclure les individus qui ne sont pas les vrais pères) qui grâce à une grande variabilité des marqueurs génétiques utilisés est de 1 (Tableau 29). Dans un second temps, la procédure de simulation intégrée à FaMoz permet de définir un seuil de LOD score. Les relations gland – père dont le LOD score dépasse ce seuil sont considérées comme fortement probables, ainsi nous avons assigné un père aux glands dont les LOD score dépassait 4.95 (Tableau 29). Avec ce seuil d'assignation, l'erreur de type I (exclure un père alors que c'est un vrai père) est d'environ 2% alors que l'erreur de type II (assigner un père qui n'est pas le vrai père) est de 7% (Tableau 29).



Le nombre de glands dont un père a été assigné varie en fonction des familles (Tableau 30) allant de 0 pour la descendance 3 échantillonnées sur la mère 367 à 24 pour la descendance 7 récoltée sur la mère 706. En moyenne, 75% des glands analysés n'ont pas de père assigné (Tableau 30). Les résultats de l'analyse de paternité, associés à ceux de l'assignation génétique, sont représentés en prenant en compte la localisation des chênes dans la parcelle (Figure 53) et synthétisés en termes d'espèce de l'arbre dont provient le pollen dans le Tableau 30. L'arbre mère est représenté par un camembert à bordure rouge dont la (les) couleur(s) symbolise(nt) l'espèce (déterminée par assignation génétique). Chaque père assigné à un gland de la famille correspondante est représenté sur la carte de la parcelle de la même façon et un trait reliant le père et la mère symbolise le trajet du grain de pollen. Parallèlement, les résultats des assignations génétiques des glands présentés dans la Figure 52 sont rappelés, en faisant apparaître deux groupes de glands : ceux ayant un père assigné (partie gauche des graphiques en barre) et ceux n'ayant pas de père assigné (partie droite des graphiques en barre). Cette représentation permet de mettre en relation l'assignation génétique des glands et l'espèce des deux parents. De plus, elle permet de comparer les caractéristiques génétiques des deux groupes de glands pour chaque famille.

Parmi les glands analysés, seul quelques-uns sont des pleins-frères, la plupart ont des pères différents. Seuls 9 pères sont assignés pour deux glands alors que dans la descendance 5 (échantillonnées sur la mère 756) un même père est assigné pour 10 glands (Figure 53). Le taux d'autofécondation détecté est très faible puisque seulement 2 glands récoltés sur 2 mères différentes sont le résultat d'autofécondation (mères 634 et 85).

Les résultats des assignations génétiques des glands concordent avec les espèces des pères et des mères. Certains chênes assignés à une espèce produisent des glands assignés à la même espèce, c'est le cas des mères pédonculés 210 et 741 ainsi que de la mère pubescent 192. On remarque cependant que, certains pères étant hybrides, certains glands peuvent posséder une petite proportion de gènes d'une autre espèce. En outre, certaines mères, bien qu'assignées à une espèce, produisent des glands ayant des caractéristiques génétiques intermédiaires dont les pères sont d'une autre espèce, démontrant des événements d'hybridation (mère pédonculé 634, mères tauzins 316 et 377). Les glands récoltés sur des chênes hybrides acceptent du pollen provenant de plusieurs espèces et produisent des glands dont la composition génétique est très variée. Un résultat systématique concerne les hybrides pédonculé x sessile qui sont fertilisés par du pollen provenant en grande majorité de chêne pédonculé (mères 461, 698 et 85). Alors que les descendants des mères 698 et 85 présentent des caractéristiques génétiques intermédiaires entre les espèces pédonculé et sessile, la mère 461 produit presque exclusivement des glands assignés à l'espèce pédonculée. Ce résultat s'explique par le fait que cette mère hybride possède elle-même une plus grande proportion de gènes pédonculés que sessiles (environ  $\frac{3}{4}$ ,  $\frac{1}{4}$ ).

Les autres hybrides peuvent se reproduire avec plusieurs espèces. Ainsi les hybrides pédonculé x pubescent (mères 101 et 39) ont des glands ayant des caractéristiques génétiques intermédiaires entre les deux espèces parentales. En effet, les pères retrouvés sont en majorité soit des chênes

pédonculés soit des chênes pubescents, bien que quelques hybrides pubescent x tauzin ou pédonculé x sessile les fertilisent à l'occasion, produisant des glands pouvant posséder des gènes provenant de trois espèces. Cependant ce phénomène reste rare. Les hybrides pédonculé x tauzin (mère 834) et pubescent x tauzin (mère 706) sont fertilisés par du pollen de chacune des deux espèces parentale mais parmi les pères retrouvés, une des deux espèces prédomine en effectifs suggérant une hybridation directionnelle. Cependant, les glands produits ne montrent pas de grande déviation vers une des deux espèces au niveau de l'assignation génétique. Enfin, les trois hybrides sessile x tauzin analysés acceptent un pollen d'espèce varié, car bien que peu de père aient été retrouvés, ils ont des caractéristiques génétiques variées et les assignations génétiques des glands témoignent d'une hybridation généralisée (mère 756, 594 et 367).

En général, l'assignation génétique entre glands ayant un père assigné et ceux n'en ayant pas est similaire (Figure 52). On note quelques légères différences chez certaines familles, en particulier, la plus forte présence de gènes pubescents pour les glands dont nous n'avons pas retrouvé le père dans la parcelle pour 9 des familles analysées (756, 834, 39, 101, 377, 634, 367, 461 et 706). Bien que les glands n'ayant pas de pères assignés sont difficilement interprétables en terme de flux de pollen provenant de l'extérieur de la parcelle (autres causes possibles de la non assignation d'un père : coupes, chênes non échantillonnés, erreur de génotypage), les chênes pubescents apparaissent comme de bons pollinisateurs.

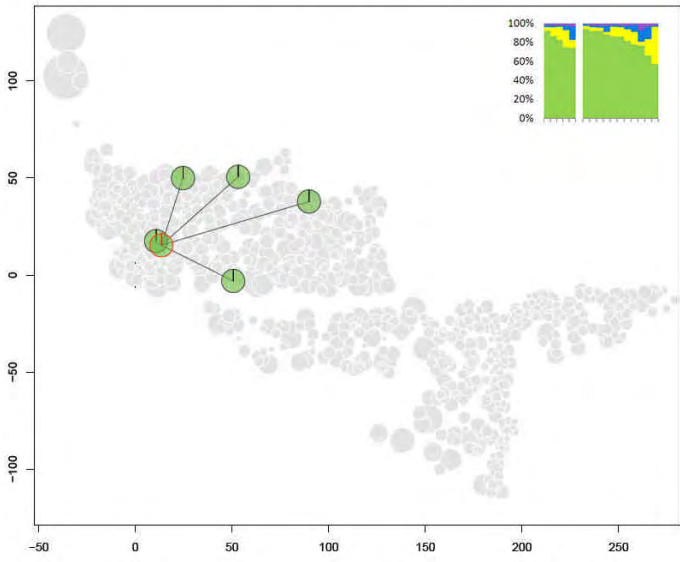
Tableau 29 : Confiance statistique des analyses de paternité et de parenté calculée à partir des fréquences alléliques des 7 microsatellites utilisés.

But	Analyse	Probabilité d'exclusion	Lod score seuil	Erreur de type I	Erreur de type II
Assigner un père connaissant la mère (glands)	Paternité	1.0	4.95	0.02	0.07
Assigner un parent (plantules)	Parenté	0.99858	5.40	0.09	0.15
Assigner un couple de parents (plantules)	Parenté	0.99995	9.50	0.02	0.02

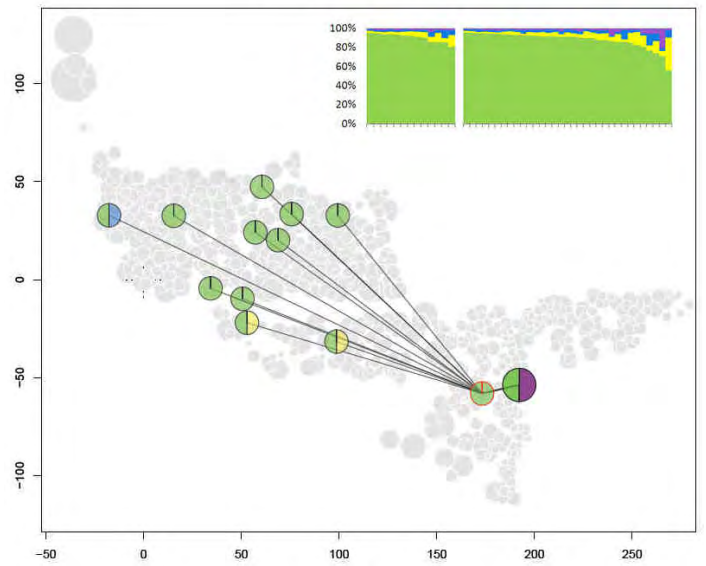
Tableau 30 : Résultats synthétiques de l'analyse de paternité. Pour chaque descendance analysée, le nombre de glands avec un père assigné, ainsi que le statut spécifique (espèces pures, hybrides) des pères est indiqué. Les arbres mères sont regroupés selon leur espèce (pure ou hybride).

Espèce mère	Mère	Desc.	Nombre glands	Nombre glands avec père	% glands sans père retrouvé	Espèce des pères									
						Ped	Ses	Pub	Toz	Hyb PedSes	Hyb PedPub	Hyb PedToz	Hyb SesPub	Hyb SesToz	Hyb PubToz
Ped	210	2	16	5	68.8%	5									
Ped	634	13	33	10	69.7%	5			2	1		1		1	
Ped	741	9	46	13	71.7%	8				2	1	2			
Pub	192	4	69	10	85.5%			7			1				2
Toz	316	1	18	1	94.4%			1							
Toz	377	16	17	6	64.7%				2						4
<b>Espèces</b>	<b>total</b>		<b>199</b>	<b>45</b>	<b>77.4%</b>	<b>18</b>	<b>0</b>	<b>8</b>	<b>4</b>	<b>3</b>	<b>2</b>	<b>3</b>	<b>0</b>	<b>1</b>	<b>6</b>
Hyb PedSes	85	8	42	10	76.2%	9				1					
Hyb PedSes	461	15	29	9	69.0%	9									
Hyb PedSes	698	6	52	11	78.8%	8			1	1		1			
Hyb PedPub	39	12	57	15	73.7%	4		6		1	1				3
Hyb PedPub	101	10	28	7	75.0%	4		2							1
Hyb PedToz	834	11	50	19	62.0%	8		1	4	2	1	1			2
Hyb SesToz	367	3	28	0	100.0%										
Hyb SesToz	594	5	43	3	93.0%			2							1
Hyb SesToz	756	14	50	12	76.0%	1		1	1					5	4
Hyb PubToz	706	7	48	24	50.0%			2	18						4
<b>Hybrides</b>	<b>total</b>		<b>427</b>	<b>101</b>	<b>74.6%</b>	<b>43</b>	<b>0</b>	<b>14</b>	<b>24</b>	<b>5</b>	<b>2</b>	<b>2</b>	<b>0</b>	<b>5</b>	<b>15</b>
<b>TOTAL</b>	<b>16 mères</b>		<b>626</b>	<b>155</b>	<b>75.2%</b>	<b>61</b>	<b>0</b>	<b>22</b>	<b>28</b>	<b>8</b>	<b>4</b>	<b>5</b>	<b>0</b>	<b>6</b>	<b>21</b>

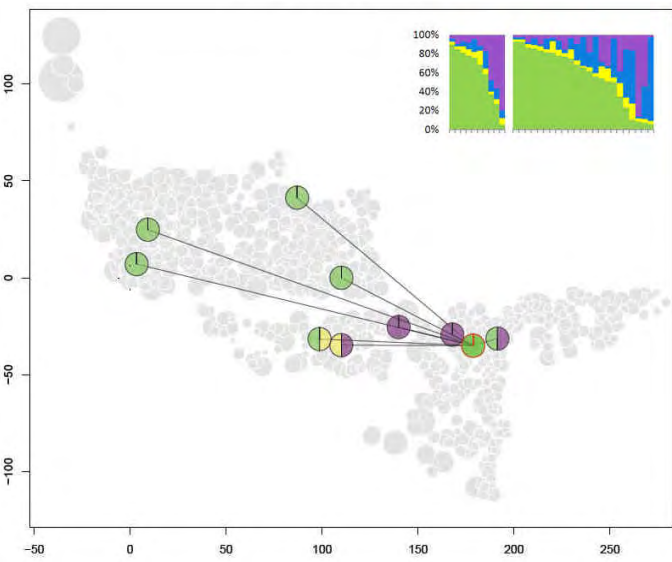
Paternity Briouant - Mother n°210



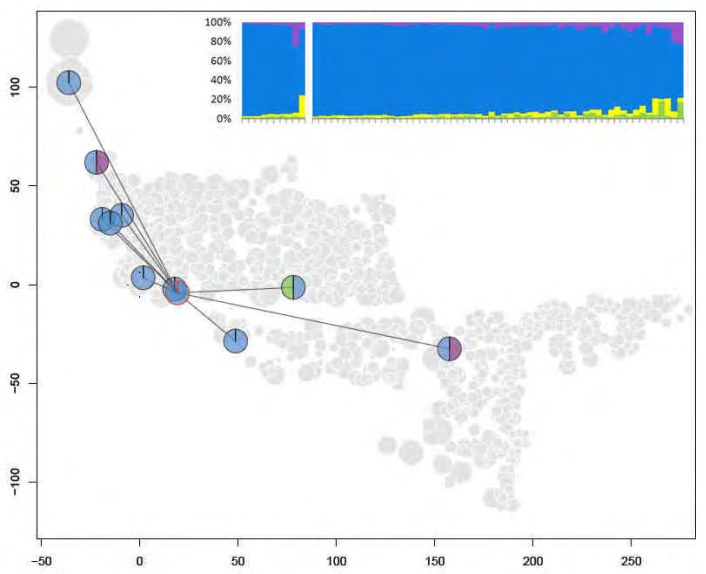
Paternity Briouant - Mother n°741



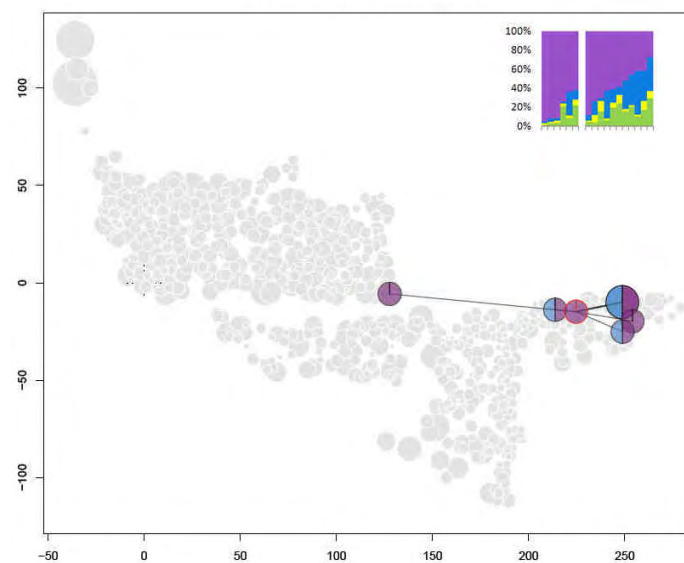
Paternity Briouant - Mother n°634



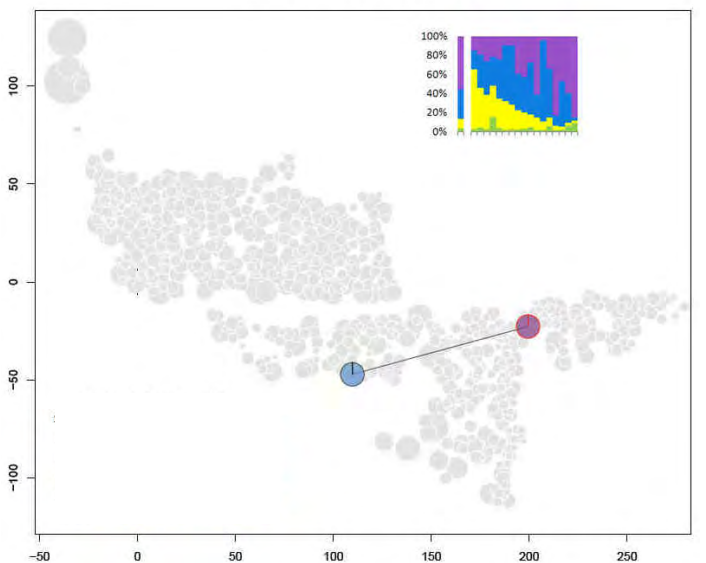
Paternity Briouant - Mother n°192



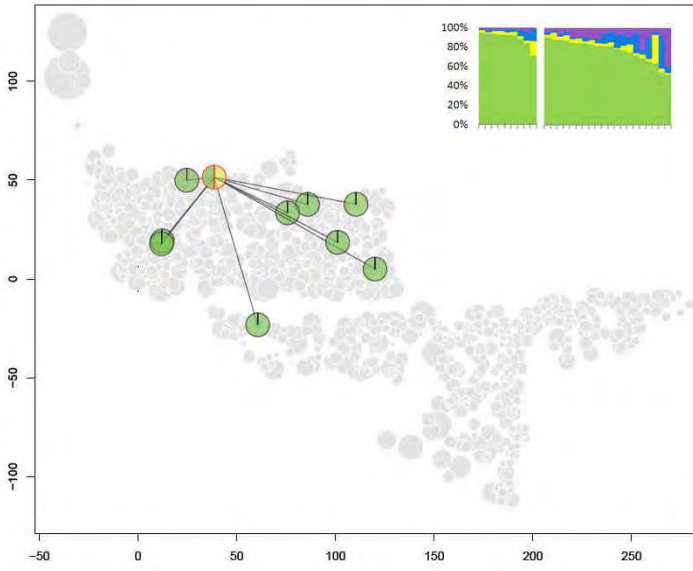
Paternity Briouant - Mother n°377



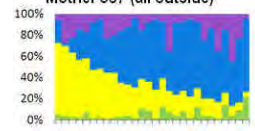
Paternity Briouant - Mother n°316



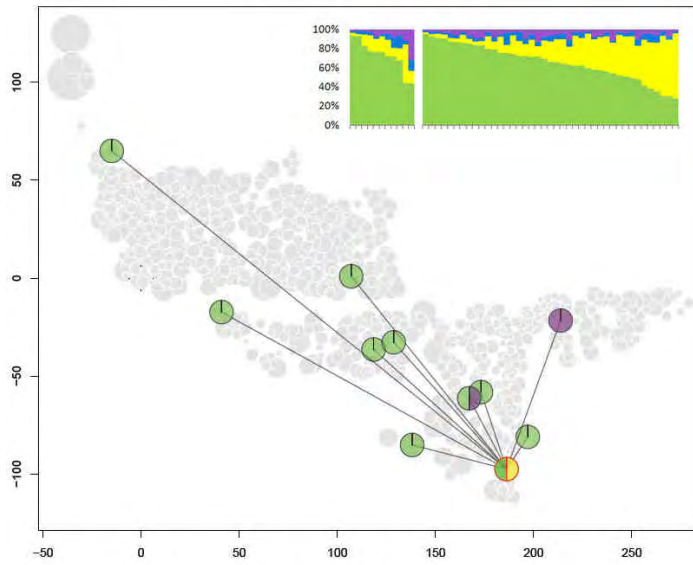
Paternity Briouant – Mother n°461



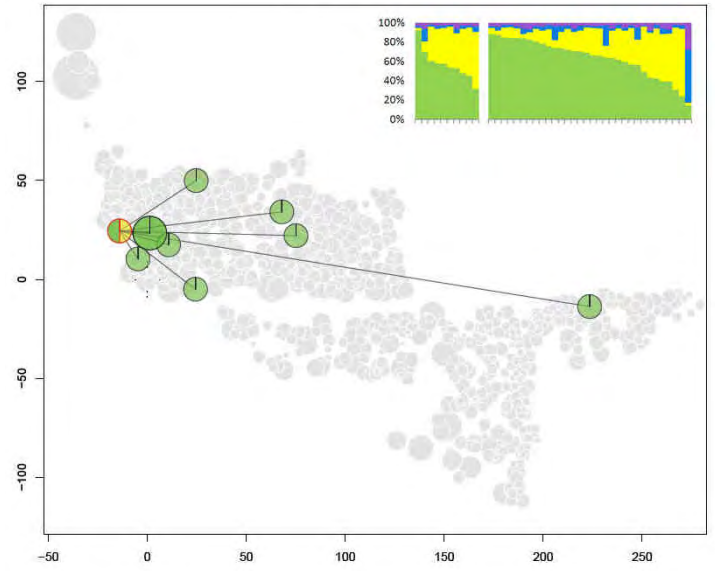
Mother 367 (all outside)



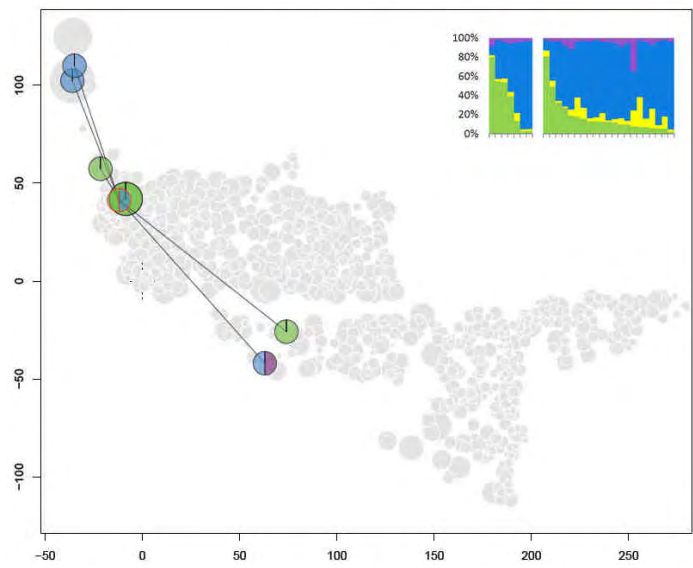
Paternity Briouant – Mother n°698



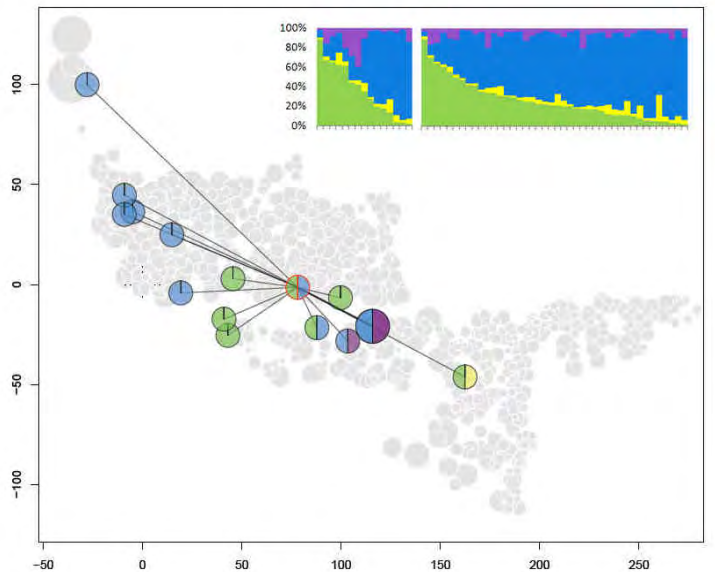
Paternity Briouant – Mother n°85



Paternity Briouant – Mother n°101



Paternity Briouant – Mother n°39



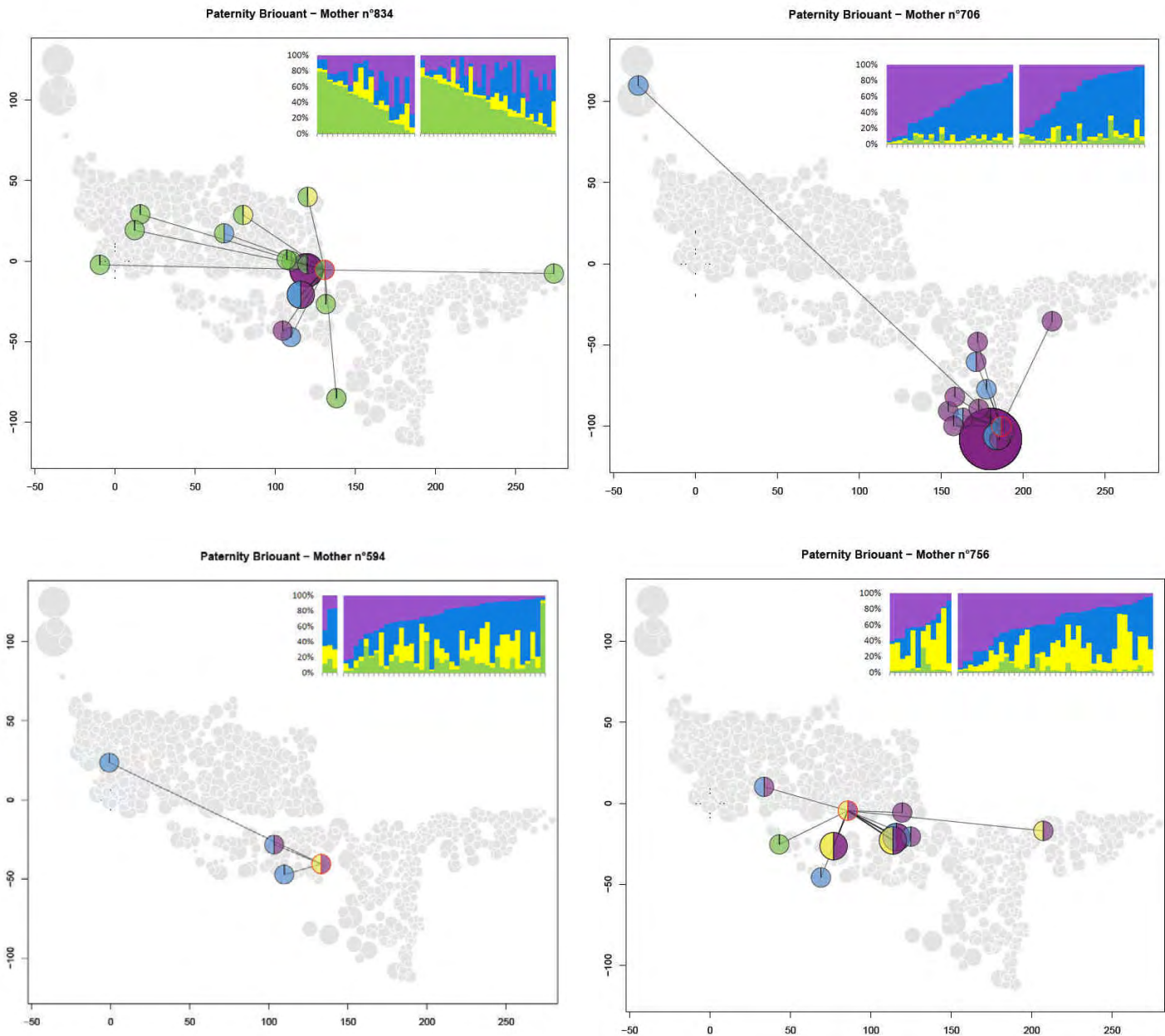


Figure 53 : Résultats des tests de paternité représentés dans le contexte de la parcelle. Chaque carte représente les résultats de la recherche de paternité pour les glands récoltés sur un arbre mère dont l'identifiant est indiqué dans le titre du graphique et la position dans la parcelle localisée par un cercle à la bordure rouge. Les pères identifiés apparaissent sous forme de cercle à bordure noire et le trait reliant père et mère symbolise le trajet parcouru par le pollen. La taille des cercle est proportionnelle au nombre de glands ayant le même père (une paternité : cercle de taille normale, deux paternités : cercle de taille moyenne et 10 paternités : cercle de grande taille). Les couleurs des cercles représentent les résultats de l'assignation génétique des arbres adultes (vert pour chêne pédonculé, jaune pour chêne sessile, bleu pour chêne pubescent et violet pour chêne tauzin). Les résultats de l'assignation génétique des glands sont représentés sous forme de deux graphiques dont chaque barre verticale représente un gland et les couleurs correspondent aux quatre espèces comme indiqué précédemment. Le premier graphique représente les glands dont les pères ont été identifiés, le second représente l'assignation génétique des glands dont le père n'a pas été identifié.



### 7.3.5. Analyse de parenté

Nous avons échantillonné et génotypé 208 plantules dans le but de retrouver un parent ou un couple de parents dont les génotypes sont compatibles avec celui de chaque plantule (test de parenté). Avec les 7 microsatellites utilisés pour cette analyse, la probabilité d'exclusion pour un couple de parents est forte (0.99995, Tableau 29) ce qui permet avec un seuil de LOD score de 9.5 d'assigner un couple de parents compatible avec une bonne précision (erreur de type I et de type II de 2%). Cependant, dans le cas où un seul parent est retrouvé, la probabilité d'exclusion est plus faible (0.9986) et les erreurs de type I et II sont plus élevées (9% et 15% respectivement). En d'autres termes, lorsqu'un seul parent compatible est retrouvé (relation avec un LOD score supérieur au seuil de 5.4), l'assignation risque d'être fautive dans 15% des cas alors que cette erreur est de seulement 2% dans le cas où un couple de parents compatible peut être assigné (Tableau 29).

Une synthèse générale de l'analyse de parenté par placette est présentée dans le tableau 6. Les placettes I, II et IV comprennent des plantules dont une forte proportion a un couple de parents ou un seul parent assigné. Pour ces trois placettes, seules 3 plantules n'ont pas de parent assigné (Tableau 31). La placette II se compose principalement de plantules dont un seul parent est compatible (73.3%), valeur un peu plus faible pour la placette V (54.4%) avec une forte proportion de plantules sans parent retrouvé (29.8%). Placette V exclue, on retrouve au moins un parent compatible pour la grande majorité des plantules analysées (Tableau 31).

Parmi les parents assignés, certains interviennent à plusieurs reprises (Figure 54). Ainsi quatre chênes (211, 443, 316 et 388) sont parents d'au moins 10 plantules. Ces arbres étant à proximité immédiate des placettes échantillonnées (Tableau 27), ce sont probablement les arbres mères dont les glands sont tombés à proximité du tronc, la distance entre les plantules et ces parents n'étant que de quelques mètres. Cette conclusion est confortée par l'analyse des distances dans le cas où un couple de parents (dont un des parents est l'un de ces quatre individus) est assigné à une plantule. Alors que la distance entre la placette et ces parents est faible, le second parent est généralement distant de plusieurs dizaines de mètres, ce qui permet de conclure que l'individu le plus lointain est le père, le pollen se disséminant par le vent sur de plus grande distance (résultats non montrés). Nous avons donc inclus ces quatre familles de demi-frères, représentant le stade post-dispersion, dans les analyses suivantes. Pour ces familles, nous avons identifié un couple de parents (donc en plus de la mère, nous avons identifié le père) pour 66% des plantules (Tableau 32). Comme pour les glands, l'espèce des pères retrouvés est présentée dans le Tableau 32. Les plantules de l'arbre mère pédonculé 211, sont toutes issues de la fertilisation par des chênes pédonculés alors que la mère 443 a été fertilisée 9 fois par des chênes pédonculés et deux fois par des chênes hybrides (Tableau 32). Une proportion non négligeable des plantules de la mère tauzin 388 est issue de croisements avec des arbres hybrides. Enfin, de façon concordante avec l'analyse des glands issus de la mère 316, peu de père pour les plantules issues de cette mère sont retrouvés et en accord avec l'assignation génétique des glands et des plantules de cette

mère (voir Figure 52), le stade plantule démontre également le fort taux d'hybridation de cet individu : parmi les sept plantules dont le père a été identifié, aucun n'est de l'espèce de la mère (6 hybrides l'ont fertilisée et un pubescent, Tableau 32). Enfin, le Tableau 33 regroupe les effectifs des différents croisements observés pour les 85 plantules dont les deux parents sont connus. Les croisements entre chênes tauzins sont les plus nombreux, suivis des croisements entre chênes pédonculés, suivis enfin par des croisements entre chênes tauzin et hybrides pubescent x tauzin, ces trois type de croisements représentant 77% des reproductions identifiées.

Tableau 31 : Résultats synthétique de l'analyse de parenté. Nombre de plantules avec un couple de parents, un seul parent ou aucun parent assignés par placette.

Placette	N	Plantules avec un couple de parents assigné	Plantules avec un parent assigné	Plantules sans parent assigné
I	81	47 (58.0%)	33 (40.7%)	1 (1.3%)
II	24	11 (45.8%)	12 (50.0%)	1 (4.2%)
III	19	2 (10.5%)	14 (73.3%)	3 (15.8%)
IV	27	16 (59.3%)	10 (37.0%)	1 (3.7%)
V	57	9 (15.8%)	31 (54.4%)	17 (29.8%)
<b>total</b>	<b>208</b>	<b>85 (40.9%)</b>	<b>100 (48.1%)</b>	<b>23 (11.0%)</b>



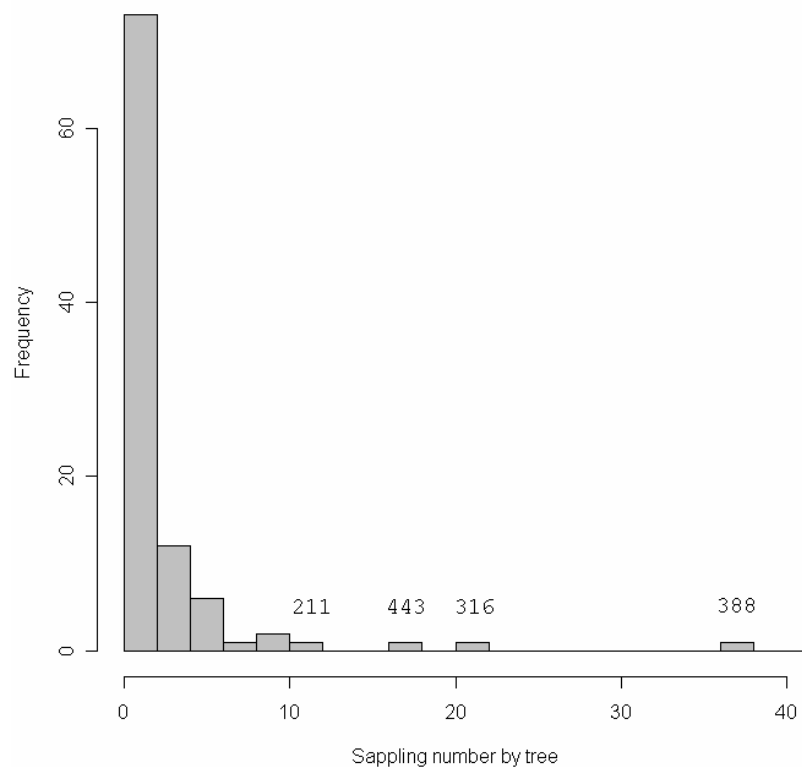


Figure 54 : Nombre de plantules par parent. Quatre parents (supposés arbres mères, voir texte) sont assignés comme parent ou couple de parents pour plus de 10 plantules. Ces quatre familles sont intégrées aux familles composées de glands dans les analyses suivantes.

Tableau 32 : Résultats synthétiques de l'analyse de parenté pour les quatre familles de plantules.

Espèce mère	Mère	Desc.	Nombre plantules	Nombre de plantules avec un couple de parents	% plantules sans père retrouvé	Espèce pères									
						Ped	Ses	Pub	Toz	Hyb PedSes	Hyb PedPub	Hyb PedToz	Hyb SesPub	Hyb SesToz	Hyb PubToz
Toz	388	17	38	33	13.2%				20			2		2	9
Ped	443	18	17	11	35.3%	9				1	1				
Ped	211	19	10	6	40.0%	6									
Toz	316	20	21	7	66.7%			1						4	2
<b>total</b>			<b>86</b>	<b>57</b>	<b>33.7%</b>	<b>15</b>	<b>0</b>	<b>1</b>	<b>20</b>	<b>1</b>	<b>1</b>	<b>2</b>	<b>0</b>	<b>6</b>	<b>11</b>

Tableau 33 : Synthèse des espèces des couples de parents assignés aux plantules.

Nombre plantules	Nombre de plantules avec un couple de parents assigné	Ped x		Pub x		Toz x		Ped x		Toz x		Toz x		hybPubToz x hybPubToz
		Ped	Pub	Pub	Pub	Toz	Toz	hybPedSes	hybPedPub	hybPedToz	hybSesToz	hybPubToz		
208	85	23	1	27	1	1	3	3	7	16	3			

### 7.3.6. Synthèse des événements de reproduction détectés

Lorsqu'un père a été assigné à un gland échantillonné sur un arbre mère (analyse de paternité) ou un couple de parents a été assigné à une plantule (analyse de parenté), nous avons directement accès au type d'événement reproducteur ayant eu lieu. Ceci est rendu possible grâce à l'assignation génétique des chênes adultes (voir chapitre 4) permettant d'assigner les chênes à une espèce ou à une catégorie d'hybride. Nous pouvons ainsi définir 5 types d'événements reproducteurs :

- croisement intra-spécifique : les parents sont de la même espèce.
- hybridation de première génération (F1) : les parents sont assignés à deux espèces différentes.
- hybridation de seconde génération (F2) : les parents sont assignés à un type d'hybride proche (au moins une espèce parentale en commun).
- rétrocroisement : un des parents est assigné à une espèce et l'autre est un hybride dont une partie du génome est composée de gènes provenant de la même espèce que le premier parent.
- tri-hybridation : un des parents est un hybride entre deux espèces et l'autre est assigné à une troisième espèce.

Au total, nous disposons de l'information des deux parents pour 155 glands et 85 plantules permettant de caractériser directement 240 événements reproducteurs (Tableau 34).

Les chênes assignés à une espèce sont en majorité fertilisés par du pollen de leur espèce : 60% des événements reproducteurs identifiés chez ces individus sont des croisements intra-spécifiques. Les rétrocroisements sont fréquents puisque environ 30% du pollen reçu par ces chênes provient de chênes hybrides possédant une partie de leur génome d'espèce commune à l'arbre mère (Tableau 34). Seulement 3 événements d'hybridation de première génération ont été détectés (soit environ 7%) impliquant une mère pédonculé (2 événements avec un individu tauzin) et une mère tauzin (un événement avec un individu pubescent). Enfin, les reproduction impliquant trois espèces (tri-hybridation) sont rares et ne représentent qu'une seule reproduction (pédonculé X hybride sessile x tauzin).

En ce qui concerne la reproduction des mères hybrides, une forte majorité des cas implique du pollen provenant d'un chêne assigné à une des deux espèces dont la mère est issue (rétrocroisement : plus de 68% des cas). Une proportion non négligeable des reproductions implique des croisements entre hybrides de types proches (F2, environ 26%). Finalement, comme dans le cas des mères assignés à une espèce, les cas de tri-hybridation sont rares (5%).

Au total les rétrocroisements représentent la majorité des événements reproducteurs (57%), alors que les croisements intra-spécifiques et entre hybrides (F2) sont identifiés pour environ 18% des cas chacun (Tableau 34). Cependant, nous avons étudié davantage de chênes hybrides, ce qui introduit un biais. La parcelle de Briouant est composée d'environ 70% de chênes assignés à une espèce pure et de 30% d'hybrides (chapitre 4). Sachant que nous avons observé 45 événements reproducteurs pour des chênes assignés à une espèce pure et 101 événements pour des chênes hybrides et en supposant que ces événements reproducteurs observés sont représentatif de ceux qui se produisent dans la parcelle, nous

pouvons calculer les pourcentages des différents événements reproducteurs rapportés à 1000 événements reproducteurs correspondant à 700 individus purs et 300 hybrides. Les proportions suivantes pour les événements reproducteurs seraient observées au stade pré-dispersion :

- croisements intra-spécifiques : 42.0%
- hybridation F1 : 4.7%
- hybridation F2 : 8.6%
- rétrocroisements : 44.1%
- tri-hybridation : 3.3%

Cependant, il ne s'agit là que d'une approximation car le comportement reproducteur apparaît très variable d'un arbre à l'autre : certains individus ont une forte propension à l'hybridation, alors que d'autres non (les trois mères pédonculés par exemple). De plus, le nombre d'arbre par espèce pure ou hybride étudié est faible au regard de la variabilité des stratégies reproductives (seulement 2 chênes tauzins étudiés, un seul chêne pubescent...).

Néanmoins, cette approximation met en évidence que les rétrocroisements sont particulièrement fréquents, que ce soit chez les hybrides ou les arbres d'espèce pures et sont aussi communs que les croisements intra-spécifiques. La formation d'hybrides de première génération concerne environ 5% des événements de reproduction, alors que des croisements plus complexes (F2 et tri-hybridation) représentent tout de même 10% des reproductions.

L'analyse au stade post-dispersion (analyse de parenté pour les plantules) permet de caractériser 85 événements de reproduction (Tableau 34). Les fréquences des différents événements sont remarquablement proches de ce qui a été observé chez les mères assignées à une espèce pure, soit environ 60% de croisements intra-spécifiques et 30% de rétrocroisements. Ce résultat concorde parfaitement avec le fait que sur 85 couples de parents observés, au moins 57 sont issus des 4 familles de demi-frères détaillées à la fin du tableau 9 et comprend seulement des mères pédonculé ou tauzin. Ainsi, l'échantillonnage en placette a privilégié la présence de nombreuses plantules issues d'arbres mères à proximité, qui sont majoritairement des individus non hybrides (voir Tableau 27). Cependant, le fait de retrouver au stade plantules sensiblement les mêmes événements de reproduction qu'au stade gland suggère qu'il n'existe pas de barrières postzygotiques fortes et systématiques, contre les hybrides. Ce résultat confirme celui de l'assignation des glands et des plantules issus de la mère 316, qui ne montrait pas de différences flagrantes (Tableau 28).

Tableau 34 : Nombre d'événements de reproduction observés pour les descendants (glands et plantules) dont les deux parents ont été identifiés.

Espèce mère	Identifiant mère	Nombre de descendants dont les 2 parents sont connus	Evénements reproducteurs				
			Intra.	F1	F2	Rétro.	Trihyb.
Ped	210	5	5	0	-	0	0
Ped	634	10	5	2	-	2	1
Ped	741	13	8	0	-	5	0
Pub	192	10	7	0	-	3	0
Toz	316	1	0	1	-	0	0
Toz	377	6	2	0	-	4	0
<b>Espèces pures</b>	<b>total</b>	<b>45</b>	<b>27</b>	<b>3</b>	<b>-</b>	<b>14</b>	<b>1</b>
			<b>60.0%</b>	<b>6.7%</b>	<b>-</b>	<b>31.1%</b>	<b>2.2%</b>
Hyb PedSes	85	10	-	-	1	9	0
Hyb_PedSes	461	9	-	-	0	9	0
Hyb PedSes	698	11	-	-	2	8	1
Hyb PedPub	39	15	-	-	5	10	0
Hyb PedPub	101	7	-	-	1	6	0
Hyb PedToz	834	19	-	-	6	12	1
Hyb SesToz	367	0	-	-	0	0	0
Hyb SesToz	594	3	-	-	1	0	2
Hyb SesToz	756	12	-	-	9	1	2
hyb PubToz	706	24	-	-	4	20	0
<b>Hybrides</b>	<b>total</b>	<b>110</b>	<b>-</b>	<b>-</b>	<b>29</b>	<b>75</b>	<b>6</b>
			<b>-</b>	<b>-</b>	<b>26.4%</b>	<b>68.2%</b>	<b>5.5%</b>
<b>Total glands</b>	<b>16 mères</b>	<b>155</b>	<b>27</b>	<b>3</b>	<b>29</b>	<b>89</b>	<b>7</b>
			<b>17.4%</b>	<b>1.9%</b>	<b>18.7%</b>	<b>57.4%</b>	<b>4.5%</b>
<b>Total plantules</b>		<b>85</b>	<b>51</b>	<b>1</b>	<b>3</b>	<b>30</b>	<b>0</b>
			<b>60.0%</b>	<b>1.2%</b>	<b>3.5%</b>	<b>35.3%</b>	<b>0%</b>
dont 4 familles de plantules :							
Toz	388	33	20	0	-	13	0
Ped	443	11	9	0	-	2	0
Ped	211	6	6	0	-	0	0
Toz	316	7	0	1	-	6	0
<b>Total partiel</b>	<b>4 mères</b>	<b>57</b>	<b>35</b>	<b>1</b>	<b>-</b>	<b>21</b>	<b>0</b>
			<b>61.4%</b>	<b>1.8%</b>	<b>-</b>	<b>36.8%</b>	<b>0%</b>

Les types d'événements reproducteurs sont décrits en comparant l'espèce ou le type d'hybride des deux parents défini par assignation génétique. Intra : croisement intra-spécifique, F1 : hybridation de première génération, F2 : reproduction entre deux parents hybrides partageant une partie de leur génome, Rétro : rétrocroisement : reproduction entre un chêne assigné à une espèce et un hybride dont une partie du génome est de la même espèce, Trihyb : tri-hybridation : reproduction entre un hybride entre deux espèces et un chêne assigné à une troisième espèce.

### 7.3.7. Corrélations de paternité intra-famille

La caractérisation des événements de reproductions détaillée précédemment tient compte d'une partie des données seulement car nous nous sommes focalisés sur les cas où les deux parents des glands ou des plantules étaient identifiés. Pour étendre l'analyse à l'ensemble des descendants, nous allons utiliser des méthodes d'analyses indirectes utilisant des modèles de reproductions corrélées. Ces méthodes utilisent les génotypes des mères et des descendants afin d'extraire les haplotypes des pollens ayant participé à la reproduction.

La première analyse présentée dans ce paragraphe concerne la corrélation de paternité intra-famille estimée grâce au logiciel KinDist (Robledo-Arnuncio *et al.*, 2006). Cette corrélation estime la relation génétique entre les pollens qui ont fertilisé chaque arbre mère. Cette corrélation estime la proportion de graines d'une mère qui partagent le même père et prend une valeur de 1 lorsque les descendants ont tous le même père (plein-frère) et une valeur de 0 lorsque chaque descendant a un père différent (Hardy *et al.*, 2004).

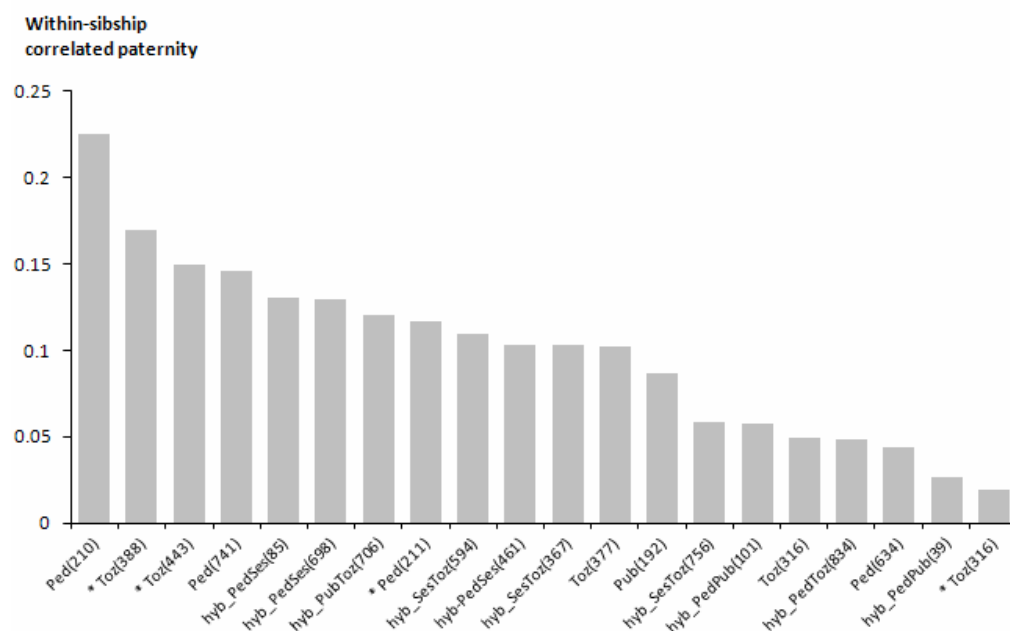


Figure 55 : Corrélations de paternité intra-famille estimées avec KinDist. L'espèce de l'arbre mère ainsi que son identifiant sont reportés en abscisse. Une étoile (\*) indique une famille de plantules, les autres familles étant composées de glands.

Les corrélations pour chaque famille (glands et plantules) analysées sont comprises entre 0.225 et 0.02 (Figure 55). Les corrélations de paternité les plus élevées concernent des familles issues de mères assignées à une espèce pure (glands récoltés sur les mères pédonculés 210 et 741, plantules des chênes tauzins 388 et 443) alors que les familles issues des chênes hybrides présentent des corrélations moyennes. Cependant certaines des familles issues d'arbres d'espèces pures ou hybrides présentent aussi des corrélations de paternité moyennes et faibles. En comparant ces valeurs ainsi que les assignations génétiques des descendants, on s'aperçoit que les corrélations de paternité sont fortes pour les arbres produisant des descendants de la même espèce que l'arbre mère (pédonculés 210 et 741,

tauzins 388 et 443) ou des hybrides subissant des rétrocroisements par du pollen majoritairement d'une seule espèce (hybrides pédonculé x sessile 85, 698 et 461 dans une moindre mesure). Le lien entre corrélation de paternité et nombre de descendants ayant le même père semble dominé par l'effet de l'hybridation comme le montre la relativement faible corrélation de paternité pour la descendance récoltée sur l'arbre 706. En effet, nous avons montré par analyse de paternité que 10 descendants de cette mère partagent le même père, ce qui devrait se traduire par une corrélation de paternité élevée, cependant elle reste moyenne, sûrement amoindrie par le fait que certains pères identifiés appartiennent à deux espèces (pubescent et tauzin, voir Figure 53 mère 706). De façon générale, les familles montrant des corrélations de paternité faibles correspondent à celles dont nous avons identifié des pères appartenant à plusieurs espèces (glands des mères 39, 634, 834, 101 et 756 par exemple, voir Figure 52) et / ou dont l'assignation génétique des descendants montre la présence de gène de deux voire trois espèces dans la descendance (glands et plantules de la mère 316, glands de l'arbre 367). Remarquons que les plantules issues de la mère 316 ont une corrélation de paternité plus faible que les glands récoltés sur le même arbre, ce qui est cohérent avec le fait que les plantules cumulent plusieurs saisons de reproduction alors que les glands récoltés sont issus d'une seule saison de reproduction. Enfin, cette analyse, qui prend en compte le génotype de tous les descendants, confirme les résultats obtenus par l'analyse de paternité et de parenté ainsi que les assignations génétiques des descendants : certains arbres se reproduisent principalement avec une seule espèce (croisement intra-spécifique ou introgression des hybrides pédonculé x sessile) alors que d'autres acceptent du pollen de plusieurs espèces (espèces pures s'hybridant ou hybrides qui se reproduisent avec leurs deux espèces parentales).

### 7.3.8. Différenciations des nuages polliniques

Etant donné que nous disposons de génotypes de descendants récoltés sur plusieurs espèces de chênes incluant des hybrides, une question importante concerne la caractéristique du pollen reçu par les chênes analysés, en particulier les hybrides. Les résultats obtenus précédemment ont montré une forte hétérogénéité dans la composition génétique des descendants en fonction des individus analysés, de leur espèce et de leur statut hybride. En prenant en compte le génotype de chaque chêne mère ainsi que ceux de ses descendants, TwoGener permet d'estimer une différenciation des nuages polliniques,  $\Phi_{FT}$ , en utilisant une approche AMOVA (Excoffier *et al.*, 1992) modifiée qui, au lieu d'utiliser la hiérarchie classique population / individu, va permettre d'analyser la différenciation des contributions paternelles entre arbres mères (hiérarchie mère / pollen) (Smouse *et al.*, 2001). La différenciation  $\Phi_{FT}$  entre nuages polliniques permet d'estimer l'hétérogénéité du pollen reçu par différentes mères.

Comme nous avons 20 descendances analysées (16 issues de glands et 4 issues de plantules), cette analyse donne une matrice de 20 lignes et 20 colonnes comprenant les valeurs de  $\Phi_{FT}$  entre chacune des 20 familles. Une telle matrice est difficilement exploitable, c'est pourquoi nous avons choisi de l'utiliser pour calculer un arbre de distances (Figure 56). Aux numéros des familles identifiant chaque branche de l'arbre, nous avons reporté les résultats des assignations génétiques des arbres mères (camembert) et des descendances (graphiques en barre). Ces informations sont complémentaires car le statut spécifique

de chaque père (auquel nous n'avons pas accès) peut être visualisé en soustrayant l'assignation génétique de la mère à celles des descendants. Les valeurs de différenciation des nuages polliniques sont comprises entre 0 et 0.14 (histogramme, Figure 56) et présentent une distribution relativement continue, ce que traduit bien la forme de l'arbre. Nous pouvons distinguer 3 branches principales dont chacune est caractéristique du pollen d'une espèce. Ainsi, comme nous l'avons vu précédemment, les familles 2 et 9 (mères 210 et 741 respectivement) se caractérisent par un nombre élevé de reproductions intra-pédonculé. La branche supérieure de l'arbre regroupe donc des mères recevant du pollen de chêne pédonculé. Suivant la même logique, la famille 4 récoltée sur le chêne pubescent 192 est composée principalement de reproductions intra-pubescent, et la branche basse de l'arbre regroupe les arbres acceptant du pollen de chêne pubescent. Enfin, la branche supérieure droite de l'arbre rassemble les chênes acceptant du pollen de chêne tauzin (dont la famille 17 représenté par des plantules issues de la mère tauzin 388). Les pollens de chênes tauzins et de chênes pubescents se différencient assez faiblement en comparaison de celui du chêne pédonculé. Ce résultat concorde parfaitement avec la différenciation entre ces trois espèces (voir chapitre 5).

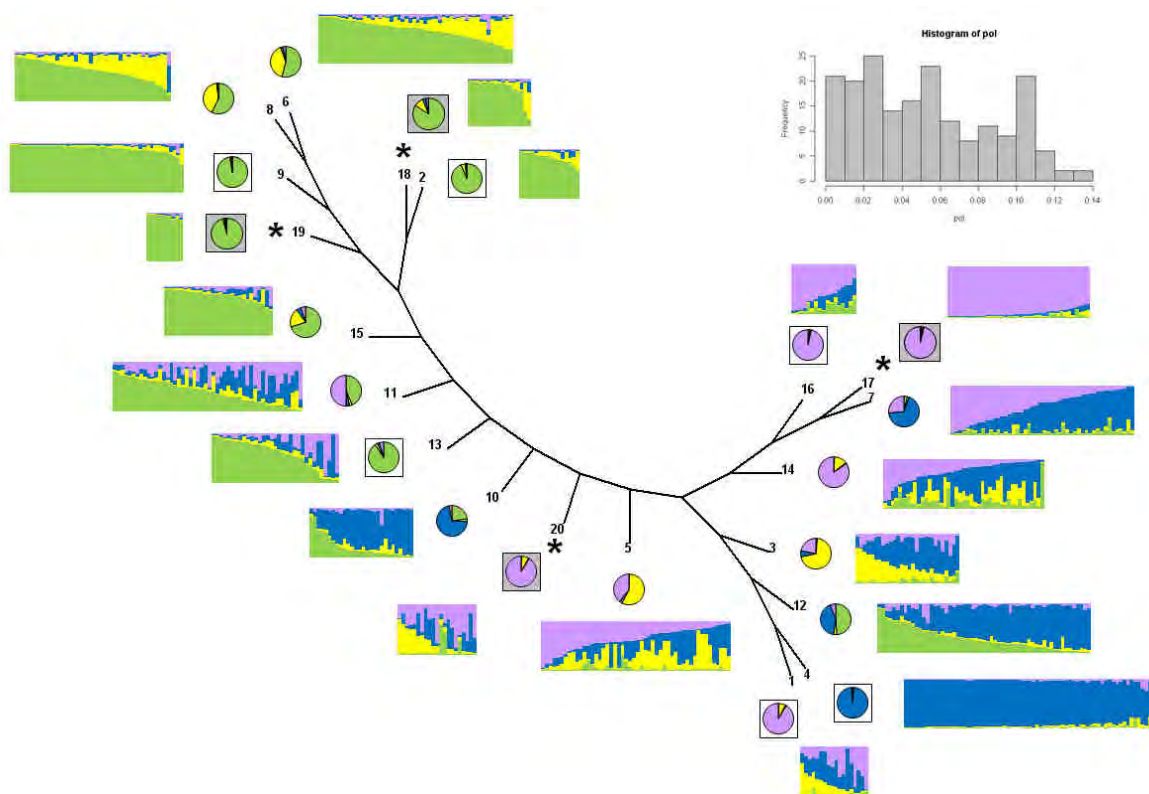


Figure 56 : Distribution des valeurs de  $\Phi_{FT}$  entre paires de familles (histogramme), résultats des assignations génétiques des arbres mères (camemberts, la présence d'un cadre indique les individus assignés à une espèce pure, les cadres grisés représentent les mères de plantules) et des descendants (graphiques en barres) et arbre non enraciné calculé par la méthode de *neighborhood joining* à partir des valeurs de  $\Phi_{FT}$  entre familles. Les feuilles de l'arbre sont identifiées par les numéros des familles (tableau 1), une étoile indique une famille composée de plantules, les autres familles correspondant à des glands. Les couleurs indiquent la proportion du génome de chaque individu (camembert ou graphique en barre) composé de gènes pédonculé (vert), sessile (jaune), pubescent (bleu) ou tauzin (violet).

Une fois la structure de l'arbre ainsi dégrossie grâce à la position de quelques familles caractéristiques, nous pouvons commenter la position des différentes mères. Conformément aux résultats des analyses précédentes, nous remarquons que les hybrides pédonculé x sessile reçoivent exclusivement du pollen de chêne pédonculé, produisant des rétrocroisements. Il en résulte une sorte de dilution de l'espèce sessile : comme nous pouvons le voir avec les résultats des assignations génétiques, la proportion de gènes sessiles dans les descendance diminue par rapport à celle des mères (descendance 6, 8, et 15, mères 698, 85 et 461). L'impact du pollen de chêne tauzin se limite à quelques arbres seulement. En effet, outre la descendance 17, seules les descendance 7 et 16 montrent une bonne correspondance entre la position sur l'arbre et les résultats d'assignation génétique. Le chêne tauzin 377 semble recevoir pour une partie du pollen de chêne tauzin mais les résultats de l'analyse de paternité ont montré qu'il recevait aussi du pollen de plusieurs chênes hybrides pubescent x tauzin et l'assignation génétique de ses glands permet de penser qu'il reçoit du pollen de chênes ayant une partie de gènes pédonculés. La descendance 7, composée de glands récoltés sur un hybride pubescent x tauzin (arbre 706, dont le génome est composé de  $\frac{1}{4}$  de gènes taurins et de  $\frac{3}{4}$  de gènes pubescents), reçoit en grande partie du pollen de chêne tauzin car la différenciation pollinique avec la famille 17 est faible. Ce résultat est confirmé par l'analyse directe (paternité) qui a montré que cet arbre se reproduisait avec une grande majorité de chênes taurins ainsi que quelques pubescents et hybrides pubescent x tauzin et la composition en gènes taurin qui augmente dans la descendance par rapport à la mère.

La branche la plus basse de l'arbre montre que la proportion de gènes pubescents augmente dans l'assignation des descendants, et ce dès la famille 14. En effet, les familles 14, 3, 12, 1, et 20 reçoivent une bonne proportion de pollen de chêne pubescent comme le montrent les assignations des descendants et leur position sur l'arbre à proximité de la branche caractéristique du pollen de chêne pubescent. Par exemple, la famille 1 (chêne tauzin 316) produit des descendants composés systématiquement d'une proportion non négligeable de gènes pubescents ce qui explique la faible différenciation entre son nuage pollinique et celui du chêne pubescent 192 (famille 4). En outre, la famille 12 issue du chêne 39 (hybride pédonculé x pubescent) qui, même s'il est pollinisé par des chênes pédonculé et pubescent, a un nuage pollinique proche du chêne pubescent, indiquant une introgression directionnelle (relative car non exclusive) des gènes pubescents dans la descendance de cet hybride (bien visible dans l'assignation de la descendance). Ainsi, il apparaît nettement que le chêne pubescent a une forte propension à polliniser des chênes de différentes espèces, en particulier les hybrides.

#### **7.4. Discussion**

Le premier objectif de ce chapitre était d'apprécier le rôle de la phénologie dans l'isolement reproducteur des espèces. L'analyse de la phénologie du débournement montre que la reprise de végétation du chêne tauzin est nettement plus tardive que les autres espèces. Ce délai de reprise de végétation entraîne un délai de floraison, ce qui peut créer une barrière à la reproduction entre espèces.



Cependant, les résultats démontrent que les périodes de floraison des espèces se chevauchent partiellement du fait de la grande variabilité du débourrement à l'intérieur des espèces. Les hybrides ont de plus une période de débourrement intermédiaire, ce qui va avoir comme conséquence de diminuer l'éventuel isolement reproducteur temporel entre espèces. Il ne semble donc pas que la phénologie soit une barrière efficace entre ces espèces. Néanmoins, la composition spécifique du pollen dispersé va varier au cours de la période de floraison, influençant la force de la compétition pollinique, ce qui peut augmenter la possibilité d'hybridation en particulier pour les individus ayant une période de floraison atypique pour leur espèce. Enfin, la floraison est précipitée lorsque de la température est élevée (printemps 2006), augmentant le chevauchement des périodes de floraison. Ainsi, l'impact de la phénologie comme barrière pré-reproductive peu dépendre de l'environnement, l'hybridation pouvant être facilitée lors de températures élevées.

Le second objectif de ce chapitre était d'estimer les flux de gènes interspécifiques contemporains en analysant le génotype de glands récoltés sur des chênes appartenant aux 4 espèces et classes d'hybrides. Deux méthodes ont été utilisées afin d'avoir une image instantanée des flux de gènes entre espèces : une méthode directe par analyse de paternité et de parenté, et une méthode indirecte utilisant des modèles de reproductions corrélées.

Connaissant le génotype de tous les chênes présents dans la parcelle (chapitre 4), une analyse de paternité a pu être menée avec l'objectif d'assigner à chaque gland un père compatible, connaissant son génotype ainsi que celui de l'arbre mère. Cette démarche se focalise sur les flux de gènes avant dispersion et germination et a été critiquée car elle pourrait surestimer le taux d'hybridation dans le cas de contre sélection des hybrides (Muir & Schlotterer, 2005). Afin d'obtenir une vision de l'hybridation après l'étape de dispersion et de germination, nous avons analysé 208 plantules afin de trouver un parent ou un couple de parents dont les génotypes sont compatibles avec celui de la plantule. Ces deux analyses, paternité et parenté, ont permis d'identifier les deux parents pour une partie des descendances. Nous obtenons ainsi une image précise de certains des événements de reproduction ayant eu lieu entre chênes adultes à l'intérieur de la parcelle.

Cependant, pour une partie des descendants un seul ou aucun parent au génotype compatible n'a été identifié, laissant de côté une bonne partie des données lors de l'interprétation des analyses directes. Afin de compléter notre étude, nous avons tenté de caractériser de façon plus complète les flux de gènes. Deux modèles de reproduction corrélées ont été utilisés pour comparer les comportements reproducteurs des différentes mères échantillonnées. Contrairement aux analyses directes, ces modèles de reproductions corrélées ne nécessitent pas d'information quant au génotype des parents potentiels. En effet, les analyses sont menées en prenant en compte les génotypes des descendants et plus précisément l'haplotype (génotype haploïde) du pollen ayant fécondé l'arbre mère. Ces méthodes cherchent à comparer la structure génétique du nuage pollinique reçu par chaque arbre mère. Cette approche, dite indirecte, permet donc de considérer les génotypes de l'ensemble des descendants, contrairement aux méthodes directes.

Bien que de méthodologie très différentes, ces deux approches nous ont donné des résultats globalement concordants mais surtout complémentaires, ce qui illustre l'utilité de leur combinaison pour obtenir une image précise de la dynamique des flux de gènes contemporains.

La combinaison de ces techniques avec celle de l'assignation génétique permet de clarifier la nature des événements reproducteurs (assignation génétique des adultes et méthodes directes) et aide grandement à l'interprétation des résultats lorsque les descendants sont également assignés génétiquement pour leur statut spécifique. En effet, l'analyse des génotypes des glands (méthodes directes et indirectes) permet d'avoir une image instantanée du comportement reproducteur d'un nombre limité de chênes, une année donnée. L'assignation génétique est elle capable de détecter un événement d'hybridation au moins deux générations dans le passé : un chêne assigné à une classe d'hybride est issu d'un événement d'hybridation au plus lors des deux générations qui le précède (voir simulations au chapitre 3). Ainsi la combinaison de l'assignation génétique des arbres adultes et des analyses des génotypes des descendants permet d'obtenir une image claire de la dynamique d'hybridation sur au moins trois générations, ce qui est particulièrement rare chez les espèces à long temps de génération comme les chênes. De plus, l'attention portée à la viabilité des glands récoltés (taille des glands, germination) ainsi que l'intégration de plantules dans les analyses permet d'étendre les conclusions au-delà de la simple quantification de l'hybridation au stade pré-dispersion.

La relation entre l'assignation génétique des arbres mères et de leurs descendants a mis en évidence une cohérence globale dans la composition en espèces au niveau du génome de chaque mère ainsi que ces descendants. Cependant, une analyse plus précise des résultats a fait apparaître certains changements de composition génétique entre les deux générations, indiquant d'une part certains événements d'hybridation chez des arbres mères assignées à une espèce pure et des phénomènes d'introggression directionnelle chez certaines mères hybrides (Tableau 28 et Figure 52).

L'identification des événements de reproduction réalisée grâce à l'analyse de paternité et de parenté (méthode directe) a permis de quantifier la proportion de différents types de reproduction chez les arbres mères d'espèces pures et hybrides. Etant donné la surreprésentation des mères hybrides dans l'échantillonnage, la proportion des différents événements de reproduction a été ramenée à la proportion d'hybrides présente dans la parcelle. Cette estimation montre que les rétrocroisements représentent la plus grosse part des événements de reproduction (44%), suivie de près par les reproductions intraspécifiques (42%). Nos résultats montrent que les événements d'hybridation de première génération constituent environ 5% des événements reproducteurs. Enfin, des événements de reproduction plus complexes, impliquant soit deux hybrides ou bien la participation de trois espèces sont assez rares puisqu'ils ne représentent qu'environ 10% des événements identifiés. Ces résultats sont partiellement compatibles avec les proportions d'hybrides trouvés parmi les adultes. En effet, les croisements intraspécifiques vont produire des individus purs, et une partie des rétrocroisements vont aussi produire des individus assignés à une espèce pure (environ 1/3 des rétrocroisements; voir simulations du chapitre 3). Ainsi, environ 57% des descendants produits devraient être assignés à une espèce pure. Les

événements de reproduction du type hybridation F1, F2 ou tri-hybridation vont majoritairement produire des individus assignés comme hybrides (15%) alors que 2/3 des rétrocroisements devraient produire aussi des individus assignés comme hybrides soit un total de 43% d'hybrides au total. Au final, parmi les événements de reproduction identifiés, 57% devrait produire des individus assignés à une espèce pure et 43% des individus assignés comme hybride. Il y a plus d'hybrides identifiés au stade pré-dispersion qu'au stade adulte, cette comparaison doit être interprétée avec précaution au regard du faible nombre d'arbres mères échantillonnés et d'une seule année d'étude pendant laquelle les conditions climatiques ont pu favoriser l'hybridation.

Les résultats des modèles de reproduction corrélées montrent une bonne cohérence avec les résultats des analyses directes. En particulier, l'analyse de la différenciation des nuages pollinique entre arbres mères montre des tendances similaires en termes de comportements reproducteurs. Certains chênes d'espèces pures ou hybrides se reproduisent principalement avec des pères d'une seule espèce (corrélation de paternité intra famille relativement élevée et position extrême sur l'arbre des différenciations des nuages pollinique, Figure 55 et Figure 56 respectivement), et certains hybrides se reproduisent principalement avec des pères d'une seule des deux espèces parentales (cas des hybrides pédonculé x sessile par exemple). A l'inverse, certains hybrides acceptent du pollen provenant de l'une des deux espèces parentales sans barrière reproductive apparente (corrélation de paternité intra famille faible et position sur l'arbre des différenciations du nuage pollinique intermédiaire, Figure 55 et Figure 56). Ces résultats font de plus apparaître que le pollen de chênes pubescents est particulièrement efficace puisque la plupart des hybrides ainsi que certains chêne tauzins sont fécondés par du pollen de cet espèce. Cette tendance apparaît particulièrement pour les descendants dont le père n'a pas été identifié dans la parcelle (voir Figure 53), ce qui pourrait indiquer que le flux de pollen en provenance de l'extérieur de la parcelle contient une bonne part de pollen de chêne pubescent. Ce résultat est concordant avec la surreprésentation de gènes pubescents parmi les hybrides par rapport au nombre d'individus de cette espèce présents dans la parcelle (chapitre 4).

En ce qui concerne la dynamique d'hybridation des autres espèces, on peut remarquer que le chêne sessile est en nette régression dans ce peuplement. D'une part, aucun événement reproducteur impliquant du pollen de chêne sessile n'a été directement détecté. D'autre part, les hybrides ayant une certaine proportion de leur génome composée de gènes sessiles sont systématiquement introgressés par du pollen provenant de l'autre espèce parentale ou de chênes pubescents (cas de tous les hybrides pédonculé x sessile ainsi que des hybrides sessile x tauzin). Les chênes pédonculés analysés montrent une proportion importante de reproduction intraspécifique, excepté un chêne pédonculé (n° 634) qui présente un taux d'hybridation non négligeable avec du pollen de chêne tauzin et de chêne pubescent. Mis à part quelques exceptions, le pollen de chênes pédonculés intervient peu dans les croisements interspécifiques ou dans les rétrocroisements (excepté dans le cas des hybrides pédonculé x sessile discuté précédemment).

Les arbres mères tauzins s'hybrident assez fréquemment et leurs hybrides sont couramment pollinisés par du pollen de chêne pubescent (excepté un arbre mère hybride pubescent x tauzin qui reçoit principalement du pollen de chêne tauzin). Peu de chênes reçoivent du pollen tauzin (Tableau 28, Figure 56). Le fort décalage phénologique de cette espèce (floraison environ 2 à 3 semaines après les autres, chapitre 7) ainsi que la particularité de la présence de cette espèce dans cette région (limite d'aire de distribution, présence de l'espèce dans les zones de pâturages de la parcelle) pourrait expliquer une telle tendance.

Bien que les résultats présentés dans ce chapitre soient majoritairement basés sur des analyses au stade pré-dispersion, l'inclusion de plantules permet de confirmer la dynamique d'hybridation. En particulier, l'analyse comparative de glands et de plantules issus du même arbre mère (hybride sessile x tauzin) ne montre pas de différence significative. Le nombre d'hybride ne varie pas entre les deux stades et de plus, les plantules ont même une composition génétique plus hétérogène que les glands produits une saison donnée (Figure 56). La pesée ainsi que le suivi de germination indiquent que les glands prélevés sur les hybrides sont au moins aussi viables et vigoureux que ceux prélevés sur les arbres d'une espèce pure (Figure 51). Enfin, le fait d'avoir pu récolter des glands viables sur des arbres mères hybrides, ainsi que d'avoir identifié des hybrides comme pères potentiels montre clairement que ces hybrides sont tout à fait fertiles pour leur fonction tant mâle que femelle.

L'ensemble des résultats présentés dans ce chapitre montre clairement que l'hybridation est un phénomène fréquent dans cette parcelle. La quantification des événements reproducteurs permet d'expliquer le taux important d'hybrides parmi les adultes. Malgré l'importance des flux de gènes entre espèces, le comportement reproducteur des hybrides, en particulier leur propension à être fécondés principalement par l'une des deux espèces parentale, conduit à l'introgession de l'une ou l'autre espèce parentale et finalement au retour vers des chênes présentant toutes les caractéristiques d'un chêne d'espèce pure en quelques générations par des rétrocroisements successifs. Ainsi, les hybrides ne se caractérisent pas par une perte des barrières reproductives (qui pourrait conduire à une homogénéisation généralisée des espèces en présence) mais plutôt par une additivité des compatibilités de reproduction entre les deux espèces parentales comme le montre la faible proportion de tri-hybridation. Globalement, la formation d'hybrides de première génération est relativement rare (5% environ), mais ces hybrides, une fois féconds, vont produire une descendance composée uniquement d'hybrides de seconde génération, expliquant le taux important d'hybrides présents parmi les chênes adultes de cette parcelle.

Les résultats présentés dans ce chapitre mettent en évidence que les hybrides sont fertiles et que les flux de gènes entre espèces de chênes sont importants. D'autre part, les flux de gènes entre espèces observés indirectement sur trois générations mettent en évidence une dynamique d'hybridation cohérente avec les résultats présentés de la composition spécifique des arbres adultes dans le peuplement (chapitre 4). Le maintien des espèces malgré des flux de gènes interspécifiques importants semble bien pouvoir s'expliquer par des phénomènes de rétrocroisements successifs, dynamique prédite par la théorie du rôle de l'hybridation comme processus de migration des espèces (Petit *et al.*, 2003).



# Chapitre 8- Conclusion générale



## 8.1. Synthèse des résultats obtenus

Au cours de ce travail, nous avons étudié l'hybridation contemporaine chez les chênes blancs grâce à des marqueurs génétiques mis au point dans ce but (partie A). Nous avons d'une part caractérisé le statut spécifique de chênes adultes en utilisant des méthodes d'assignation génétique (partie B). D'autre part, nous avons analysé le système de reproduction des espèces aussi bien en croisements contrôlés qu'en reproduction libre en forêt (partie C). Ces deux approches complémentaires ont démontré la présence de flux de gènes entre espèces de chênes et ont permis de caractériser la dynamique d'introgession récente.

### 8.1.1. Délimitation des espèces et assignation

Le développement d'un protocole de génotypage rapide a permis d'acquérir suffisamment d'information pour délimiter les espèces de chênes, malgré leur faible différenciation génétique (partie A). Les tests de la méthode d'assignation mise en œuvre au chapitre 3 montrent qu'un petit nombre de marqueurs différenciés sont suffisants pour distinguer les espèces. La réussite de la méthode d'assignation génétique nécessite des données sur au moins une soixantaine d'individus par espèces, ce qui peut clairement poser des difficultés dans les cas d'une espèce peu échantillonnée. Or, dans de nombreuses populations de chênes, plusieurs espèces coexistent mais une ou deux dominent face aux autres, plus rares. Pour résoudre cette difficulté, le regroupement de plusieurs jeux de données (chapitre 4) a été une étape cruciale pour l'obtention d'une assignation génétique cohérente. Il a permis de rassembler suffisamment d'individus pour délimiter chacune des espèces, même celles faiblement représentées dans certaines populations.

Nous avons choisi le critère génétique pour identifier les espèces lors de cette thèse. Nous n'avons donc utilisé aucun autre a priori que le génotype des individus (*blind test approach*, Duminil *et al.*, 2006). Nous avons pu également vérifier la concordance entre les assignations génétiques et morphologiques qui conduisent globalement à la même délimitation des espèces. Cependant, contrairement à la morphologie, l'analyse génétique est plus précise et permet d'identifier les individus hybrides. Les simulations du chapitre 3 montrent qu'il est ainsi possible de distinguer les hybrides de première génération (F1) et de seconde génération (rétrocroisements, F2) des individus d'espèce pure. Il n'est cependant pas possible de différencier les hybrides de première génération de ceux de seconde génération, ce qui nécessiterait un plus grand nombre de marqueurs génétiques. L'assignation génétique avec dix marqueurs microsatellites permet donc d'identifier les événements d'hybridation remontant aux deux générations précédentes mais pas au-delà. Notre méthode d'assignation, adaptée à notre contexte grâce aux simulations, nous a ainsi permis de quantifier l'hybridation récente.

Suite aux déterminations génétiques du statut spécifique des individus, une analyse morphologique fine a pu être conduite (chapitre 5). Les hybrides ont globalement une morphologie intermédiaire entre les espèces parentales. Cependant, nous avons également observé une grande variabilité de formes : certains hybrides ont une forme typique d'une espèce et d'autres individus d'espèce pure présentent une



morphologie intermédiaire. Le seul critère morphologique n'est donc pas adapté pour une étude détaillée de l'hybridation.

### 8.1.2. Hybridation – introgression

Nous avons trouvé entre 15 et 30% d'hybrides dans les peuplements adultes étudiés (chapitre 4). Sur l'ensemble des populations analysées, nous avons identifié des hybrides impliquant les quatre espèces. Toutes les espèces deux à deux peuvent donc former des hybrides dans la nature. Nous nous sommes ensuite intéressés à l'effet des effectifs des espèces sur l'hybridation dans trois parcelles. Nous avons mis en évidence que les espèces dominantes s'hybridaient relativement peu entre elles (Briouant : pédonculé et tauzin ; Petite Charnie : pédonculé et sessile ; Paguères : pédonculé et pubescent ; Figure 57 flèche 3), par contre ces dernières formaient beaucoup d'hybrides avec les espèces plus rares (Briouant et Paguères : sessile ; Petite Charnie : pubescent ; Figure 57 flèches 2 et 4). Nous avons aussi démontré la présence de certaines espèces uniquement à l'état hybride (Paguères et Petite Charnie : tauzin), parfois à plusieurs dizaines de kilomètres de l'aire de distribution connue de l'espèce (Figure 57 flèche 1). Ces résultats peuvent être liés à une dispersion récente sur de longues distances (par graine ou plus probablement par pollen) ou bien à une rémanence de gènes, témoins d'une ancienne aire plus étendue (Briouant : sessile).

Afin de caractériser plus finement la dynamique d'hybridation, nous avons effectué des analyses globales pour parer aux limites de l'assignation génétique (distinction impossible des différentes générations d'hybrides au niveau individuel). Les caractéristiques génétiques des hybrides identifiés dans les parcelles montrent qu'il s'agit pour la plupart d'hybrides de seconde génération (chapitre 4) et que la composition génétique des hybrides est influencée par l'effectif relatif des espèces parentales en présence. Une espèce fréquente va produire davantage de rétrocroisements qu'une espèce rare ce qui va contribuer à une l'introgression directionnelle voire unidirectionnelle (Figure 57 points 4 et 2 respectivement).

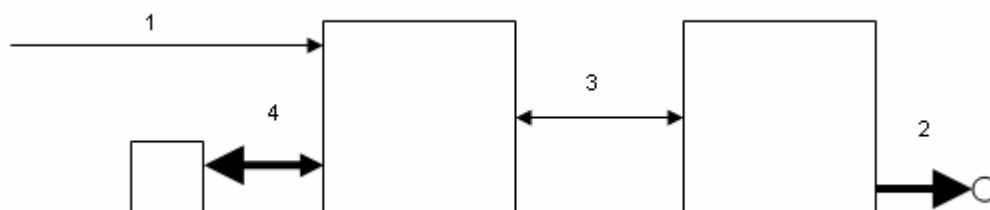


Figure 57 : Schéma synthétique des flux de gènes interspécifiques en lien avec l'effectif relatif des espèces. Les formes représentent des espèces de chênes différentes dont la taille est proportionnelle à leur effectif (les carrés représentent des espèces composées de nombreux individus, le cercle représente un ou quelques individus). Les flèches symbolisent les flux de gènes interspécifiques. 1 : hybridation à longue distance (dispersion de pollen ou migration de graine) ; 2 : hybridation d'une espèce très rare par les espèces fréquentes (introgression unilatérale) ; 3 : hybridation limitée entre espèces fréquentes, introgression bidirectionnelle (équilibre) ; 4 : hybridation d'une espèce minoritaire et d'une espèce fréquente, introgression directionnelle (déséquilibre).

L'analyse de paternité en population naturelle a permis de préciser les raisons du fort pourcentage d'hybrides détecté dans la parcelle de Briouant (chapitre 7). Ces analyses ont confirmé que la fréquence d'hybridation F1 était faible (5% des événements de reproduction à l'échelle de la parcelle) alors que les rétrocroisements étaient aussi fréquents que les reproductions intraspécifiques pour les 16 descendances analysées au stade graines. Les hybrides F1 sont plus rares au stade plantule (1.2% sur 4 descendances) qu'au stade graine (6.7% sur 6 descendances) ce qui pourrait indiquer un rôle précoce de la sélection contre les hybrides. Cependant, ces différences pourraient aussi être liées à des biais d'estimation (hybridations F1 particulièrement fréquentes sur les mères analysées pour une seule année de récolte de glands, plantules échantillonnées sous certains arbres mères, ...). La comparaison des descendants d'un même arbre mère au stade graine et plantule ne montre pas de signe de sélection contre les hybrides. Bien que les hybrides de première génération soient rares, certains atteignent l'âge adulte et sont capable de se reproduire, en formant de nombreux hybrides de seconde génération (rétrocroisements). L'hybridation n'est pas une impasse évolutive chez les chênes puisque les hybrides transmettent leurs gènes à la génération suivante, assurant un flux de gènes interspécifique.

### 8.1.3. Isolement reproducteur

La caractérisation génétique de chênes adultes et de leurs descendances a permis de retracer a posteriori les événements de reproduction ayant eu lieu lors des expériences de croisements contrôlés ou dans la nature.

Le faible nombre de reproductions interspécifiques (hybridation F1) identifiées en forêt (chapitre 7) montre que les espèces sont partiellement isolées reproductivement. Grâce aux croisements contrôlés (chapitre 6), nous avons identifié certains facteurs impliqués dans l'isolement reproducteur, en particulier l'isolement prézygotique. Seuls deux croisements n'ont pas produit de descendance (sessile x pédonculé et sessile x tauzin) et paraissent totalement incompatibles en conditions contrôlées. En population naturelle ces croisements ont pu être observés, ce qui montre que les résultats obtenus en conditions contrôlées ne sont pas forcément généralisables. Pour les autres combinaisons, la comparaison des croisements intra- et interspécifiques a montré qu'une incompatibilité génétique partielle freinait la productivité des croisements interspécifiques (soit passivement : dysfonctionnement, non coordination liée à la différence génétique ; ou activement : mécanisme de discrimination du pollen ; Figure 58 points 2 et 3). Nous avons d'autre part démontré l'existence d'un mécanisme actif grâce aux expériences de compétition pollinique. Lorsque plusieurs espèces de pollen sont disponibles, la préférence au pollen conspécifique réduit le taux d'hybridation (chapitre 6 ; Figure 58 point 2). Les taux d'hybridation F1 sont de 22% sur les chênes pédonculé, et 40% chez les chênes sessiles et tauzins en conditions contrôlées. Ces chiffres sont nettement supérieurs aux 6.7% estimés en population (chapitre 7). Cette différence s'explique par le fait que nous avons appliqué un mélange équilibré des pollens des différentes espèces. Dans ces conditions, le pollen conspécifique est minoritaire ce qui peut expliquer le taux d'hybridation élevé. De plus, l'existence en population de barrières pré-reproductives peut aussi expliquer le faible taux d'hybridation (Figure 58 point 1). Nous avons montré que la période de floraison est différente entre les espèces, ce qui pourrait contribuer à l'isolement pré-reproducteur.

Cependant, ce caractère est très variable entre individus d'une même espèce ce qui se traduit par un chevauchement des périodes de floraison. Ainsi, le pollen libéré dans l'air sera constitué d'une proportion variable dans le temps et l'espace de mélanges de plusieurs espèces. Chaque arbre mère expérimentera donc un degré de compétition pollinique différent en fonction de sa localisation spatiale et de la période de réceptivité de ses fleurs femelles. Les individus présentant une phénologie moyenne auront davantage de pollen conspécifique à leur disposition et la compétition pollinique sera alors favorable aux reproductions intraspécifiques. A l'opposé, les chênes qui présentent une floraison extrême (ou isolés géographiquement) auront relativement peu de pollen conspécifique à disposition, ce qui augmentera le taux d'hybridation. Nos résultats suggèrent que la proportion de pollen libéré (qui est liée aux effectifs des espèces parentales et au déroulement temporel des floraisons) influence la dynamique d'hybridation. Ainsi les barrières pré-reproductives telles que la phénologie et l'environnement vont influencer l'intensité de la compétition pollinique, la fréquence d'hybridation F1 ainsi que la dynamique d'introgression (interaction entre barrières pré-reproductives et prézygotiques Figure 58). Par exemple, une température atmosphérique élevée au cours de la période de floraison va avoir tendance à précipiter la floraison et diminuer les différences phénologiques entre espèces (comparaison de la phénologie des années 2006 et 2007, chapitre 7). Les barrières pré-reproductives seront alors moindres lors de printemps chauds, ce qui va intensifier la compétition pollinique et potentiellement augmenter la fréquence d'hybridation. Une analyse génétique des pollens pourrait permettre de mieux comprendre la dynamique de reproduction des espèces. Une expérience préliminaire a montré qu'il était possible d'identifier les allèles présents dans un mélange pollinique de 5 à 20 individus (mélanges utilisés en croisements contrôlés, résultats non présentés). Le piégeage du pollen circulant en forêt à différents moments de la période de reproduction (de la Cruz *et al.*, 2008) permettrait d'accéder aux fréquences alléliques du mélange pollinique grâce à un développement méthodologique approprié et de donner une idée de sa composition spécifique. Ainsi il serait possible d'estimer l'impact de la compétition pollinique dans l'isolement pré-reproducteur entre espèces. Deux approches sont possibles : l'analyse génétique sur grain de pollen unique (Matsuki *et al.*, 2007) ou de mélange de pollens (*pool DNA genotyping* ; Khatib *et al.*, 1994 ; Band & Ron, 1998 ; Breen *et al.*, 1999 ; Checa *et al.*, 2000 ; Hillel *et al.*, 2003 ; Skalski *et al.*, 2006). Une étude prospective devrait être entreprise pour estimer la faisabilité de cette approche.

Les analyses des systèmes de reproduction en populations naturelles (chapitre 7) démontrent que les hybrides sont reproductivement compatibles avec leurs deux espèces parentales alors qu'ils restent isolés des autres espèces. Ce relâchement limité des barrières reproductives chez les hybrides conduit à de nombreux rétrocroisements. Cette caractéristique de la reproduction des hybrides intervient probablement dans le maintien des espèces parentales.

Nous avons également estimé un faible impact des barrières postzygotiques en conditions artificielles lors des premiers stades de vie des descendances. Cependant, la sélection liée au milieu peut être un facteur important en population bien qu'elle soit difficile à estimer (Figure 58 point 4 et 5). Nous avons néanmoins observé que des individus hybrides participaient à la reproduction. Certains hybrides passent

donc le stade de la sélection et sont suffisamment viables et adaptés pour transmettre leurs gènes à la génération suivante. Cependant des études supplémentaires, telles que des transplantations réciproques des espèces et des hybrides dans les environnements caractéristiques de chaque espèce (Rieseberg & Carney, 1998), permettraient de mieux appréhender l'importance de l'environnement dans l'isolement postzygotique et son impact dans la sélection contre les hybrides. Cependant, de telles expériences sont difficiles à mettre en œuvre chez des espèces à long temps de génération comme les chênes.

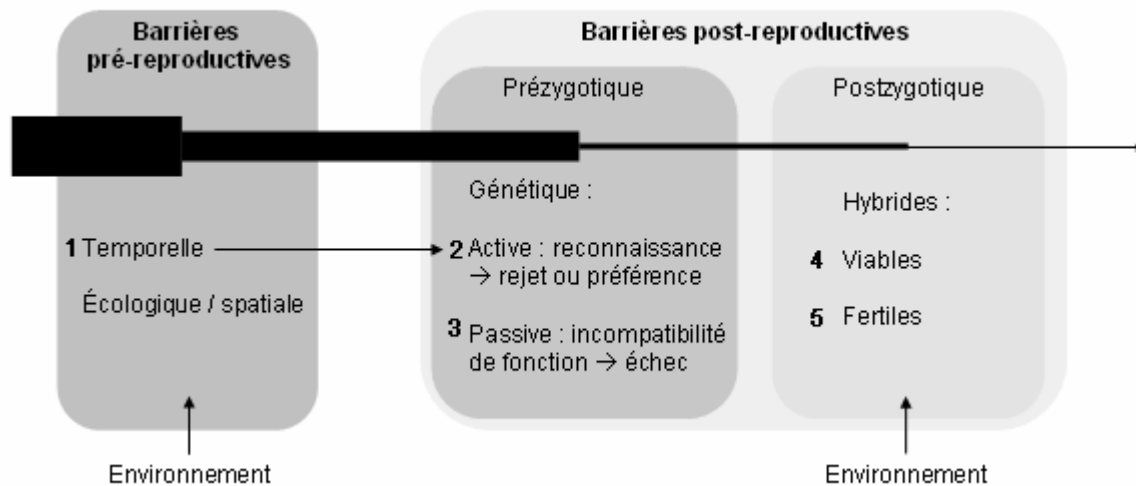


Figure 58 : Schéma synthétique des barrières reproductives entre espèces de chênes mises en évidence lors de ce travail. L'épaisseur de la flèche indique le taux d'hybridation potentiel diminuant sous l'effet des différentes barrières reproductives. Les numéros indiquent les différentes étapes mises en évidences dans notre étude (voir texte).

## 8.2. Conséquences de l'hybridation sur l'évolution du complexe d'espèces

Nous avons montré que l'hybridation est un phénomène qui existe chez les chênes blancs, qui reste relativement rare mais qui a des conséquences évolutives importantes car les frontières entre espèces sont poreuses aux flux de gènes. L'introgression est dans certains cas directionnelle voire unilatérale ce qui confirme le rôle possible de l'hybridation comme processus de dispersion des espèces (Petit *et al.*, 2003). Cependant, le modèle d'hybridation unilatéral proposé entre chênes sessile et pédonculé n'est pas confirmé par nos résultats. Nous avons en effet trouvé que l'hybridation entre ces deux espèces pouvait avoir lieu dans les deux sens. Plus généralement, nous avons montré que la directionnalité de l'introgression dépendait fortement des conditions locales (les effectifs des espèces ont un fort impact sur ce phénomène) et probablement des effets environnementaux. Les taux d'hybridation ainsi que la directionnalité d'introgression sont très variables en fonction des populations. Les conditions européennes postglaciaires ont probablement favorisé une hybridation unidirectionnelle entre chêne pédonculés et sessiles. Nos résultats montrent de plus que l'hybridation et l'introgression sont des processus contemporains à l'œuvre parmi les espèces du complexe des chênes blancs.

La présence de flux de gènes interspécifique est en accord avec l'hypothèse que la différenciation des espèces est liée à une sélection divergente. Certaines zones du génome contenant des gènes impliqués dans l'adaptation au milieu seraient fortement différenciées entre espèces et imperméables à

l'introgression. A l'opposé, la différenciation génétique dans le reste du génome serait faible et les gènes présents dans ces zones perméables aux flux de gènes sont potentiellement échangeables entre espèces car n'intervenant pas dans l'expression des caractères adaptatifs différenciant les espèces. Etant donné le niveau de flux de gènes détecté, la sélection divergente doit jouer un rôle important dans le maintien des espèces. Cette forte sélection pourrait concerner certains gènes impliqués dans l'adaptation des espèces à leur milieu (humidité et nature chimique du sol, tolérance à la sécheresse, ...). Les gènes liés à l'isolement entre espèces (phénologie, reconnaissances des gamètes, ...) seraient intéressants à étudier pour comprendre la différenciation moléculaire entre espèces. Des zones génomiques particulièrement différenciées et contrôlant des caractères morphologiques différenciant les espèces ont été localisées sur plusieurs chromosomes (Saintagne *et al.*, 2004). Une prochaine étape consisterait à identifier certains gènes impliqués dans l'adaptation au milieu ou dans l'isolement reproducteur afin de les localiser sur la carte génétique et de comparer leur position par rapport à ceux impliqués dans les différences morphologiques. Des tests de détection d'événements de sélection pourraient ensuite être entrepris pour préciser si les différences morphologiques et l'isolement reproducteur sont des sous-produits de la sélection divergente pour des caractères écologiques adaptatifs.

L'évolution des espèces du complexe des chênes blancs a été marquée par les changements climatiques qui se sont succédés au cours du quaternaire (de 2.4 Ma BP à nos jours). Lors des 800000 dernières années, huit alternances de phases glaciaires et interglaciaires (Louergue *et al.*, 2008) ont profondément contraint l'évolution des espèces (Hewitt, 2000 ; Davis & Shaw, 2001). Face à ces changements climatiques de grande amplitude et parfois rapides (Steffensen *et al.*, 2008), les espèces se sont adaptées aux nouvelles conditions et ont migré vers des zones où les conditions climatiques étaient supportables. Ainsi, lors des périodes glaciaires (d'une durée de 100000 ans environ) les espèces de chênes étaient localisées au sud des péninsules européennes, isolées les unes des autres, permettant une divergence. A l'inverse, alors que lors des périodes interglaciaires (d'une durée de 10000 ans) l'aire de répartition des chênes était nettement plus étendue et les populations provenant de plusieurs refuges ont été remises en contact. Ces alternances de phases d'extension et de contraction des aires de distribution ont probablement participé à l'évolution des espèces. Les méthodes utilisées lors de cette thèse pourraient être appliquées à du matériel ancien retrouvé dans des couches sédimentaires datées (pollen : Bennett & Parducci, 2006 ; ou bois : Liepelt *et al.*, 2006) afin de mieux comprendre l'évolution génétique de ces espèces, leur répartition lors du dernier épisode glaciaire ou à d'autres époques, ainsi que l'impact de l'hybridation lors des phases d'expansion et de contraction des aires. L'analyse génétique à partir d'ADN ancien (Parducci & Petit, 2004 ; Gugerli *et al.*, 2005) a montré son potentiel chez quelques espèces d'arbres (sapin : Suyama *et al.*, 1996 ; pin sylvestre : Paffetti *et al.*, 2007; Parducci *et al.*, 2005 ; hêtre : Paffetti *et al.*, 2007) et pourrait apporter des résultats nouveaux sur l'évolution du complexe des chênes blancs européens.

## Références bibliographiques

- Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS (2007) Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology Notes* **7**, 10-14.
- Anderson E (1948) Hybridization of the habitat. *Evolution* **2**, 1-9.
- Anderson E, Hubricht L (1938) Hybridization in *Tradescantia*. III. The evidence for introgressive hybridization. *American Journal of Botany* **25**, 396-402.
- Anderson E, Stebbins J (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.
- Arnold J (1993) Cytonuclear Disequilibria in Hybrid Zones. *Annual Review of Ecology and Systematics* **24**, 521-553.
- Arnold M, Robinson J, Buckner C, Bennet B (1992) Pollen dispersal and interspecific gene flow in Louisiana irises. **68**, 399-404.
- Arnold ML (1992) Natural Hybridization as an Evolutionary Process. *Annual Review of Ecology and Systematics* **23**, 237-261.
- Arnold ML (1997) *Natural hybridization and evolution* Oxford University Press, New York, USA.
- Arnold ML (2006) *Evolution through genetic exchange* Oxford University Press, USA, Oxford.
- Arnold ML, Hodges SA (1995) Are natural hybrids fit or unfit relative to their parents? *Trends in Ecology & Evolution* **10**, 67-71.
- Austerlitz F, Smouse PE (2001) Two-generation analysis of pollen flow across a landscape. II. Relation between  $\Phi$ (ft), pollen dispersal and interfemale distance. *Genetics* **157**, 851-857.
- Austerlitz F, Smouse PE (2002) Two-generation analysis of pollen flow across a landscape. IV. Estimating the dispersal parameter. *Genetics* **161**, 355-363.
- Avise JC (2004) *Molecular markers, natural history, and evolution*, Second Edition edn. Sinauer Associates, Inc., Sunderland MA.
- Bacilieri R, Ducouso A, Kremer A (1995) Genetic, Morphological, Ecological and Phenological Differentiation between *Quercus-Petraea* (Matt) Liebl and *Quercus-Robur* L in a Mixed Stand of Northwest of France. *Silvae Genetica* **44**, 1-10.
- Bacilieri R, Ducouso A, Kremer A (1996a) Comparison of morphological characters and molecular markers for the analysis of hybridization in sessile and pedunculate oak. *Annales Des Sciences Forestieres* **53**, 79-91.
- Bacilieri R, Ducouso A, Petit RJ, Kremer A (1996b) Mating system and asymmetric hybridization in a mixed stand of European oaks. *Evolution* **50**, 900-908.
- Bacles CFE, Ennos RA (2008) Paternity analysis of pollen mediated gene flow for *Fraxinus excelsior* L. in a chronically fragmented landscape. *Heredity in press*.
- Band M, Ron M (1998) Determination of allele frequency from DNA pools using bovine trinucleotide microsatellites. *Animal Biotechnology* **9**, 35-45.
- Banks MA, Eichert W (2000) WHICHRUN (version 3.2): A computer program for population assignment of individuals based on multilocus genotype data. *Journal of Heredity* **91**, 87-89.
- Banks MA, Eichert W, Olsen JB (2003) Which genetic loci have greater population assignment power? *Bioinformatics* **19**, 1436-1438.
- Baric S, Monschein S, Hofer M, Grill D, Via JD (2008) Comparability of genotyping data obtained by different procedures an inter-laboratory survey. *Journal of Horticultural Science & Biotechnology* **83**, 183-190.
- Barreneche T, Bodenes C, Lexer C, Trontin JF, Fluch S, Streiff R, Plomion C, Roussel G, Steinkellner H, Burg K, Favre JM, Glossl J, Kremer A (1998) A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers. *Theoretical and Applied Genetics* **97**, 1090-1103.
- Barreneche T, Casasoli M, Russell K, Akkak A, Meddour H, Plomion C, Villani F, Kremer A (2004) Comparative mapping between *Quercus* and *Castanea* using simple-sequence repeats (SSRs). *Theoretical and Applied Genetics* **108**, 558-566.
- Barton NH, Hewitt GM (1985) Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics* **16**, 113-148.
- Belkir K, Borsa P, Chiki L, Raufaste N, Bonhomme F (1996-2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).
- Bennett KD, Parducci L (2006) DNA from pollen: principles and potential. *Holocene* **16**, 1031-1034.
- Beurton PJ (2002) Ernst Mayr through time on the biological species concept - a conceptual analysis. *Theory in Biosciences* **121**, 81-98.
- Bigelow RS (1965) Hybrid Zones and Reproductive Isolation. *Evolution* **19**, 449-458.
- Boccacci P, Akkak A, Marinoni DT, Bounous G, Botta R (2004) Typing European chestnut (*Castanea sativa* Mill.) cultivars using oak simple sequence repeat markers. *Hortscience* **39**, 1212-1216.
- Bodenes C, Joandet S, Laigret F, Kremer A (1997) Detection of genomic regions differentiating two closely related oak species *Quercus petraea* (Matt) Liebl and *Quercus robur* L. *Heredity* **78**, 433-444.

- Bomblies K (2006) Hybrid incompatibility: When opposites attract with a fatal outcome. *Current Biology* **16**, R542-R544.
- Bomblies K, Weigel D (2007) Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nature Reviews Genetics* **8**, 382-393.
- Bonin A, Bellemain E, Eidesen PB, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* **13**, 3261-3273.
- Bossera I (1979) Jays and oaks: an eco-ethological study of a symbiosis. *Behaviour* **70**, 1-116.
- Bousquet J, Simon L, Lalonde M (1990) DNA amplification from vegetative and sexual tissues of trees using polymerase chain reaction. *Canadian Journal of Forest Research* **20**, 254-257.
- Bradshaw HD, Schemske DW (2003) Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature* **426**, 176-178.
- Breen G, Sham P, Li T, Shaw D, Collier DA, St Clair D (1999) Accuracy and sensitivity of DNA pooling with microsatellite repeats using capillary electrophoresis. *Molecular and Cellular Probes* **13**, 359-365.
- Brewer S, Cheddadi R, de Beaulieu JL, Reille M (2002) The spread of deciduous *Quercus* throughout Europe since the last glacial period. *Forest Ecology and Management* **156**, 27-48.
- Broquet T, Petit E (2004) Quantifying genotyping errors in noninvasive population genetics. *Molecular Ecology* **13**, 3601-3608.
- Bruschi P, Vendramin GG, Bussotti F, Grossoni P (2000) Morphological and molecular differentiation between *Quercus petraea* (Matt.) Liebl. and *Quercus pubescens* Willd. (Fagaceae) in Northern and Central Italy. *Annals of Botany* **85**, 325-333.
- Buchan JC, Archie EA, Van Horn RC, Moss CJ, Alberts SC (2005) Locus effects and sources of error in noninvasive genotyping. *Molecular Ecology Notes* **5**, 680-683.
- Buggs RJA (2007) Empirical study of hybrid zone movement. *Heredity* **99**, 301-312.
- Burger WC (1975) The species concept in *Quercus*. *Taxon* **24**, 45-50.
- Burgess KS, Morgan M, Deverno L, Husband BC (2005) Asymmetrical introgression between two *Morus* species (*M. alba*, *M. rubra*) that differ in abundance. *Molecular Ecology* **14**, 3471-3483.
- Burke JM, Arnold ML (2001) Genetics and the fitness of hybrids. *Annual Review of Genetics* **35**, 31-52.
- Camus A (1936) *Monographie du genre Quercus*.
- Carney SE, Hodges SA, Arnold ML (1996) Effects of Differential Pollen-Tube Growth on Hybridization in the Louisiana Irises. *Evolution* **50**, 1871-1878.
- Casasoli M, Derory J, Morera-Dutrey C, Brendel O, Porth I, Guehl JM, Villani F, Kremer A (2006) Comparison of quantitative trait loci for adaptive traits between oak and chestnut based on an expressed sequence tag consensus map. *Genetics* **172**, 533-546.
- Cavalli-Sforza LL, Edwards AW (1967) Phylogenetic analysis. Models and estimation procedures. *Am J Hum Genet* **19**.
- Chagné D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera MT, Vendramin GG, Garcia V, Frigerio JMM, Echt C, Richardson T, Plomion C (2004) Cross-species transferability and mapping of genomic and cDNA SSRs in pines. *Theoretical and Applied Genetics* **109**, 1204-1214.
- Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* **24**, 621-631.
- Checa LM, Carleos C, Baro A J, Dunner S, Canon J (2000) Estimation of allele frequencies in pools of DNA by processing the results of capillary electrophoresis. *ITEA Produccion Animal* **96A**.
- Chen C, Durand E, Forbes F, Francois O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes* **7**, 747-756.
- Chen CB, Forbes F, Francois O (2006) FASTER: model-based clustering made faster. *Molecular Ecology Notes* **6**, 980-983.
- Clark NL, Aagaard JE, Swanson WJ (2006) Evolution of reproductive proteins from animals and plants. *Reproduction* **131**, 11-22.
- Coart E, Lamote V, De Loose M, Van Bockstaele E, Lootens P, Roldan-Ruiz I (2002) AFLP markers demonstrate local genetic differentiation between two indigenous oak species *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl in Flemish populations. *Theoretical and Applied Genetics* **105**, 431-439.
- Comes HP, Kadereit JW (1998) The effect of quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science* **3**, 432-438.
- Cooper A, Poinar HN (2000) Ancient DNA: Do it right or not at ALL. *Science* **289**, 1139-1139.
- Corander J, Waldmann P, Martinen P, Sillanpaa MJ (2004) BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics* **20**, 2363-2369.
- Cornuet JM, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* **153**, 1989-2000.
- Cottam WP, Tucker JM, Santamour FS (1982) *Oak hybridization at thee University of Utah State Arboretum of Utah, Salt Lake City, USA*.
- Coyne JA, Orr HA (1997) "Patterns of speciation in *Drosophila*" revisited. *Evolution* **51**, 295-303.
- Craft KJ, Ashley MV, Koenig WD (2002) Limited hybridization between *Quercus lobata* and *Quercus douglasii* (Fagaceae) in a mixed stand in central coastal California. *American Journal of Botany* **89**, 1792-1798.

- Curtu AL, Gailing O, Finkeldey R (2007) Evidence for hybridization and introgression within a species-rich oak (*Quercus* spp.) community. *BMC Evolutionary Biology* **7**, 218.
- Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. *Heredity* **93**, 504-509.
- Darwin CR (1863) Letter 3917 to Alphonse de Candolle. In: *Darwin correspondance project*.
- Davis MB, Shaw RG (2001) Range shifts and adaptive responses to Quaternary climate change. *Science* **292**, 673-679.
- Davison A, Chiba S (2003) Laboratory temperature variation is a previously unrecognized source of genotyping error during capillary electrophoresis. *Molecular Ecology Notes* **3**, 321-323.
- Day T, Schluter D (1995) The fitness of hybrids. *Trends in Ecology & Evolution* **10**, 288.
- de Candolle A (1862a) Etude sur l'espèce à l'occasion d'une révision de la faamille des Cupulifères. *Bibliothèque Universelle et Revue Suisse (Archives des Sciences Physiques et Naturelles)* **15**, 326-365.
- de Candolle A (1862b) Letter 3603 to Darwin, 13 June 1862. In: *Darwin Correspondance Project*.
- de la Cruz DR, Reyes ES, Sanchez JS (2008) Aerobiological study of Fagaceae pollen in the middle-west of Spain. *Aerobiologia* **24**, 67-76.
- De Warnaffe GD, Deconchat M, Ladet S, Balent G (2006) Variability of cutting regimes in small private woodlots of south-western France. *Annals of Forest Science* **63**, 915-927.
- Degen B, Streiff R, Ziegenhagen B (1999) Comparative study of genetic variation and differentiation of two pedunculate oak (*Quercus robur*) stands using microsatellite and allozyme loci. *Heredity* **83**, 597-603.
- Delmotte F, Leterme N, Simon JC (2001) Microsatellite allele sizing: Difference between automated capillary electrophoresis and manual technique. *Biotechniques* **31**, 810-818.
- den Ouden J, Jansen PA, Smit R (2005) Jays, mice and oaks: Predation and dispersal of *Quercus robur* and *Q. petraea* in north-western Europe. In: *Seed fate prédation, dispersal and seedling establishment*, pp. 223-240. CABI Publishing.
- Depamphilis CW, Wyatt R (1989) Hybridization and Introgression in Buckeyes (*Aesculus*, Hippocastanaceae) - a Review of the Evidence and a Hypothesis to Explain Long-Distance Gene Flow. *Systematic Botany* **14**, 593-611.
- Depamphilis CW, Wyatt R (1990) Electrophoretic Confirmation of Interspecific Hybridization in *Aesculus* (Hippocastanaceae) and the Genetic-Structure of a Broad Hybrid Zone. *Evolution* **44**, 1295-1317.
- DeWoody J, Nason JD, Hipkins VD (2006) Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes* **6**, 951-957.
- Dobzhansky T (1937) *Genetics and the origin of species*. Columbia University Press, New York, USA.
- Dodd RS, Afzal-Rafii Z (2004) Selection and dispersal in a multispecies oak hybrid zone. *Evolution* **58**, 261-269.
- Ducouso A, Petit RJ (1994) Le geai des chênes, premier reboiseur européen. *Forêt-entreprise* **97**, 60-64.
- Duhamel du Monceau HL (1755) *Traité des arbres et arbustes qui se cultivent en France en pleine terre*, Paris.
- Duminil J, Caron H, Scotti I, Cazal SO, Petit RJ (2006) Blind population genetics survey of tropical rainforest trees. *Molecular Ecology* **15**, 3505-3513.
- Dumolin S, Demesure B, Petit RJ (1995) Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theoretical and Applied Genetics* **91**, 1253-1256.
- Dumolin-Lapègue S, Demesure B, Fineschi S, LeCorre V, Petit RJ (1997) Phylogeographic structure of white oaks throughout the European continent. *Genetics* **146**, 1475-1487.
- Dumolin-Lapegue S, Kremer A, Petit RJ (1999) Are chloroplast and mitochondrial DNA variation species independent in oaks? *Evolution* **53**, 1406-1413.
- Dupouey JL, Badeau V (1993) Morphological variability of oaks (*Quercus robur* L, *Quercus petraea* (Matt) Liebl, *Quercus pubescens* Willd) in northeastern France: preliminary results. *Annales Des Sciences Forestieres* **50**, 35s-40s.
- Dzialuk A, Chybicki I, Burczyk J (2005) PCR multiplexing of nuclear microsatellite loci in *Quercus* species. *Plant Molecular Biology Reporter* **23**, 121-128.
- Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. *Heredity* **99**, 125-132.
- Endler JA, Houde AE (1995) Geographic-Variation in Female Preferences for Male Traits in *Poecilia-Reticulata*. *Evolution* **49**, 456-468.
- Ewen KR, Bahlo M, Treloar SA, Levinson DF, Mowry B, Barlow JW, Foote SJ (2000) Identification and analysis of error types in high-throughput genotyping. *American Journal of Human Genetics* **67**, 727-736.
- Excoffier L, Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics* **7**, 745-758.
- 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.
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.
- FIF-ENGREF P (2006) Caractéristiques de la parcelle de Briouant, Exposé de restitution à l'INRA de Toulouse, UMR DYNAFOR.
- Gaget M, Villar M, Dumas C (1989) The mentor pollen phenomenon in poplars: a new concept. *Theoretical and Applied Genetics* **78**, 129-135.
- Gerard PR, Klein EK, Austerlitz F, Fernandez-Manjarres JF, Frascaria-Lacoste N (2006) Assortative mating and differential male mating success in an ash hybrid zone population. *Bmc Evolutionary Biology* **6**.



- 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.
- Ghosh S, Karanjawala ZE, Hauser ER, Ally D, Knapp JI, Rayman JB, Musick A, Tannenbaum J, Te C, Shapiro S, Eldridge W, Musick T, Martin C, Smith JR, Carpten JD, Brownstein MJ, Powell JI, Whiten R, Chines P, Nylund SJ, Magnuson VL, Boehnke M, Collins FS (1997) Methods for precise sizing, automated binning of alleles, and reduction of error rates in large-scale genotyping using fluorescently labeled dinucleotide markers. *Genome Research* **7**, 165-178.
- Gilder JR, Doom TE, Inman K, Krane DE (2007) Run-specific limits of detection and quantitation for STR-based DNA testing. *Journal of Forensic Sciences* **52**, 97-101.
- Gomory D, Yakovlev I, Zhelev P, Jedinakova J, Paule L (2001) Genetic differentiation of oak populations within the *Quercus robur/Quercus petraea* complex in Central and Eastern Europe. *Heredity* **86**, 557-563.
- Gonzalez M (2006) *Effet des pratiques de gestion forestière et de l'organisation des paysages agricoles sur la structure et la dynamique de la végétation des îlots boisés.*, Thèse de l'Université Toulouse III.
- Gonzalez-Rodriguez A, Arias DM, Valencia S, Oyama K (2004) Morphological and RAPD analysis of hybridization between *Quercus affinis* and *Q. laurina* (fagaceae), two Mexican red oaks. *Am. J. Bot.* **91**, 401-409.
- Good TP, Ellis JC, Annett CA, Pierotti R (2000) Bounded hybrid superiority in an avian hybrid zone: effects of mate, diet and habitat choice. *Evolution* **54**, 1774-1783.
- Govaerts R, Andrews S, Coombes A, Gilbert M, Hunt D, Nixon K, Thomas M (2006) World Checklist of Fagaceae. The Board of Trustees of the Royal Botanic Gardens, Kew. Published on the Internet; <http://www.kew.org/wcsp/> accessed 30 May 2008; 15:00 GMT.
- Grant PR, Grant BR (1997) Genetics and the origin of bird species. *Proceedings of the National Academy of Sciences* **94**, 7768-7775.
- Gugerli F, Parducci L, Petit RJ (2005) Ancient plant DNA: review and prospects. *New Phytologist* **166**, 409-418.
- Guillot G, Estoup A, Mortier F, Cosson JF (2005a) A spatial statistical model for landscape genetics. *Genetics* **170**, 1261-1280.
- Guillot G, Mortier F, Estoup A (2005b) GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes* **5**, 712-715.
- Guyon JP, Bessière F, Faignaert O, Bois JC (1996) Evolution des formations boisées sur le territoire de deux communes du canton d'Aurignac (Haute Garonne) de 1942 à 1992. *Etudes et Recherches sur les Systèmes Agraires et le Développement* **29**, 139-152.
- Gyllensten U, Wilson AC (1987) Interspecific mitochondrial DNA transfer and the colonization of Scandinavia by mice. *Genetical Research* **49**, 25-29.
- Hancock CN, Kondo K, Beecher B, McClure B (2003) The S-Locus and Unilateral Incompatibility. *Philosophical Transactions: Biological Sciences* **358**, 1133-1140.
- Hardy OJ, Gonzalez-Martinez SC, Colas B, Freville H, Mignot A, Olivieri I (2004) Fine-scale genetic structure and gene dispersal in *Centaurea corymbosa* (Asteraceae). II. Correlated paternity within and among sibships. *Genetics* **168**, 1601-1614.
- 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.
- Harrison RG (1993) *Hybrid zones and the evolutionary process* Oxford University Press, New York, USA.
- Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution* **59**, 1633-1638.
- Hempel G, Wilhelm K (1889) *Die Bäume und Sträucher des Waldes.*, Vienne.
- Heslop-Harrison J, Heslop-Harrison Y (1970) Evaluation of pollen viability by enzymatically induced fluorescence; intracellular hydrolysis of fluorescein diacetate. *Stain Technology* **45**, 115-120.
- Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907-913.
- Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68**, 87-112.
- Hillel J, Groenen MAM, Tixier-Boichard M, Korol AB, David L, Kirzhner VM, Burke T, Dirie AB, Crooijmans R, Elo K, Feldman M, Freidlin PJ, Maki-Tanila A, Oortwijn M, Thomson P, Vignal A, Wimmers K, Weigend S (2003) Biodiversity of 52 chicken populations assessed by microsatellite typing of DNA pools. *Genetics Selection Evolution* **35**, 533-557.
- Hoffman JI, Amos W (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology* **14**, 599-612.
- Holderegger R, Wagner HH (2006) A brief guide to landscape genetics. *Landscape Ecology* **21**, 793-796.
- Howard DJ (1999) Conspecific sperm and pollen precedence and speciation. *Annual Review of Ecology and Systematics* **30**, 109-132.
- Howard DJ, Preszler RW, Williams J, Fenchel S, Boecklen WJ (1997) How discrete are oak species? Insights from a hybrid zone between *Quercus grisea* and *Quercus gambelii*. *Evolution* **51**, 747-755.
- Huelsenbeck JP, Andolfatto P (2007) Inference of population structure under a Dirichlet process model. *Genetics* **175**, 1787-1802.
- Ishida TA, Hattori K, Sato H, Kimura MT (2003) Differentiation and hybridization between *Quercus crispula* and *Q. dentata* (Fagaceae): Insights from morphological traits, amplified fragment length polymorphism markers, and leafminer composition. *American Journal of Botany* **90**, 769-776.

- Jensen JS, Olrik DC, Siegismund HR, Lowe AJ (2003) Population genetics and spatial autocorrelation in an unmanaged stand of *Quercus petraea* in Denmark. *Scandinavian Journal of Forest Research* **18**, 295-304.
- Jin L, Chakraborty R (1994) Estimation of Genetic-Distance and Coefficient of Gene Diversity from Single-Probe Multilocus DNA-Fingerprinting Data. *Molecular Biology and Evolution* **11**, 120-127.
- Johnson MA, Preuss D (2002) Plotting a Course: Multiple Signals Guide Pollen Tubes to Their Targets. *Developmental Cell* **2**, 273-281.
- Johnson PCD, Haydon DT (2007) Maximum-likelihood estimation of allelic dropout and false allele error rates from Microsatellite genotypes in the absence of reference data. *Genetics* **175**, 827-842.
- Jones EW (1959) Biological flora of the British Isles - *Quercus* L. *The Journal of Ecology* **47**, 169-222.
- Kalinowski ST (2004) Counting alleles with rarefaction: Private alleles and hierarchical sampling designs. *Conservation Genetics* **5**, 539-543.
- Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* **5**, 187-189.
- Kampfer S, Lexer C, Glossl J, Steinkellner H (1998) Characterization of (GA)<sub>n</sub> microsatellite loci from *Quercus robur*. *Hereditas* **129**, 183-186.
- Kanno M, Yokoyama J, Suyama Y, Ohyama M, Itoh T, Suzuki M (2004) Geographical distribution of two haplotypes of chloroplast DNA in four oak species (*Quercus*) in Japan. *Journal of Plant Research* **117**, 311-317.
- Kay KM (2006) Reproductive isolation between two closely related hummingbird-pollinated neotropical gingers. *Evolution* **60**, 538-552.
- Khatib H, Darvasi A, Plotski Y, Soller M (1994) Determining Relative Microsatellite Allele Frequencies in Pooled DNA Samples. *Pcr-Methods and Applications* **4**, 13-18.
- Kimpton CP, Gill P, Walton A, Urquhart A, Millican ES, Adams M (1993) Automated DNA Profiling Employing Multiplex Amplification of Short Tandem Repeat Loci. *Pcr-Methods and Applications* **3**, 13-22.
- Kleinschmit J, Kleinschmit JGR (2000) *Quercus robur* - *Quercus petraea*: A critical review of the species concept. *Glasnik Za Smske Pokuse* **37**, 441-452.
- Knox RB, Willing RR, Ashford AE (1972) Role of Pollen-Wall Proteins as Recognition Substances in Interspecific Incompatibility in Poplars. **237**, 381-383.
- Koumi P, Green HE, Hartley S, Jordan D, Lahec S, Livett RJ, Tsang KW, Ward DM (2004) Evaluation and validation of the ABI 3700, ABI 3100 and the MegaBACE 1000 capillary array electrophoresis instruments for use with short tandem repeat microsatellite typing in a forensic environment. *Electrophoresis* **25**, 2227-2241.
- Kraak SBM, Mundwiler B, Hart PJB (2001) Increased number of hybrids between benthic and limnetic three-spined sticklebacks in Enos Lake, Canada; the collapse of a species pair? *Journal of Fish Biology* **58**, 1458-1464.
- Kremer A (2007) La migration des chênes. In: *Les Amphis de France 5*. SIS - Grand Nancy Terre de Sciences et d'Innovation - Université Nancy 2 - Vidéoscop - <http://www.canal-u.education.fr>.
- Kremer A, Dupouey JL, Deans JD, Cottrell J, Csaikl U, Finkeldey R, Espinel S, Jensen J, Kleinschmit J, Van Dam B, Ducouso A, Forrest I, de Heredia UL, Lowe AJ, Tutkova M, Munro RC, Steinhoff S, Badeau V (2002a) Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands. *Annals of Forest Science* **59**, 777-787.
- Kremer A, Petit RJ, Ducouso A (2002b) Biologie évolutive et diversité génétique des chênes sessile et pédonculé. *Revue Forestière Française* **2**, 111-130.
- Langella O (1999) Populations, Version 1.2.28 - <http://www.pge.cnrs-gif.fr/bioinfo/populations/index.php>.
- Lanner RM, Phillips AM (1992) Natural Hybridization and Introgression of Pinyon Pines in Northwestern Arizona. *International Journal of Plant Sciences* **153**, 250-257.
- Lavabre J (2005) Conséquences de la gestion forestière sur la structure génétique d'une chênaie, p. 21. UMR BIOGECO 1202 - INRA & Université Bordeaux 1, Cestas.
- LeCorre V, Machon N, Petit RJ, Kremer A (1997) Colonization with long-distance seed dispersal and genetic structure of maternally inherited genes in forest trees: a simulation study. *Genetical Research* **69**, 117-125.
- Leigh F, Lea V, Law J, Wolters P, Powell W, Donini P (2003) Assessment of EST- and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat. *Euphytica* **133**, 359-366.
- Lepais O, Lavabre J, Gonzalez M, Willm J, Cabanettes A, Gerber S (2006a) Diversité et structuration génétique des chênes à l'échelle de deux paysages : impact de l'écologie, de l'histoire et de la gestion. *Les Actes du BRG* **6**, 543-557.
- Lepais O, Leger V, Gerber S (2006b) Short note: High throughput microsatellite genotyping in oak species. *Silvae Genetica* **55**, 238-240.
- Lexer C, Kremer A, Petit RJ (2006) Shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Molecular Ecology* **15**, 2007-2012.
- Liepert S, Sperisen C, Deguilloux MF, Petit RJ, Kissling R, Spencer M, Beaulieu JL, Taberlet P, Gielly L, Ziegenhagen B (2006) Authenticated DNA from ancient wood remains. *Annals of Botany* **98**, 1107-1111.
- Lin TH, Myers EW, Xing EP (2006) Interpreting anonymous DNA samples from mass disasters - probabilistic forensic inference using genetic markers. *Bioinformatics* **22**, E298-E306.
- Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial Genetic-Structure of a Tropical Understory Shrub, *Psychotria Officinalis* (Rubiaceae). *American Journal of Botany* **82**, 1420-1425.

- Loulergue L, Schilt A, Spahni R, Masson-Delmotte V, Blunier T, Lemieux B, Barnola J-M, Raynaud D, Stocker TF, Chappellaz J (2008) Orbital and millennial-scale features of atmospheric CH<sub>4</sub> over the past 800,000 years. *Nature* **453**, 383-386.
- Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: matching biological questions techniques with appropriate. *Trends in Ecology & Evolution* **20**, 136-142.
- Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**, 189-197.
- Manos PS (2001) The Historical Biogeography of Fagaceae: Tracking the Tertiary History of Temperate and Subtropical Forests of the Northern Hemisphere. *International Journal of Plant Sciences* **162**, S77-S93.
- Manos PS, Doyle JJ, Nixon KC (1999) Phylogeny, Biogeography, and Processes of Molecular Differentiation in Quercus Subgenus Quercus (Fagaceae). *Molecular Phylogenetics and Evolution* **12**, 333-349.
- Mariette S, Cottrell J, Csaikl UM, Goikoechea P, Konig A, Lowe AJ, Van Dam BC, Barreneche T, Bodenes C, Streiff R, Burg K, Groppe K, Munro RC, Tabbener H, Kremer A (2002) Comparison of levels of genetic diversity detected with AFLP and microsatellite markers within and among mixed Q-petraea (MATT.) LIEBL. and Q-robur L. stands. *Silvae Genetica* **51**, 72-79.
- Martin NH, Bouck AC, Arnold ML (2007) The genetic architecture of reproductive isolation in Louisiana irises: Flowering phenology. *Genetics* **175**, 1803-1812.
- Martin NH, Sapir Y, Arnold ML (2008) The genetic architecture of reproductive isolation in Louisiana irises: Pollination syndromes and pollinator preferences. *Evolution* **62**, 740-752.
- Masly JP, Jones CD, Noor MAF, Locke J, Orr HA (2006) Gene Transposition as a Cause of Hybrid Sterility in *Drosophila* 10.1126/science.1128721. *Science* **313**, 1448-1450.
- Matsuki Y, Isagi Y, Suyama Y (2007) The determination of multiple microsatellite genotypes and DNA sequences from a single pollen grain. *Molecular Ecology Notes* **7**, 194-198.
- Mayfield JA, Fiebig A, Johnstone SE, Preuss D (2001) Gene Families from the Arabidopsis thaliana Pollen Coat Proteome. *Science* **292**, 2482-2485.
- Mayr E (1942) *Systematics and the origin of species from the viewpoint of a zoologist*. Columbia University Press, New York.
- Mayr E (1963) *Animal species and evolution*, Cambridge, Massachusetts.
- Mayr E (1992) A Local Flora and the Biological Species Concept. *American Journal of Botany* **79**, 222-238.
- Moore WS (1977) An evaluation of narrow hybrid zones in vertebrates. *The Quarterly Review of Biology* **52**, 263-277.
- Muir G, Fleming CC, Schlotterer C (2000) Taxonomy - Species status of hybridizing oaks. *Nature* **405**, 1016-1016.
- Muir G, Lowe AJ, Fleming CC, Vogl C (2004) High nuclear genetic diversity, high levels of outcrossing and low differentiation among remnant populations of Quercus petraea at the margin of its range in Ireland. *Annals of Botany* **93**, 691-697.
- Muir G, Schlotterer 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.
- Muir G, Schlotterer C (2006) Moving beyond single-locus studies to characterize hybridization between oaks (Quercus spp.). *Molecular Ecology* **15**, 2301-2304.
- Muller CH (1952) Ecological control of hybridization in Quercus: a factor in the mechanism of evolution. *Evolution* **6**, 147-161.
- Müntzing A (1930) Outlines to a genetic monograph of the genus Galeopsis. *Hereditas* **13**, 185-341.
- Nason JD, Ellstrand NC, Arnold ML (1992a) Patterns of hybridization and introgression in populations of oaks, manzanitas and irises. *American Journal of Botany* **79**, 101-111.
- Nason JD, Ellstrand NC, Arnold ML (1992b) Patterns of Hybridization and Introgression in Populations of Oaks, Manzanitas, and Irises. *American Journal of Botany* **79**, 101-111.
- Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversities. *Ann Hum Genet* **47**.
- Nielsen EE, Hansen MM, Ruzzante DE, Meldrup D, Gronkjaer P (2003) Evidence of a hybrid-zone in Atlantic cod (Gadus morhua) in the Baltic and the Danish Belt Sea revealed by individual admixture analysis. *Molecular Ecology* **12**, 1497-1508.
- Nielsen EEG, Bach LA, Kotlicki P (2006) HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples. *Molecular Ecology Notes* **6**, 971-973.
- Niemiller ML, Fitzpatrick BM, Miller BT (2008) Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae : Gyrinophilus) inferred from gene genealogies. *Molecular Ecology* **17**, 2258-2275.
- Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology* **17**, 2103-2106.
- Olrik DC, Kjaer ED (2007) The reproductive success of a Quercus petraea x Q-robur F1-hybrid in back-crossing situations. *Annals of Forest Science* **64**, 37-45.
- Orr HA (1998) Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics* **149**, 2099-2104.
- Orr HA (2005) The genetic basis of reproductive isolation: Insights from Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 6522-6526.
- Orr HA, Masly JP, Presgraves DC (2004) Speciation genes. *Current Opinion in Genetics & Development* **14**, 675-679.

- Paabo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M (2004) Genetic analyses from ancient DNA. *Annual Review of Genetics* **38**, 645-679.
- Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite Analysis of Population-Structure in Canadian Polar Bears. *Molecular Ecology* **4**, 347-354.
- Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* **13**, 55-65.
- Paffetti D, Vettori C, Caramelli D, Vernesi C, Lari M, Paganelli A, Paule L, Giannini R (2007) Unexpected presence of *Fagus orientalis* complex in Italy as inferred from 45,000-year-old DNA pollen samples from Venice lagoon. *Bmc Evolutionary Biology* **7**.
- Pandey KK (1977) Mentor pollen: Possible role of wall-held pollen growth promoting substances in overcoming intra- and interspecific incompatibility. *Genetica* **47**, 219-229.
- Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289-290.
- Parducci L, Petit RJ (2004) Ancient DNA - unlocking plants' fossil secrets. *New Phytologist* **161**, 335-339.
- Parducci L, Suyama Y, Lascoux M, Bennett KD (2005) Ancient DNA from pollen: a genetic record of population history in Scots pine. *Molecular Ecology* **14**, 2873-2882.
- Pashley CH, Ellis JR, McCauley DE, Burke JM (2006) EST databases as a source for molecular markers: Lessons from *Helianthus*. *Journal of Heredity* **97**, 381-388.
- Pasqualotto AC, Denning DW, Anderson MJ (2007) A cautionary tale: Lack of consistency in allele sizes between two laboratories for a published multilocus microsatellite typing system. *Journal of Clinical Microbiology* **45**, 522-528.
- Pernès J (1984) *Gestion des ressources génétiques. Tome 2: Manuel*. Agence de Coopération Culturelle et Technique, Paris, France.
- Perron M, Bousquet J (1997) Natural hybridization between black spruce and red spruce. *Molecular Ecology* **6**, 725-734.
- Petit RJ, Bodenes C, Ducouso A, Roussel G, Kremer A (2003) Hybridization as a mechanism of invasion in oaks. *New Phytologist* **161**, 151-164.
- Petit RJ, Brewer S, Bordacs S, Burg K, Cheddadi R, Coart E, Cottrell J, Csaikl UM, van Dam B, Deans JD, Espinel S, Fineschi S, Finkeldey R, Glaz I, Goicoechea PG, Jensen JS, König AO, Lowe AJ, Madsen SF, Matyas G, Munro RC, Popescu F, Slade D, Tabbener H, de Vries SGM, Ziegenhagen B, de Beaulieu JL, Kremer A (2002a) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management* **156**, 49-74.
- Petit RJ, Csaikl UM, Bordacs S, Burg K, Coart E, Cottrell J, van Dam B, Deans JD, Dumolin-Lapegue S, Fineschi S, Finkeldey R, Gillies A, Glaz I, Goicoechea PG, Jensen JS, König AO, Lowe AJ, Madsen SF, Matyas G, Munro RC, Olalde M, Pemonge MH, Popescu F, Slade D, Tabbener H, Turchini D, de Vries SGM, Ziegenhagen B, Kremer A (2002b) Chloroplast DNA variation in European white oaks - Phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management* **156**, 5-26.
- Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* **12**, 844-855.
- Petit RJ, Kremer A, Wagner DB (1993) Geographic Structure of Chloroplast DNA Polymorphisms in European Oaks. *Theoretical and Applied Genetics* **87**, 122-128.
- Petit RJ, Latouche-Halle C, Pemonge MH, Kremer A (2002c) Chloroplast DNA variation of oaks in France and the influence of forest fragmentation on genetic diversity. *Forest Ecology and Management* **156**, 115-129.
- Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducouso A, Kremer A (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 9996-10001.
- Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004) GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity* **95**, 536-539.
- Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: Causes, consequences and solutions. *Nature Reviews Genetics* **6**, 847-859.
- Potts BM, Reid JB (1988) Hybridization as a Dispersal Mechanism. *Evolution* **42**, 1245-1255.
- Presgraves DC (2007) Speciation genetics: Epistasis, conflict and the origin of species. *Current Biology* **17**, R125-R127.
- Presgraves DC, Balagopalan L, Abmayr SM, Orr HA (2003) Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. **423**, 715-719.
- Presson AP, Sobel E, Lange K, Papp JC (2006) Merging microsatellite data. *Journal of Computational Biology* **13**, 1131-1147.
- Primmer CR, Koskinen MT, Piironen J (2000) The one that did not get away: individual assignment using microsatellite data detects a case of fishing competition fraud. *Proceedings of the Royal Society of London Series B-Biological Sciences* **267**, 1699-1704.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

- Pritchard JK, Wen X, Falush D (2007) Documentation for structure software: Version 2.2., p. <http://pritch.bsd.uchicago.edu/software/structure22/readme.pdf>. Department of Human Genetics, University of Chicago, Chicago, Illinois.
- Rameau JC, Mansion D, Dumé G (1994) *Flore forestière française - Tome 1, Plaines et Collines*.
- Ramsey J, Bradshaw HD, Schemske DW (2003) Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution* **57**, 1520-1534.
- Randi E (2008) Detecting hybridization between wild species and their domesticated relatives. *Molecular Ecology* **17**, 285-293.
- Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 9197-9201.
- Rattenbury JA (1962) Cyclic hybridization as a survival mechanism in the New Zealand forest flora. *Evolution* **16**, 348-363.
- Rice WR, Hostert EE (1993) Laboratory Experiments on Speciation - What Have We Learned in 40 Years. *Evolution* **47**, 1637-1653.
- Rieseberg LH (1991) Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes. *American Journal of Botany* **78**.
- Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution* **16**, 351-358.
- Rieseberg LH, Burke JM (2001) A genic view of species integration - Commentary. *Journal of Evolutionary Biology* **14**, 883-886.
- Rieseberg LH, Carney SE (1998) Plant hybridization. *New Phytologist* **140**, 599-624.
- Rieseberg LH, Carter R, Zona S (1990) Molecular tests of the hypothesized hybrid origin of two diploid *Helianthus* species (Asteraceae). *Evolution* **44**.
- Rieseberg LH, Wood TE, Baack EJ (2006) The nature of plant species. *Nature* **440**, 524-527.
- Robledo-Arnuncio JJ, Austerlitz F, Smouse PE (2006) A new method of estimating the pollen dispersal curve independently of effective density. *Genetics* **173**, 1033-1045.
- Robledo-Arnuncio JJ, Austerlitz F, Smouse PE (2007) POLDISP: a software package for indirect estimation of contemporary pollen dispersal. *Molecular Ecology Notes* **7**, 763-766.
- Rosenberg NA (2005) Algorithms for selecting informative marker panels for population assignment. *Journal of Computational Biology* **12**, 1183-1201.
- Roussel G (1999) Injecteur de pollen à deux voies et à air reconstitué. *Cahier des techniques INRA* **42**.
- Roussel G (2002) Techniques de pollinisation sur les chênes européens au laboratoire de génétique et d'amélioration des arbres forestiers de Pierroton Cestas France, p. 71. INRA Bordeaux - UMR BIOGECO 1202, Cestas.
- Ruane LG, Donohue K (2007) Environmental effects on pollen-pistil compatibility between *Phlox cuspidata* and *P. drummondii* (Polemoniaceae): Implications for hybridization dynamics. *American Journal of Botany* **94**, 219-227.
- Ruane LG, Donohue K (2008) Pollen competition and environmental effects on hybridization dynamics between *Phlox drummondii* and *Phlox cuspidata*. *Evolutionary Ecology* **22**, 229-241.
- Rungis D, Berube Y, Zhang J, Ralph S, Ritland CE, Ellis BE, Douglas C, Bohlmann J, Ritland K (2004) Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. *Theoretical and Applied Genetics* **109**, 1283-1294.
- Rushton BS (1977) Artificial hybridization between *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. *Watsonia* **11**.
- Saintagne C, Bodenes C, Barreneche T, Pot D, Plomion C, Kremer A (2004) Distribution of genomic regions differentiating oak species assessed by QTL detection. *Heredity* **92**, 20-30.
- Scareli-Santos C, Herrera-Arroyo MaL, Sanchez-Mondragan MaL, Gonzalez-Rodríguez A, Bacon J, Oyama K (2007) Comparative analysis of micromorphological characters in two distantly related Mexican oaks, *Quercus conzattii* and *Q. eduardii* (Fagaceae), and their hybrids. *Brittonia*, 37-48.
- Scotti-Saintagne C, Bodenes C, Barreneche T, Bertocchi E, Plomion C, Kremer A (2004a) Detection of quantitative trait loci controlling bud burst and height growth in *Quercus robur* L. *Theoretical and Applied Genetics* **109**, 1648-1659.
- Scotti-Saintagne C, Mariette S, Porth I, Goicoechea PG, Barreneche T, Bodenes K, Burg K, Kremer A (2004b) 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 (2006) Conservation: Losing biodiversity by reverse speciation. *Current Biology* **16**, R334-R337.
- Seehausen O, vanAlphen JJM, Witte F (1997) Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**, 1808-1811.
- Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* **9**, 615-629.
- Simpson GG (1951) The species concept. *Evolution* **5**, 285-298.
- Skalski GT, Couch CR, Garber AF, Weir BS, Sullivan CV (2006) Evaluation of DNA pooling for the estimation of microsatellite allele frequencies: A case study using striped bass (*Morone saxatilis*). *Genetics* **173**, 863-875.
- Slavov GT, Howe GT, Gyaourova AV, Birkes DS, Adams WT (2005) Estimating pollen flow using SSR markers and paternity exclusion: accounting for mistyping. *Molecular Ecology* **14**, 3109-3121.



- Slotte T, Huang H, Lascoux M, Ceplitis A (2008) Polyploid speciation did not confer instant reproductive isolation in *Capsella* (Brassicaceae). *Molecular Biology and Evolution* **25**, 1472-1481.
- Smith MW, Lautenberger JA, Shin HD, Chretien JP, Shrestha S, Gilbert DA, O'Brien SJ (2001) Markers for mapping by admixture linkage disequilibrium in African American and Hispanic populations. *American Journal of Human Genetics* **69**, 1080-1094.
- Smouse PE, Dyer RJ, Westfall RD, Sork VL (2001) Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. *Evolution* **55**, 260-271.
- Soulsbury CD, Iossa G, Edwards KJ, Baker PJ, Harris S (2007) Allelic dropout from a high-quality DNA source. *Conservation Genetics* **8**, 733-738.
- Sourdriil A (2007) *Frontières sociales, frontières spatiales. Approche anthropologique de l'exploitation et de la transmission des petites forêts privées fragmentées des Coteaux de Gascogne (Haute-Garonne)*. Thèse de l'Université de Paris X (Nanterre).
- Steffensen JP, Andersen KK, Bigler M, Clausen HB, Dahl-Jensen D, Fischer H, Goto-Azuma K, Hansson M, Johnsen SJ, Jouzel J, Masson-Delmotte V, Popp T, Rasmussen SO, Rothlisberger R, Ruth U, Stauffer B, Siggaard-Andersen M-L, Sveinbjornsdottir AE, Svensson A, White JWC (2008) High-Resolution Greenland Ice Core Data Show Abrupt Climate Change Happens in Few Years. *Science*, DOI: 10.1126/science.1157707.
- Steinhoff S (1993) Results of species hybridization with *Quercus robur* L and *Quercus petraea* (Matt) Liebl. *Annales Des Sciences Forestieres* **50**, 137s-143s.
- Steinhoff S (1998) Controlled crosses between pendunculate and sessile oak: results and conclusion. *Allgemeine Forst Und Jagdzeitung* **169**, 163-168.
- Steinkellner H, Fluch S, Turetschek E, Lexer C, Streiff R, Kremer A, Burg K, Glossl J (1997) Identification and characterization of (GA/CT)(n)-microsatellite loci from *Quercus petraea*. *Plant Molecular Biology* **33**, 1093-1096.
- Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, Dezzani R, Delmelle E, Vierling L, Waits LP (2007) Putting the 'landscape' in landscape genetics. *Heredity* **98**, 128-142.
- Strasburg JL, Rieseberg LH (2008) Molecular Demographic History of the Annual Sunflowers *Helianthus Annuus* and *H. Petiolaris*-Large Effective Population Sizes and Rates of Long-Term Gene Flow. *Evolution*.
- Streiff R (1998) *Structure génétique spatiale et flux polliniques dans une chenaie mixte : Quercus petraea (Matt.) Liebl., Quercus robur L.*, Université de Montpellier 2.
- Streiff R, Ducouso A, Lexer C, Steinkellner H, Glossl J, Kremer A (1999) Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L-and *Q-petraea* (Matt.) Liebl. *Molecular Ecology* **8**, 831-841.
- Streiff R, Labbe T, Bacilieri R, Steinkellner H, Glossl J, Kremer A (1998) Within-population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. *Molecular Ecology* **7**, 317-328.
- Streiff R, Veyrier R, Audiot P, Meusnier S, Brouat C (2005) Introgression in natural populations of bioindicators: a case study of *Carabus splendens* and *Carabus punctatoauratus*. *Molecular Ecology* **14**, 3775-3786.
- Suyama Y, Kawamuro K, Kinoshita I, Yoshimura K, Tsumura Y, Takahara H (1996) DNA sequence from a fossil pollen of *Abies* spp from Pleistocene peat. *Genes & Genetic Systems* **71**, 145-149.
- Swanson R, Edlund AF, Preuss D (2004) Species specificity in pollen-pistil interactions. *Annual Review of Genetics* **38**, 793-818.
- Swanson WJ, Vacquier VD (2002) The rapid evolution of reproductive proteins. *Nature Reviews Genetics* **3**, 137-144.
- Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7**, 453-464.
- Taylor EB, Boughman JW, Groenenboom M, Sniatynski M, Schluter D, Gow JL (2006) Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology* **15**, 343-355.
- Topchy A, Scribner K, Punch W (2004) Accuracy-driven loci selection and assignment of individuals. *Molecular Ecology Notes* **4**, 798-800.
- Tovar-Sanchez E, Oyama K (2004) Natural hybridization and hybrid zones between *Quercus crassifolia* and *Quercus crassipes* (Fagaceae) in Mexico: morphological and molecular evidence. *Am. J. Bot.* **91**, 1352-1363.
- Vaha JP, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology* **15**, 63-72.
- 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.
- Valbuena-Carabana M, Gonzalez-Martinez SC, Sork VL, Collada C, Soto A, Goicoechea PG, Gil L (2005) Gene flow and hybridisation in a mixed oak forest (*Quercus pyrenaica* Willd. and *Quercus petraea* (Matts.) Liebl.) in central Spain. *Heredity* **95**, 457-465.
- Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.
- Van Valen L (1976) Ecological species, multispecies, and oaks. *Taxon* **25**, 233-239.
- Vandeputte M, Mauger S, Dupont-Nivet M (2006) An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by exclusion. *Molecular Ecology Notes* **6**, 265-267.

- Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology* **13**, 921-935.
- Vemireddy LR, Archak S, Nagaraju J (2007) Capillary electrophoresis is essential for Microsatellite marker based detection and quantification of adulteration of basmati rice (*Oryza sativa*). *Journal of Agricultural and Food Chemistry* **55**, 8112-8117.
- Viscosi V, Fortini P, Slice DE, Loy A, Blasi C (2008a) Geometric morphometric analyses of leaf variation in four oak species of subgenus *Quercus* (Fagaceae). *Plant Biosystems* **in press**.
- Viscosi V, Lepais O, Gerber S, Fortini P (2008b) Leaf morphological analyses in four European oak species (*Quercus*) and their hybrids: a comparison of traditional and geometric morphometric methods. *Plant Biosystems* **in press**.
- Wagner WHJ (1970) Biosystematics and evolutionary noise. *Taxon* **19**, 146-151.
- Wandeler P, Hoeck PEA, Keller LF (2007) Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution* **22**, 634-642.
- Wang JL (2003) Maximum-likelihood estimation of admixture proportions from genetic data. *Genetics* **164**, 747-765.
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**.
- Williams JH, Boecklen WJ, Howard DJ (2001) Reproductive processes in two oak (*Quercus*) contact zones with different levels of hybridization. *Heredity* **87**, 680-690.
- Winge Ö (1917) The chromosomes: their number and general importance. *Comptes Rendues des Travaux du Laboratoire Carlesberg* **13**, 131-275.
- Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**, 851-865.
- Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH (2007) Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics* **175**, 1883-1893.
- Zaffarano PL, McDonald BA, Linde CC (2008) Rapid speciation following recent host shifts in the plant pathogenic fungus *Rhynchosporium*. *Evolution* **62**, 1418-1436.

# ANNEXES



## ANNEXE 1 - Résultats complets des croisements contrôlés.

Tableau : Effectifs des paternités pour chaque croisement contrôlé déterminé par 5 microsattellites. (\* : pollen non attendu lié à une inversion des pollens lors de l'injection ; souligné : croisements interspécifiques).

Mère Pollen injecté	11P					A4						3P					A3	QS30					QS32					QS11				QS27
	Me	Pe	Se	Pu	Ta	Me	Pe	Se	Pu	Ta	Me	Pe	Se	Pu	Ta	Me	Me	Se	Pe	Pu	Ta	Me	Se	Pe	Pu	Ta	Se	Pe	Pu	Ta	Me	
Père déterminé autof.	2		7													46								1								
11P						9	3	24			4	11				38																
A4			73*									9				6																
A6	3		107*			15	5	34			2	22				60																
3P	9	2	63*			7	3	30								26																
27P	89	2	70*			3	3	1			2	22				24																
QS27	<u>2</u>	<u>23*</u>	<u>30</u>			<u>3</u>	<u>1</u>	<u>156</u>					<u>40</u>			<u>10</u>	5	1					17	15*			3					
QS30			<u>1</u>										<u>0</u>										1									
QS32			<u>1</u>			<u>1</u>		<u>26</u>			<u>1</u>		<u>7</u>			<u>6</u>	10										7				2	
QS31											<u>1</u>		<u>3</u>				2						2	4*			3				1	
QS11	<u>1</u>	<u>1*</u>					<u>1</u>	<u>1</u>					<u>4</u>				1		1*				2								4	
PU1			<u>16</u>								<u>4</u>		<u>3</u>				<u>4</u>							<u>1</u>				<u>9</u>			<u>1</u>	
PU2	<u>3</u>		<u>6</u>								<u>1</u>		<u>2</u>				<u>5</u>							<u>1</u>				<u>6</u>				
PU7			<u>15</u>										<u>2</u>				<u>6</u>							<u>11</u>				<u>8</u>		<u>1</u>		
Bra55			<u>14</u>										<u>1</u>				<u>2</u>											<u>12</u>				
Bra58			<u>3</u>								<u>1</u>		<u>5</u>				<u>7</u>							<u>13</u>				<u>14</u>				
T12														<u>9</u>		<u>1</u>																
T16											<u>1</u>			<u>5</u>		<u>1</u>																
T25	<u>2</u>													<u>1</u>		<u>1</u>	1															
T26														<u>0</u>																		
T27	<u>2</u>					<u>1</u>								<u>3</u>		<u>1</u>																
total déterminé	113	28	352	54	0	39	16	89	183	0	0	17	64	54	13	18	220	43	1	1	0	0	0	22	20	26	0	13	0	49	0	9
Pollution	118	5	39	0	1	0	16	25	8	0	0	0	25	18	1	1	9	1	0	0	0	0	0	0	1	1	0	1	0	1	0	1
Non Amplifié	2	7	27	0	0	1	13	29	21	0	0	0	5	10	0	0	11	0	0	0	0	0	0	5	2	4	0	0	0	2	0	0
Total	233	40	418	54	1	40	45	143	212	0	0	17	94	82	14	19	240	44	1	1	0	0	0	27	23	31	0	14	0	52	0	10

ANNEXE 1 - Résultats complets des croisements contrôlés.

Tableau (suite)

Mère Pollen injecté	T9					T29					T30					Bra2					Bra55					Bra56				
	Me	Ta	Pe	Se	Pu	Me	Ta	Pe	Se	Pu	Me	Ta	Pe	Se	Pu	Me	Pu	Pe	Se	Ta	Me	Pu	Pe	Se	Ta	Me	Pu	Pe	Se	Ta
Père autof.													1		1									1						
11P						1		4					6	1																
A4													2																	
A6	1		3			1		4					10	1																
3P						1							2																	
27P			1					4					10	1																
QS27				7		2			4		1				9															
QS30																														
QS32				2		1			6						7															
QS31															2															
QS11				1					1																					
PU1					1	4				1					3							1			2					
PU2										1					2		1													
PU7						1					1				5	1						1			1					
Bra55					1					2					4															
Bra58	1				3	1				4					7							1								
T12	1					4	13				3																			
T16		1				2	9				2																			
T25	1					2	12				1																			
T26							1																							
T27	1					3	24					2																		
total dét.	5	1	4	10	5	23	59	12	11	8	8	2	31	21	22	1	1	0	0	0	0	3	0	1	3	0	0	0	0	0
pollution	1	2	1	0	0	3	10	0	13	2	0	2	7	1	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
na	0	4	0	1	1	5	8	2	11	2	1	0	2	3	8	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0
Total	6	7	5	11	6	31	77	14	35	12	9	4	40	25	35	1	1	0	0	0	0	3	0	1	4	0	0	1	0	0

Les individus d'espèces pures sont correctement assignés dans 97,2 à 99,6 % des cas (tableaux 1 et 2 en annexe). Les individus purs mal classés sont généralement assignés à une des classes d'hybrides de seconde génération de leur espèce (tableau 1). Il n'y a pas d'erreur d'assignation entre espèces au niveau des individus purs. Les hybrides de première génération ont des taux d'assignation correcte allant de 47 % à 67 % alors que les hybrides de seconde génération présentent des taux plus faibles entre 32 % et 60 %. Les mauvaises classifications des hybrides résultent majoritairement de la difficulté de dissocier les différentes générations d'hybrides (par exemple confondre un F1 pédonculé x sessile avec un rétrocroisement pédonculé x sessile). En effet les erreurs d'assignation entre paires d'espèces sont rares (confondre un F1 pédonculé x sessile avec un F1 pédonculé x tauzin). Les précisions (tableau 2) montrent des valeurs faibles pour les hybrides F1, ce qui s'explique par le fait que les classes F1 contiennent en grande proportion des hybrides de seconde génération (tableau 1). Néanmoins, les classes espèces pures et hybrides de seconde génération contiennent bien une majorité d'individus appartenant à ces deux classes respectives (environ 94% des classes espèces pures et environ 65% des classes d'hybrides de seconde génération). La performance globale de l'assignation est assez bonne pour les espèces pures mais ces classes contiennent entre 5 et 10 % d'hybrides de seconde génération. En outre, la performance globale est mauvaise pour les différentes classes d'hybrides. Ceci provient du fait que de nombreux F1 sont identifiés comme hybrides de seconde génération et parallèlement, un nombre non négligeable d'hybrides de seconde génération est confondu avec des F1 ou les espèces pure parentales.

Tableau 1 : Efficacité, précision et performance globale de l'assignation génétique.

Classe	Efficacité	Précision	Performance
Ped	99.60%	95.04%	94.66%
Ses	98.80%	94.36%	93.23%
Toz	99.20%	94.57%	93.81%
Pub	97.20%	93.46%	90.84%
F1_PedSes	46.67%	40.00%	18.67%
F1_PedToz	60.00%	42.86%	25.71%
F1_PedPub	66.67%	64.52%	43.01%
F1_SesToz	53.33%	42.11%	22.46%
F1_SesPub	63.33%	42.22%	26.74%
F1_ToPub	53.33%	55.17%	29.43%
bc_PedSes	51.67%	63.27%	32.69%
bc_SesPed	38.33%	58.97%	22.61%
bc_PedToz	46.67%	71.79%	33.50%
bc_ToPed	46.67%	71.79%	33.50%
bc_PedPub	60.00%	83.72%	50.23%
bc_PubPed	58.33%	74.47%	43.44%
bc_SesToz	53.33%	66.67%	35.56%
bc_ToSes	31.67%	59.38%	18.80%
bc_SesPub	35.00%	70.00%	24.50%
bc_PubSes	30.00%	45.00%	13.50%
bc_ToPub	55.00%	66.00%	36.30%
bc_PubToz	36.67%	55.00%	20.17%

## ANNEXE 2 - Identification des deux générations d'hybrides et des espèces pures

Tableau 2 : Répartition des individus simulés dans chaque classe d'assignation.

SimuléAssigné	Ped	Ses	Toz	Pub	F1 PedSes	F1 PedToz	F1 PedPub	F1 SesToz	F1 SesPub	F1 TozPub	bc PedSes	bc SesPed	bc PedToz	bc TozPed	bc PedPub	bc PubPed	bc SesToz	bc TozSes	bc SesPub	bc PubSes	bc TozPub	bc PubToz	total
Ped	996										3		1										1000
Ses		988			1							6					4		1				1000
Toz			992											3				2			3		1000
Pub				972				1	1							4				13		9	1000
F1_PedSes		3			14						7	6											30
bc_PedSes	19	1			7						31		2										60
bc_SesPed		20			13			1			2	23					1						60
F1_PedToz			1			18							6	4							1		30
bc_PedToz	15					11					2		28	1	3								60
bc_ToizPed	1		16			13								28							1	1	60
F1_PedPub				1			20								4	5							30
bc_PedPub	17						2				4		1		36								60
bc_PubPed				11			9		1			1				35						3	60
F1_SesToz				1			16										7	6					30
bc_SesToz		16					6	1				2		1			32		2				60
bc_ToizSes			23				13							2				19			3		60
F1_SesPub									19							1			5	5			30
bc_SesPub		19		1			1	14				1					3		21				60
bc_PubSes				29				9								2				18		2	60
F1_ToizPub										16							1			2	8	3	30
bc_ToizPub			16				1			4			1					3	1		33	1	60
bc_PubToz			1	25						8								1		2	1	22	60
total	1048	1047	1049	1040	35	42	31	38	45	29	49	39	39	39	43	47	48	32	30	40	50	40	4900

## Short Note: High Throughput Microsatellite Genotyping in Oak Species

By O. LEPAIS, V. LÉGER and S. GERBER\*

UMR BIOGECO, INRA-Université Bordeaux I, 69 route d'Arcachon, 33612 Cestas Cedex, France

(Received 21<sup>st</sup> March 2006)

### Abstract

Microsatellites are widely used markers for multiple purposes in oaks. We describe a complete procedure for cheap DNA extraction and fast microsatellites genotyping by multiplex PCR. 10 loci were selected to form two multiplex kits including three loci that show a high differentiation between *Quercus robur* and *Q. petraea*. The loci were tested in three oak species and show a high mean genetic diversity of 0.84. The cumulative exclusion probability for parentage analysis was 0.999977 for single parent and 1.0 for paternity. Finally, the relatively high differentiation coefficient ( $G_{st} = 0.04$ ) will facilitate species assignment based on genotypes in oaks.

**Key words:** microsatellites; multiplex PCR; *Quercus*; genetic assignment; parentage analysis.

Oaks are common species in a large part of the world and are involved in an increasing number of genetic studies concerning especially population genetic structure and gene flow. New powerful methods for analysing multilocus genetics data allow inference about past demographic events, genetics assignment or detection of selection (CHIKHI and BRUFORD, 2005). We took advantage of technical laboratory progresses to increase availability of multilocus microsatellites data. Combining multiplex PCR and automated capillary sequencer speed up the genotyping compared to silver staining gel or gel plate automated sequencer. In the same time it decreases cost of genotyping and allows to reach high loci number analysis which is needed for high resolution studies in natural populations. In this short note, we present a protocol that allows fast DNA extraction and genotyping for 10 microsatellites in two multiplex PCR.

We used a modified protocol based on CTAB/dichloromethane (DUMOLIN et al., 1995) allowing DNA extraction in 96 wells plates format. This modification provides high quality DNA extraction from 192 individuals in one day at a low cost compared to commercial kits.

One centimeter square of leave or three to five buds are cut into small pieces and put in a 1.2 mL microtube (Qiagen, Cat. No. 19560) with two 2 mm steel balls, a small amount of insoluble PVP and alumina, 100  $\mu$ L of extraction buffer (CTAB 20 mg.mL<sup>-1</sup>, EDTA (pH 8) 0.02M, Tris HCl (pH 8) 0.1M, NaCl 1.4M, soluble PVP 10 mg.mL<sup>-1</sup>). Vegetable tissues and 300  $\mu$ L of extraction buffer are ground in a Retsch-Mill at 30 vibrations per second for 1 min for each side. The microtubes are incubated at 55°C for 1 hour in a vertical position with shaking. After cooling, 300  $\mu$ L of dichloromethane is added and microtubes are shaken. After 20 minutes of centrifugation at 6200 rpm and 4°C, the upperphase is transferred in new microtubes, 300  $\mu$ L of cold isopropanol is added and after shaking the microtubes are placed at -20°C for 1 hour in a vertical position. After a centrifugation step of 20 minutes at 6200 rpm and 4°C, the supernatant is removed, 500  $\mu$ L of ethanol is added and microtubes are shaken before centrifugation for 20 minutes at 6200 rpm and 4°C. Supernatant is then removed and the pellet is dried in a vacuum system concentrator for 15 minutes at 55°C. Finally the pellet is resuspended in 50  $\mu$ L of pure water. To avoid lost of DNA pellet when removing the upperphase after each centrifugation step, microtubes are kept on ice.

We selected microsatellites loci that are highly differentiated between *Quercus robur* and *Q. petraea* (SCOTTI-SANTAGNE et al., 2004; P. Goikoetxea, unpublished data). Before multiplexing, primers were checked for primer-primer interaction using the AutoDimer software (VALLONE and BUTLER, 2004). One set, called KIT1, comprises the five following loci QrZAG11, QrZAG39, QrZAG96, QrZAG112 and QpZAG110 and the second one, five others (KIT2): QrZAG5, QrZAG7, QrZAG20, QrZAG65 and QrZAG87 (KAMPFER et al., 1997; STEINKELLNER et al., 1997). On the 10 selected microsatellites, 7 are unlinked (Table 1). Each forward primer was marked with a fluorescent dye (MWG and Applied Biosystems) as shown in Table 1. PCR were carried out in a final volume of 15  $\mu$ L including 5  $\mu$ L of 1/50 to 1/200 diluted DNA, depending on DNA concentration

\*) Communicating author: SOPHIE GERBER. Telephone: 0033-557122830. Fax: 0033-557122881. Email: [gerber@pierroton.inra.fr](mailto:gerber@pierroton.inra.fr)

Table 1. – Characteristics of the ten microsatellites for a sample of 90 oaks (30 trees by species), Aurignac, France.

Locus	Primer Sequence	LG	[Primer] ( $\mu$ M)	Dye	Allele Number	Size Range (pb)	<i>Q. robur</i>			<i>Q. petraea</i>			<i>Q. pubescens</i>			EP(sp)	EP(p)	EP(pp)	Gst
							Ho	He	Pval	Ho	He	Pval	Ho	He	Pval				
KIT1 QrZAG11	CCTTGAACTCGAAGGTGTCCCTT	10	0.350	HEX	16	242-286	0.75	0.88	0.285	0.79	0.84	0.095	0.74	0.76	0.562	0.67	0.80	0.94	0.082
	GTAGGTCAAAAACCATTTGGTTGACT																		
QrZAG39	CACCGCTGGAATTTTAAAGGGA	5	0.025	NED	17	103-139	0.86	0.85	0.728	0.67	0.92	0.003	0.83	0.93	0.309	0.71	0.83	0.95	0.015
	GACCTAAGCCAAAGTGTGGGC																		
QrZAG96	CCAGTCACATCCACTACTGTCC	10	0.120	HEX	19	137-179	0.37	0.37	0.377	0.83	0.85	0.100	0.81	0.91	0.536	0.56	0.72	0.90	0.157
	GGTTGGGAAAAGGAGATCAGA																		
QrZAG112	TCTTGTCTTTGGTGCCCG	12	0.040	FAM	15	72-106	0.74	0.66	0.931	0.47	0.61	0.005	0.81	0.85	0.216	0.36	0.55	0.77	0.042
	GTGGTCAGAGACTCGSTAAGTATTC																		
QpZAG110	GGAGGCTTCCTTCAACCTACT	8	0.300	FAM	18	193-235	0.71	0.66	0.807	0.83	0.85	0.348	0.92	0.86	0.976	0.49	0.67	0.86	0.017
	GATCTCTTGTGTGCTGTATTT																		
KIT1 all loci			1.670		85		0.68	0.68		0.70	0.81		0.82	0.86		0.986	0.998	0.999	0.063
KIT2 QrZAG5b	TGAAGAGTAAGACCATTCACATCA	6	0.400	HEX	22	217-263	1.00	0.91	0.997	0.89	0.92	0.574	0.87	0.94	0.075	0.76	0.87	0.97	0.018
	GTATGTGAGTGTGTTGTGGTTGG																		
QrZAG7	CAACTTGGTGTTCGGATCAA	2	0.120	FAM	21	109-152	0.92	0.91	0.364	0.79	0.94	0.064	0.93	0.92	0.748	0.77	0.87	0.97	0.018
	GTGCATTTCTTTTATAGCATTAC																		
QrZAG20	CCATTAAGAAGCAGTATTTTGT	1	0.150	HEX	17	155-195	0.78	0.78	0.472	0.90	0.88	0.807	0.83	0.90	0.509	0.56	0.72	0.89	0.010
	GCAACACTCAGCCTATATCTAGAA																		
QrZAG65	CAGTGGTGTCAACTCCCTCCAG	10	0.500	FAM	27	249-306	0.79	0.94	0.030	0.75	0.96	0.001	0.71	0.94	0.000	0.81	0.89	0.98	0.001
	GTCAGGTGACCATTCAAACCTAGAA																		
QrZAG87	TOCCACCCTTTGGTCTCTCA	2	0.025	NED	18	101-141	0.81	0.85	0.231	0.73	0.88	0.053	0.78	0.81	0.506	0.64	0.78	0.93	0.051
	GTTGTGAGCAGTGGGATGGGTA																		
KIT2 all loci			2.390		105		0.86	0.88		0.81	0.92		0.82	0.90		0.998	0.999	1.000	0.019

Legend – LG: linkage group, Ho: observed heterozygosity, He: expected heterozygosity, EP(sp): single parent exclusion probability, EP(p): paternity exclusion probability, EP(pp): parent pair exclusion probability, Gst: differentiation coefficient.

and quality. For routine analysis, we advise to test amplification with several DNA dilution to find the one producing a complete amplification in all individuals. PCR reaction buffer contains 1.2X buffer (Invitrogen), 2mM MgCl<sub>2</sub>, 0.2mM dNTPs, 0.025 unit of *Taq* polymerase (Invitrogen) for one individual and each primer at different concentrations (Table 1) to increase homogeneity of amplification intensity. Even if primer concentrations produced a balanced signal with several DNA extractions series, it can be empirically adjusted to homogenize loci amplification by increasing the primer concentration of weaker loci and at the same time decreasing the primer concentration of stronger loci. The amplification is performed with a MJResearch DNA Engine Tetrad2 (Biorad) thermocycler in ABgene PCR plates (AB-800) covered with adhesive sealing sheets (AB-0558) to avoid evaporation during PCR. The cycling condition consists of a denaturation step of 3 min at 94°C, following by 30 cycles of 30 s at 94°C, 30s at 50°C, 90 s at 65°C and a final extension of 15 min at 65°C. We found that elongation at 65°C reduces stutter phenomena and at the same time increases PCR efficiency during multiplex PCR of microsatellites (HENEGARIU et al., 1997).

After amplification, 8  $\mu$ L of PCR products were mixed with 3  $\mu$ L of bromophenol blue and migrated on 3% agarose gels for 20 minutes at 100V to check for a correct amplification of loci. PCR products were diluted six-fold in pure water and 2 $\mu$ L were combined with 10  $\mu$ L of diluted size standards ET-400Rox (Amersham Bio-

sciences). Separation of fragments and alleles detection was carried out on a MegaBACE 1000 96 capillaries automated sequencer (Amersham Biosciences). Injection parameters were 60 s at 3 kV and the run was performed during 75 min at 10 kV and a constant temperature of 45°C. Data were then scored with the Fragment Profiler v1.2 software (Amersham Biosciences).

The genotyping procedure was tested in several oak species. It gave consistent results for *Q. robur*, *Q. petraea*, *Q. pubescens*, *Q. pyrenaica* (European white oaks species) and also for *Q. laurina/Q. affinis* (Mexican red oaks) and *Q. resinosa* (Mexican white oaks). In *Q. suber* and *Q. ilex*, only few loci were successfully amplified. This shows the transferability limit of microsatellites due to divergence of these species belonging respectively to *Cerris* and *Ilex* groups. Loci and multiplex PCR are therefore easily transferable in other white and red oak species.

Both microsatellites multiplex have been tested in 90 oaks of three species: *Q. robur*, *Q. petraea* and *Q. pubescens* from Aurignac (Haute Garonne, France). Allele number, observed and expected heterozygosities and P value for Hardy-Weinberg exact test (Genepop, RAYMOND and ROUSSET, 1995) are given in Table 1. Genetic diversity is high with a mean of 0.78 for KIT1 and 0.90 for KIT2. QrZAG65 shows significant heterozygote deficit for all species and some other loci show a significant heterozygote deficit or excess in one of the species.

To test for potential null alleles we genotyped one pedigree (30 offsprings) obtained by controlled crosses of



two trees (data not shown). All loci followed Mendelian inheritance and none of them, including QrZAG65, showed missing allele. So, we believe that QrZAG65 results indicate that panmixia is not expected in a sample of such a limited size because of the regional sampling and hybridization between species.

The 10 loci could be used for species assignment because they differentiate well between *Q. robur* and *Q. petraea* ( $G_{st} = 0.028$ ) but also *Q. petraea* and *Q. pubescens* ( $G_{st} = 0.019$ ) and *Q. robur* and *Q. pubescens* ( $G_{st} = 0.045$ ). Moreover, KIT1 contains QrZAG96, QrZAG112 and QrZAG11, the three highest differentiating microsatellites between *Q. robur* and *Q. petraea* (SCOTTI-SAINTEAGNE et al., 2004). These loci are located close to QTLs involve in leaf morphology that differentiate the two species (SAINTEAGNE et al., 2004).

To measure the efficiency of the loci for parentage analysis, we computed exclusion probabilities for single parent (EP(sp)), paternity (EP(p): mother known) and parent pair (EP(pp)) using FaMoz (GERBER et al., 2003). The two sets of microsatellites show high exclusion probabilities which could be slightly overestimated due to partial linkage between some of the loci (Table 1). However, considering such high probabilities, these loci are highly advisable for contemporary gene flow analyses. KIT2 is composed of 0 loci showing high allelic richness and is routinely used for controlled crosses checking in the lab. The DNA extraction procedure and the multiplex PCR described here allow a fast and convenient genotyping for 10 microsatellites in oaks for a wide variety of applications.

#### Acknowledgements

We thank PATRICK LÉGER and PIERRE-YVES DUMOLIN for the technical support, ERWAN GUICHOUX for helpful comments on protocols, ANTONIO GONZALEZ RODRIGUEZ for Mexican oaks DNA; JOHAN FOGELQVIST, SUSANNE GUSTAFSSON and MARTIN LASCoux for invitation at the Evolutionary Functional Genomics Laboratory of Uppsala University under OAKFLOW financial support. This work was founded by the Conseil Régional d'Aquitaine (Convention 20030304002FA), FEDER (Convention n°2003227) and the "Bureau des Ressources Génétiques" ("Impact de

l'écologie, de l'histoire et du mode de gestion des populations sur la structure de la diversité en forêts fragmentées" project, 2003-2004 call).

#### Literature

- CHIKHI, L. and M. BRUFORD (2005): Mammalian population genetics and genomics, pp. 539–584. In: *Mammalian Genomics*, edited by A. RUVINSKY and J. MARSHALL Graves, CAB International.
- DUMOLIN, S., B. DEMESURE and R. J. PETIT (1995): Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theoretical and Applied Genetics* **91**: 1253–1256.
- GERBER, S., P. CHABRIER and A. KREMER (2003): FaMoz: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. *Molecular Ecology Notes* **3**: 479–481.
- HENEGARIU, O., N. A. HEEREMA, S. R. DLOUHY, G. H. VANCE and P. H. VOGT (1997): Multiplex PCR: critical parameters and step-by-step protocol. *Biotechniques* **23**: 504–511.
- KAMPFER, S., C. LEXER, J. GLÖSSL and H. STEINKELLNER (1998): Characterization of (GA)<sub>n</sub> microsatellite loci from *Quercus robur*. *Hereditas* **129**: 183–186.
- RAYMOND, M. and F. ROUSSET (1995): GENEPOP version 1.2. population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**: 248–249.
- SAINTEAGNE, C., C. BODÈNES, T. BARRENECHE, D. POT, C. PLOMION and A. KREMER (2004): Distribution of genomic regions differentiating oak species assessed by QTL detection. *Heredity* **92**: 30–30.
- SCOTTI-SAINTEAGNE, C., S. MARIETTE, I. PORTH, P. G. GOICOECHEA, T. BARRENECHE, C. BODÈNES, K. BURG and A. KREMER (2004): Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics* **168**: 1615–1626.
- STEINKELLNER, H., S. FLUCH, E. TURETSCHKE, C. LEXER, R. STREIFF, A. KREMER, K. BURG and J. GLÖSSL (1997): Identification and characterization of (GA/CT)<sub>n</sub> – microsatellite locus from *Quercus petraea*. *Plant Molecular Biology* **33**: 1093–1096.
- VALLONE, P. M. and J. M. BUTLER (2004): Autodimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques* **37**(2): 226–231.

## Diversité et structuration génétique des chênes à l'échelle de deux paysages : impact de l'écologie, de l'histoire et de la gestion

Olivier LEPAIS<sup>(1)</sup>, Jessica LAVABRE<sup>(1)</sup>, Maya GONZALEZ<sup>(2)</sup>,  
Jérôme WILLM<sup>(2)</sup>, Alain CABANETTES<sup>(2)</sup>, Sophie GERBER<sup>(1)\*</sup>

<sup>(1)</sup> UMR Biogeco, INRA-Université Bordeaux I, 69 route d'Arcachon,  
33612 Cestas Cedex, France

<sup>(2)</sup> UMR Dynafor INRA-INPT/ENSAT, chemin de Borde Rouge, BP 27,  
31326 Castanet Tolosan cedex, France

**Abstract:** Genetic diversity and genetic structure of oaks in two landscapes: impact of ecology, history and management. Genetic diversity is a prerequisite for species adaptation to a changing world. We studied chloroplast and nuclear genetic diversity of oaks in two fragmented woodlands in south-western France. The chloroplast DNA polymorphism shows that oaks have a common population history, that is an Iberian postglacial maternal origin. However haplotype structures are totally different between the two studied areas due to contrasted landscape dynamics. Intense haplotype sharing between species in the Auzignac area shows that unidirectional hybridization and introgression may have lead to migration by pollen swap for at least pubescent oak. The comparison of nuclear diversity between two stands, one with an even-aged management in a continuous habitat and the other with a coppice with standard management in a fragmented landscape, shows that management does not affect nuclear diversity. The overlapping generations and the frequent partial harvest in the case of coppice with standard can explain the maintenance of high diversity level. We also show that isolation by distance have no effect on the genetic diversity of the studied oaks, characterized by long generation time and high pollen flow. The ecological background seems to have a great effect on interspecific diversity in the Auzignac region, where four species of oaks are sometimes co-occurring in the same stand. Interspecific gene flow likely increases intraspecific genetic diversity, as must be the case for pubescent oaks.

genetic diversity/ oak/ forest management/ fragmentation/ landscape

**Résumé :** Nous avons étudié la diversité génétique chloroplastique et nucléaire des chênes appartenant à deux paysages forestiers fragmentés du sud-ouest de la France. L'histoire de ces populations a été retracée grâce au polymorphisme de l'ADN chloroplastique. Alors que ce marqueur démontre que les chênes ont une origine maternelle ibérique, sa structuration géographique contrastée entre les deux sites

---

\* Correspondance et tirés à part : gerber@pierroton.inra.fr



témoigne de dynamiques paysagères différentes. Les diversités nucléaires des chênes d'une forêt gérée en futaie régulière en peuplement continu et de ceux d'une forêt gérée en taillis sous futaie en peuplement fragmenté sont équivalentes, le mode de gestion n'a pas ici de conséquence détectable sur la diversité intraspécifique. Le contexte écologique semble jouer un rôle prédominant dans la diversité interspécifique ce qui pourrait influencer, par l'hybridation, la diversité intraspécifique. Le chêne pubescent semble en effet avoir recolonisé les milieux par introgression récurrente, ce qui expliquerait l'importante diversité nucléaire de cette espèce.

diversité génétique/ chêne/ gestion forestière/ fragmentation/ paysage

## 1. INTRODUCTION

La diversité génétique est au centre des préoccupations actuelles concernant l'avenir des forêts car on prévoit pour les espèces qui les composent des pressions de sélection liées au changement climatique d'une rapidité inédite. La migration et l'adaptation sont deux stratégies permettant à des espèces longévives telles que les chênes de faire face à ces changements. Chez les chênes, les forêts étudiées concernent principalement des peuplements gérés en futaie régulière [21] dans des massifs continus. Dans le sud-ouest de la France, les populations de chênes sont souvent représentées par des petites forêts fragmentées dans le paysage. Cette fragmentation peut induire un isolement géographique réduisant les flux de gènes entre parcelles pouvant entraîner une baisse de diversité dans les populations et réduire leur capacité à s'adapter aux changements environnementaux. En outre, ces forêts régénérées naturellement sont gérées pour la plupart en taillis ou taillis sous futaie, qui associe à la futaie (reproduite par graine) une reproduction végétative par rejet de souche (le taillis). Cette gestion particulière peut avoir des conséquences sur la composition des peuplements et sur la structure génétique des populations. L'objectif de cette étude est d'observer l'effet de la fragmentation et de la gestion irrégulière sur la diversité et la structuration génétique des peuplements de chênes. Nous avons étudié deux sites composés de peuplements irréguliers comprenant plusieurs espèces de chênes. Nous avons utilisé des marqueurs chloroplastiques pour retracer l'histoire des peuplements étudiés ainsi que des marqueurs microsatellites nucléaires pour estimer la diversité des espèces. Les résultats obtenus sont comparés à ceux observés en peuplement régulier situé dans un massif continu [21].

## 2. MATÉRIEL ET MÉTHODES

### 2.1. Zones d'études

#### 2.1.1. Coteaux de Gascogne

La zone d'étude est localisée dans le bassin versant de la Nère, dans le canton d'Aurignac. Elle s'étend sur 8 000 ha et comprend 12 % de surface boisée sous forme de forêts communales et de nombreux bosquets gérés par des agriculteurs. Ces surfaces boisées sont insérées dans un paysage agricole de cultures et de prairies. Trois espèces de chênes sont représentées : chêne pédonculé (*Quercus robur*), chêne sessile (*Q. petraea*) et chêne pubescent (*Q. pubescens*). Ces espèces se répartissent dans les formations boisées suivant leurs préférences écologiques et la maturité des peuplements. Les chênes pédonculés préfèrent les sols profonds, riches et frais, les chênes pubescents se retrouvent préférentiellement sur sol calcaire et sec, majoritairement en lisière ou dans les milieux jeunes alors que les chênes sessiles ont une affinité pour les sols à tendance acide et se retrouvent au milieu de massifs plutôt âgés en accord avec leur caractère plus forestier. Trois échantillonnages à différentes échelles ont été effectués. Au niveau régional, 66 arbres ont été échantillonnés à raison de 3 chênes tous les 10 km dans un rayon de 30 km autour du canton d'Aurignac. A l'échelle du canton, 250 chênes ont été récoltés à raison de 3 arbres par population tous les 500 m. Enfin, dans un seul bois (la Paguères), 220 chênes ont été échantillonnés.

#### 2.1.2. Landes de Gascogne

La zone d'étude est localisée dans le bassin versant du Tagon, dans la partie nord du Parc Naturel Régional des Landes de Gascogne. Elle s'étend sur 6 000 ha et se compose principalement de plantations de pins maritimes (*Pinus pinaster*) d'âges variés. On trouve cependant des forêts de feuillus ou des forêts mixtes dans la partie ouest le long du bassin d'Arcachon ainsi que le long des cours d'eau où ils forment des ripisylves. Certains chênes se retrouvent aussi disséminés le long des plantations de pins et des pistes forestières ainsi que dans les espaces en déprise. Deux espèces de chênes sont présentes. Le chêne tauzin (*Q. pyrenaica*) est typique des sols acides et sa répartition est limitée dans la zone par une trop forte humidité du sol. Le chêne pédonculé (*Q. robur*) se retrouve partout et constitue l'espèce typique des ripisylves. L'échantillonnage comprend 250 chênes prélevés le long d'un maillage régulier de 500 m. A chaque point de prélèvement un chêne par espèce présente a été échantillonné.

O. Lepais et al.

## 2.2. Marqueurs moléculaires chloroplastiques et nucléaires

Le génome chloroplastique transmis maternellement et de façon clonale chez les chênes [14] permet d'étudier la dispersion des graines. Une technique de PCR-RFLP [16] a été utilisée pour étudier le polymorphisme de 4 régions intergéniques du génome chloroplastique et déterminer l'haplotype de chaque chêne.

Parmi la centaine de marqueurs microsatellites nucléaires disponibles chez les chênes, 10 ont été sélectionnés sur la base de leur différenciation interspécifique (projet européen OAKFLOW) et de leur nombre d'allèles pour génotyper les 220 arbres du bois de la Paguères et environ 30 chênes de chaque espèce de l'échantillonnage régional d'Aurignac : QpZAG110, QrZAG11, QrZAG112, QrZAG39, QrZAG96, QpZAG108, QrZAG20, QrZAG5, QrZAG65 et QrZAG87 [7], [20]. L'amplification PCR a été réalisée en combinant cinq paires d'amorces marquées correspondant à cinq locus microsatellites [11]. Ces locus ont été utilisés pour compléter le génotypage des chênes de la forêt de la Petite Charnie. L'analyse des fragments ainsi amplifiés s'est faite avec un séquenceur 96 capillaires.

## 2.3. Analyses des données chloroplastiques et nucléaires

Pour chaque zone, une carte géographique des haplotypes chloroplastiques est construite pour chaque espèce avec le logiciel ARCVIEW. Au Tagon, la structuration spatiale des haplotypes est testée grâce à l'indice I de Moran calculé pour différentes classes de distance à l'aide du logiciel SPAGEDI [6]. La zone est composée de cinq types de milieux : milieu ouvert herbacé, milieu ouvert arbustif, bois de feuillus ou bois mixte, plantation de pin et environnement urbain. Ces habitats diffèrent en terme d'âge, de stabilité dans le temps et d'influence anthropique. Nous avons comparé les compositions haplotypiques de ces milieux afin de comprendre si la structure géographique des haplotypes est liée aux caractéristiques des différents milieux. Les coordonnées spatiales des chênes pédonculés ainsi que les données haplotypiques et des milieux ont été utilisées pour construire trois matrices de distances. Les corrélations entre les différentes matrices prises deux à deux ont été analysées par des tests de Mantel. La corrélation entre haplotypes et milieux avec contrôle des distances spatiales a été étudiée avec un test de Mantel partiel [9].

Des indices de diversité des marqueurs nucléaires (hétérozygotie He sans biais [11]) sont calculés grâce au logiciel GENETIX [2] pour chaque espèce et comparés selon les dispositifs. La structuration spatiale des données microsatellites est testée en calculant les coefficients de différenciation Fst [22] pour différentes classes de distance à l'aide du logiciel SPAGEDI [6].

## Diversité génétique des chênes

Tableau I : Fréquence des haplotypes pour chaque espèce dans les deux sites étudiés.

	N	f haplo 1	f haplo 7	f haplo 10a	f haplo 10b	f haplo 11	f haplo 12
Arrignac	<i>Q. robur</i>	136	0,01	0,97	0,01	-	-
	<i>Q. petraea</i>	88	-	0,93	0,06	-	-
	<i>Q. pubescens</i>	95	0,04	0,88	0,03	0,01	-
	<b>Total</b>	<b>319</b>	<b>0,02</b>	<b>0,93</b>	<b>0,03</b>	<b>0,00</b>	<b>-</b>
Tagon	<i>Q. robur</i>	177	-	0,21	0,03	0,66	0,10
	<i>Q. pyrenaica</i>	81	-	0,07	0,02	0,85	0,05
	<b>Total</b>	<b>258</b>	<b>-</b>	<b>0,17</b>	<b>0,03</b>	<b>0,72</b>	<b>0,09</b>

O. Lepais *et al.*

### 3. RÉSULTATS

#### 3.1. Diversité chloroplastique

##### 3.1.1. Cartes des haplotypes

À Auignac, au niveau régional, l'haplotype 10a est le plus fréquent (tabl. I). Les trois espèces ont trois haplotypes en commun (haplotypes 10a, 10b et 7) alors que les haplotypes 1 et 11 sont spécifiques des chênes pubescents. Les haplotypes trouvés représentent tous les refuges glaciaires connus pour le chêne : haplotype 1 (péninsule italienne), haplotype 7 (Balkans), haplotypes 10a, 10b et 11 (péninsule ibérique) [16]. La partie nord-est de la zone est plus diversifiée que la partie sud-ouest, fixée pour l'haplotype 10a : un seul chêne pubescent parmi les 253 chênes analysés dans le canton ne possède pas cet haplotype (fig. 1). Cet arbre a été échantillonné dans le bois de la Paguères distant de 30 km de la plus proche population possédant le même haplotype.

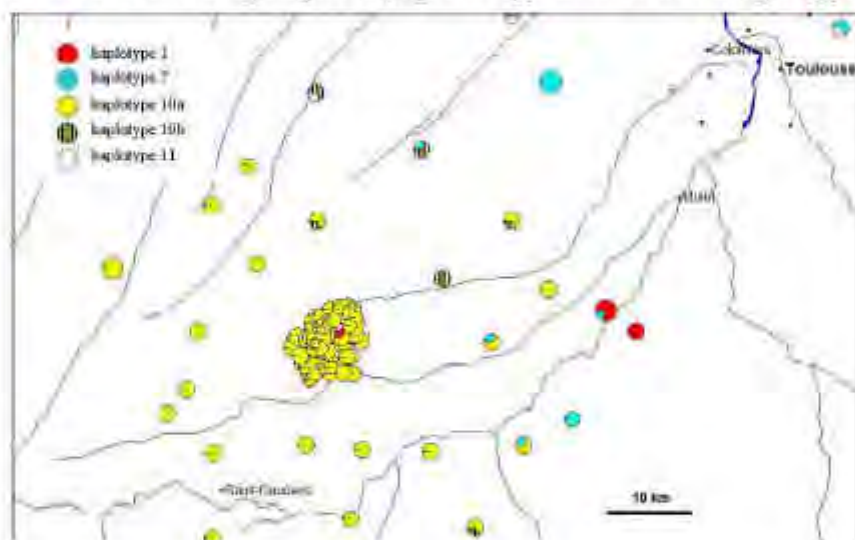


Figure 1 : Carte des haplotypes dans la région d'Auignac toutes espèces confondues. Chaque population représentée par un cercle correspond à au moins trois chênes.

Au Tagon, quatre haplotypes présents chez les deux espèces ont été identifiés parmi les 258 chênes échantillonnés dans la zone (fig. 2). Ces haplotypes, très proches génétiquement, sont d'origine ibérique [16]. L'haplotype 11 est le plus fréquent, en particulier chez les chênes tauzins (tabl. I). Les deux espèces diffèrent pour la configuration spatiale des haplotypes, contrairement à ce qui est observé à Auignac.

## Diversité génétique des chênes

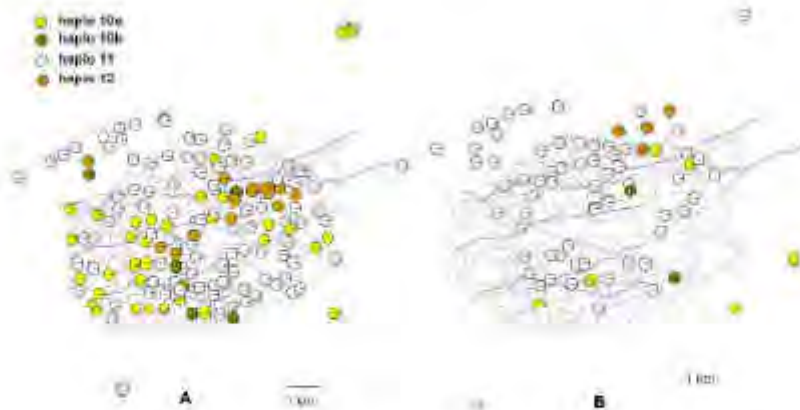


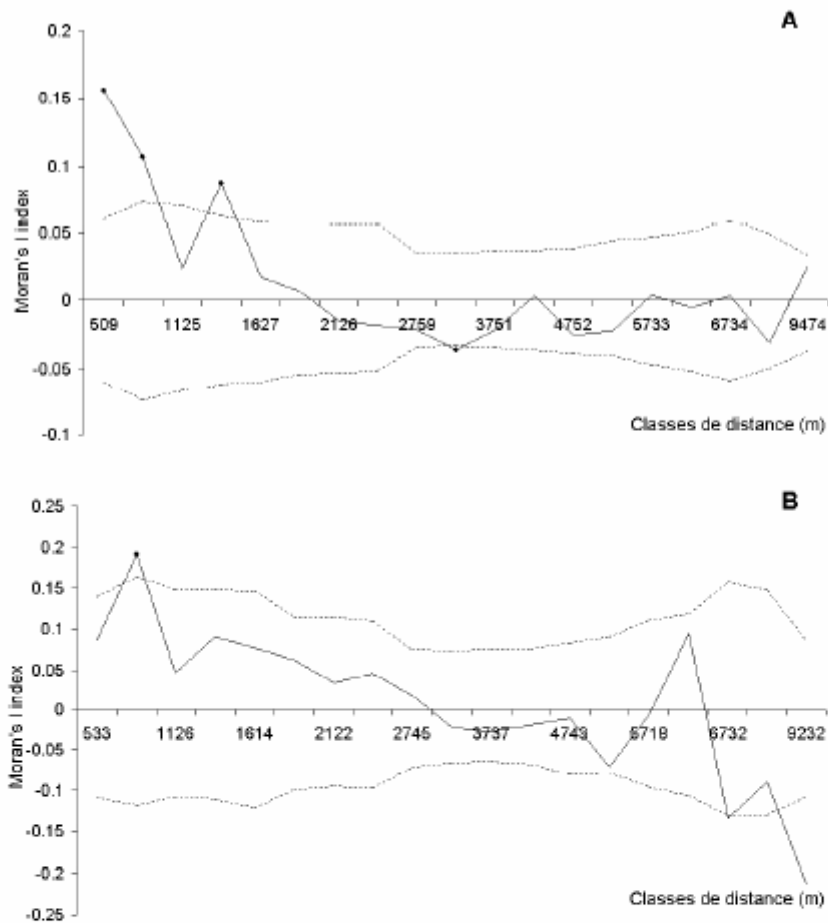
Figure 2 : Carte des haplotypes au Tagon pour les chênes pédonculés (à gauche) et les chênes tauzins (à droite). Chaque cercle correspond à un chêne.

### 3.1.2. Structuration spatiale des haplotypes

À Aunagnac, la structuration haplotypique apparaît clairement sur la carte : l'haplotype 10a est fixé dans une zone géographique s'étalant sur au moins 40 km, indépendamment de l'espèce considérée (fig. 1). Au Tagon, les chênes pédonculés montrent une forte structure spatiale pour de faibles classes de distances, jusqu'à 1 000 m, significative pour chaque haplotype, regroupé en tache. Les chênes tauzins sont plus faiblement structurés (fig. 3).

### 3.1.3. Corrélation entre haplotypes et milieux au Tagon

Les résultats des tests de Mantel mettent en évidence une corrélation forte entre matrice des milieux et matrice des distances géographiques ( $r = 0,087$ ,  $P < 0,001$ ) ainsi qu'entre matrice des haplotypes et matrice des milieux ( $r = 0,062$ ,  $P < 0,01$ ). Un test de Mantel partiel, qui contrôle l'effet des distances géographiques, montre une corrélation entre haplotypes et milieux significative ( $r = 0,061$ ,  $P = 0,01$ ). Cette corrélation est due à trois interactions. Les haplotypes 10a et 11 sont corrélés au milieu urbain (respectivement,  $r = 0,23$ ,  $P = 0,001$  et  $r = 0,01$ ,  $P = 0,01$ ) et l'haplotype 12 est lié au milieu ouvert arbustif ( $r = 0,16$ ,  $P = 0,001$ ).

O. Lepais *et al.*

**Figure 3** : Indice I de Moran calculé avec les données haplotypiques du Tagon pour les chênes pédonculés (A) et les chênes tauzins (B) pour différentes classes de distance. Les lignes pointillées montrent l'enveloppe à 95 % des valeurs de I sous l'hypothèse nulle d'absence de structuration spatiale obtenue après 1 000 permutations des données haplotypiques.

### 3.2. Diversité nucléaire

#### 3.2.1. Comparaison entre espèces

Au niveau de la région d'Auignac, les chênes renferment une diversité génétique élevée ( $He = 0,88 \pm 0,07$ ). Les chênes pubescents sont significativement plus diversifiés que les chênes pédonculés (respectivement,  $He = 0,88 \pm 0,06$  et  $He = 0,78 \pm 0,17$ ; test F :  $p < 0,01$ ) alors que les chênes sessiles présentent une diversité intermédiaire, non significativement

## Diversité génétique des chênes

différente des deux autres espèces ( $He = 0,86 \pm 0,010$  ; respectivement  $p = 0,17$  et  $p = 0,10$ ). A l'échelle de la parcelle de la Paguères, la diversité génétique des espèces est comparable à celle au niveau régional ( $0,81 \pm 0,15$ ,  $0,88 \pm 0,06$  et  $0,82 \pm 0,09$  respectivement pour les chênes pédonculé, pubescent et sessile) et le chêne pubescent présente un niveau de diversité significativement supérieur au chêne pédonculé (test F :  $p < 0,01$ ) alors que le chêne sessile a une diversité intermédiaire.

## 3.2.2. Comparaison entre types de gestion

La Paguères est une forêt du canton d'Aurignac gérée en taillis sous futaie, isolée d'au moins 400 m des autres bois, alors que la Petite Charnie est une parcelle gérée en futaie régulière qui appartient à un massif forestier continu de la Sarthe. La diversité de ces deux peuplements est comparée pour les dix même marqueurs microsatellites. Les chênes pédonculés des deux forêts présentent les mêmes niveaux de diversité ( $He = 0,81 \pm 0,15$  et  $He = 0,83 \pm 0,16$  ; test F,  $p = 0,5$ ) ainsi qu'un nombre moyen d'allèles par locus et d'allèles efficaces équivalents. Le faible effectif en chênes sessiles à la Paguères ne permet pas de comparer la diversité génétique de cette espèce. Le déficit d'hétérozygotes est significatif chez toutes les espèces dans tous les sites étudiés excepté chez les chênes pédonculés d'Aurignac. Si les chênes pédonculés de la Paguères de type franc-pied (issus de graine) et cépées (issus de rejet de souche) sont distingués, l'indice de fixation pour ces deux sous-populations devient non significatif : ces deux groupes appartiennent à des générations différentes et leur mélange conduit à un déficit apparent en hétérozygotes.

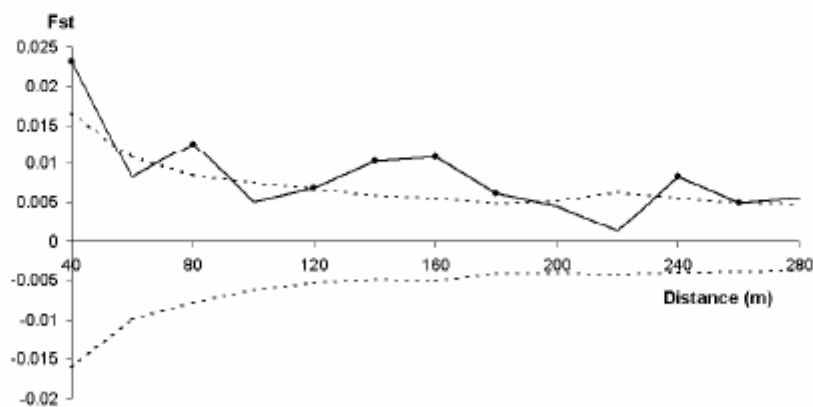


Figure 4 : Variation de  $F_{st}$  calculé à partir des génotypes de dix microsatellites pour différentes classes de distance chez les chênes pédonculés de la Paguères. Les lignes pointillées montrent l'enveloppe à 95 % des valeurs de  $F_{st}$  sous l'hypothèse nulle d'absence de structuration spatiale obtenue après 1 000 permutations des génotypes.



O. Lepais et al.

### 3.2.3. Structure génétique spatiale

Les chênes pédonculés de la Paguères présentent une structuration génétique significative pour toutes les classes de distance (fig. 4). Cette structuration diminue avec la distance même si elle reste significative. Lorsque les analyses sont effectuées séparément entre les chênes pédonculés de type franc-pied et cépée, la structuration reste significative mais faible pour les franc-pieds alors que les cépées ne montrent pas de structure (fig. 5).

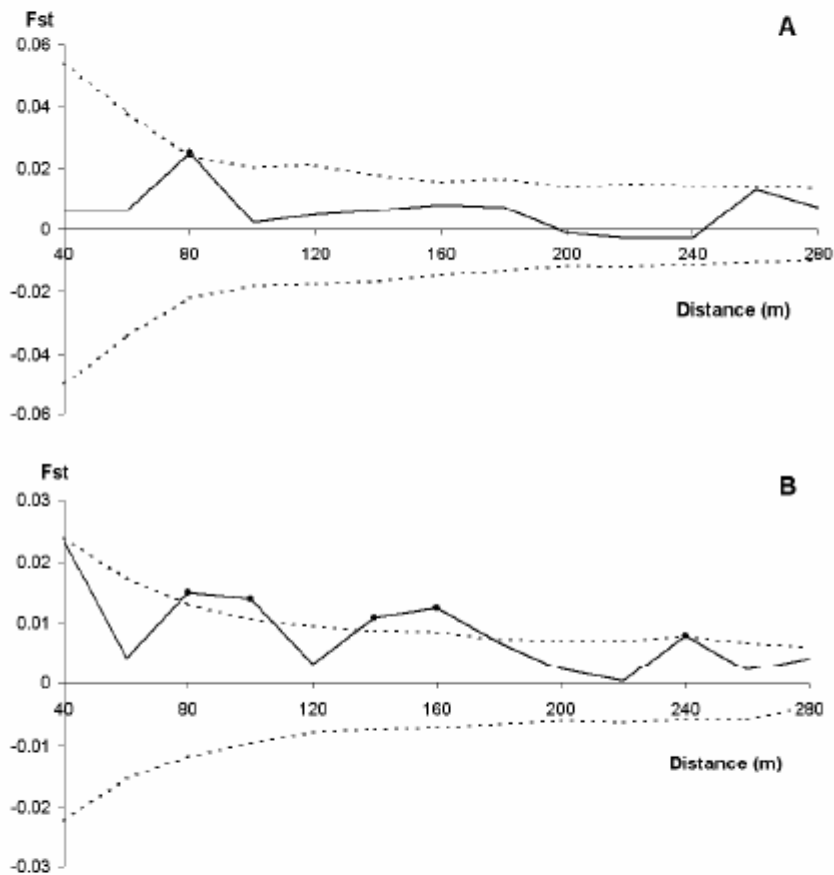


Figure 5 : Variation de  $F_{st}$  calculé à partir des génotypes de dix microsatellites pour différentes classes de distance chez les chênes pédonculés cépées (A) et franc-pieds (B) de la Paguères. Les lignes pointillées montrent l'enveloppe à 95 % des valeurs de  $F_{st}$  sous l'hypothèse nulle d'absence de structuration spatiale obtenue après 1 000 permutations des génotypes.

## 4. DISCUSSION

### 4.1. Diversité et structuration chloroplastique

Les grandes différences observées pour le génome chloroplastique entre le Tagon et Aurignac s'expliquent en grande partie par des raisons historiques. La fixation de l'haplotype 10a dans un rayon d'au moins 50 km chez toutes les espèces à Aurignac est un témoin de la recolonisation postglaciaire par les chênes [14]. Les chênes de la région sont issus des glands de quelques chênes venus des refuges glaciaires ibériques, portant l'haplotype 10a. Ces premiers chênes sont à l'origine d'un peuplement continu possédant une même lignée chloroplastique maternelle, saturant tous les sites disponibles et empêchant ainsi l'installation ultérieure de chênes provenant de refuges différents. Cette forêt a été ensuite fragmentée jusqu'à obtenir la configuration paysagère actuelle tout en conservant cette structuration haplotypique. Les bois de la région sont transmis de génération en génération [5], cette stabilité s'accompagne d'une gestion très locale basée sur la régénération naturelle. Le seul chêne ayant un haplotype différent dans notre échantillonnage d'Aurignac est certainement lié à une graine transportée naturellement à grande distance.

Le site du Tagon présente une configuration très différente avec une structuration des haplotypes à une échelle beaucoup plus fine, de l'ordre de 1 km. Tous les haplotypes appartiennent à la même lignée post-glaciaire d'origine ibérique. Cette structure fine pourrait être expliquée par une migration quasi-simultanée de ces haplotypes. L'exploitation du pin maritime au Tagon conduit à une dynamique paysagère qui se traduit par des cycles de colonisation – extinction pour les chênes, espèces non cibles de la gestion forestière dans cette région. Leur structure haplotypique dans la zone est alors constamment remaniée, comme en témoignent les corrélations entre haplotypes et milieux. Contrairement à la région d'Aurignac, le Tagon est une zone instable où la dispersion des chênes par graine est plus intense, en relation avec les perturbations anthropiques du milieu.

### 4.2. Relation entre haplotypes et espèces de chênes

Les trois espèces de chênes d'Aurignac partagent l'haplotype 10a. Le partage d'un même haplotype par différentes espèces de chêne d'une même population a été mis en évidence lors d'analyses de la diversité chloroplastique à l'échelle de la France entière [17] ou de l'Europe [3], [15]. Ce phénomène a été expliqué par la capacité du chêne sessile à se disséminer par introgression génétique dans les peuplements de chêne pédonculé. Le croisement du chêne sessile avec le chêne pédonculé (mère) aurait formé des hybrides interspécifiques, et des rétrocroisements successifs et dirigés (sessile

vers hybride), permettraient la redifférenciation d'individus de type sessile dans la population initiale de pédonculés [8]. Dans le cas d'Aurignac, ce partage d'haplotype est très accentué et concerne l'ensemble des espèces de chênes de la zone à haplotype 10a. Cet haplotype est absent d'une étude de l'ADN chloroplastique du chêne pubescent menée dans la péninsule ibérique [13]. L'abondance de cet haplotype parmi les chênes pubescents de notre étude pourrait s'expliquer par la colonisation de cette espèce par son pollen grâce à des hybridations récurrentes. Cette hypothèse est confortée par la diversité plus élevée de ces chênes, qui auraient profité des diversités des autres espèces. Le chêne pubescent apprécie les sols des coteaux calcaires le long des vallées (Garonne, Dordogne, Charente), zones où l'on retrouve les haplotypes 7 ou 10a qui pourraient résulter d'événements d'introgression génétique [15]. Une fois que tous les sites disponibles ont été colonisés par les chênes, la dispersion des glands devient limitée ce qui favorise la migration par pollen ([19], [18]) du chêne pubescent dans les milieux qui lui sont favorables. La coexistence de plusieurs espèces connectées par des flux de gènes est particulièrement intéressante dans la perspective d'évolution des peuplements face aux changements environnementaux biotiques comme abiotiques. En effet, ces espèces ont des préférences écologiques marquées et des sensibilités différentes aux pathogènes. Une plus grande richesse spécifique augmente alors les possibilités d'adaptation des populations de chênes directement par remplacement d'une espèce par une autre mieux adaptée ou indirectement par introgression de gènes à fort potentiel adaptatif [4].

Au Tagon, à l'opposé, lorsque les chênes pédonculés et tauzins sont présents au même endroit, ils ne partagent pas un même haplotype et il apparaît nettement que la structuration haplotypique est différente chez les deux espèces. Le potentiel d'hybridation entre ces deux espèces est peu connu mais cette absence de partage d'haplotype ne peut seule indiquer une hybridation impossible. En effet, les perturbations importantes de ce milieu pourraient effacer cette éventuelle structuration.

#### 4.3. Effet de la fragmentation et du mode de gestion sur la diversité

La fragmentation peut avoir deux conséquences sur les populations forestières : une diminution de la taille des populations ainsi qu'un isolement géographique qui peuvent induire des baisses de diversité. Les diversités génétiques, comparables entre une parcelle comprise dans un peuplement continu (la Petite Charrie) et un bois isolé au sein d'un paysage (la Paguères), montrent que la fragmentation ne se traduit pas par une perte de diversité chez les chênes. L'effet de la réduction de la taille des populations pourrait être contre-

## Diversité génétique des chênes

balancé par une augmentation de la distance des flux de gènes dans un paysage fragmenté comme cela a été montré chez le frêne [1].

Les analyses de structuration spatiale menées dans un peuplement géré en futaie régulière (Petite Charnie) montraient une structuration plus forte chez les chênes sessiles que chez le chêne pédonculé [21]. Dans la forêt de la Paguères, gérée en taillis sous futaie, une structuration génétique plus forte est observée pour les chênes pédonculés. La structuration observée à la Petite Charnie résulte d'une seule classe d'âge alors que celle mise en évidence à la Paguères résulte de la superposition de différentes classes d'âges. Les cépées regroupant des chênes plus âgés d'au moins une génération ont une structuration faible alors que les franc-pieds montrent une structuration forte. Ce chevauchement des générations dans un peuplement en taillis sous futaie ainsi que la régénération naturelle [4] renforce la structuration génétique spatiale. Néanmoins, il semble nécessaire d'analyser un bois échantillonné exhaustivement pour pouvoir comparer finement l'effet de la gestion sur la structure génétique du peuplement. Bien que la diversité génétique ne soit pas affectée par le mode de gestion au niveau intraspécifique, il pourrait avoir des conséquences sur la diversité en espèces des peuplements. Les bois gérés en taillis subissent des coupes rases partielles qui créent sporadiquement des zones ouvertes à l'intérieur du bois. Le chêne sessile, espèce typiquement forestière, se retrouve alors parfois en lisière, situation non favorable à cette espèce qui peut créer des dépérissements. Il est donc important de pérenniser les peuplements de taille importante (forêt de Fabas et Maubuisson dans la région d'Aurignac) afin de conserver une taille de population de chêne sessile suffisante. En outre, dans les bois gérés en taillis simple, les récoltes de bois se font souvent par coupe rase d'une petite partie du bois. La pression pour la ressource en bois est de plus en plus forte et a pour effet d'augmenter les surfaces récoltées. Une perte de diversité rapide pourrait alors intervenir dans le cas où la taille de la population reproductrice viendrait à sérieusement diminuer. Une gestion en taillis sous futaies qui conserve davantage d'arbres reproducteurs aurait le double avantage de réduire le temps de régénération du peuplement et de prévenir d'éventuelles pertes de diversité génétique. Néanmoins, malgré une gestion ancienne de ces bois, aucune perte de diversité n'a été mise en évidence lors de cette étude. De plus, la fragmentation n'a pas d'effet chez ces espèces à fort flux de pollen et pourrait même augmenter les flux de gènes à longues distances assurant une bonne connectivité entre fragments boisés.

La forte hétérogénéité qui caractérise la région d'Aurignac, zone de transition entre diverses influences, permet la cohabitation d'une grande diversité spécifique pour les chênes à l'intérieur d'un même peuplement. Cette caractéristique augmente les possibilités d'évolution des peuplements face à un environnement changeant. Une étude détaillée de la dynamique

O. Lepais et al.

d'hybridation au sein de ce complexe d'espèce dans la région aiderait à mieux comprendre la dynamique de ces populations et le lien entre diversité intra- et interspécifique.

## REMERCIEMENTS

Nous souhaitons remercier ici Jean-Marc LOUVET et Jean-Pierre ROSSI pour leur aide à différentes étapes de ce travail.

## RÉFÉRENCES

- [1] Bacles C.F.E., Lowe A.J., Ennos R.A., Effective seed dispersal across a fragmented landscape, *Science* 311 (2006) 628.
- [2] Belkhir K., Borsa P., Chikhi L., Raufaste N., Bonhomme F., GENETIX, logiciel sous Windows™ pour la génétique des populations, (1996), Laboratoire Génome, Populations, Interactions CNRS UMR 5000, Université de Montpellier II, Montpellier (France).
- [3] Dumolin-Lapègue S., Demesure B., Fineschi S., Le Corre V., Petit R.J., Phylogeographic structure of white oaks throughout the European continent, *Genetics* 146 (1997) 1475-1487.
- [4] Finkeldey R., Ziehe M., Genetic implications of silvicultural regimes, *Forest Ecology and Management* 197 (2004) 231-244.
- [5] Guyon J.P., Bessière F., Faignaert O., Bois J.C., Evolution des formations boisées sur le territoire de deux communes du canton d'Auzignac (Haute Garonne) de 1942 à 1992, *Etudes et Recherches sur les Systèmes Agraires et le Développement* 29 (1996) 139-152.
- [6] Hardy O.J., Vekemans X., SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels, *Molecular Ecology Notes* 2 (2002) 618-620.
- [7] Kampfner S., Lexer C., Glössl J., Steinkellner H., Characterization of (GA)<sub>n</sub> microsatellite loci from *Quercus robur*, *Hereditas* 129 (1998) 183-186.
- [8] Kremer A., Petit R.J. et Ducouso A., Biologie évolutive et diversité génétique des chênes sessile et pédonculé, *Revue Forestière Française* 2 (2002) 111-130.
- [9] Legendre P., Legendre L., Causal modelling: partial Mantel analysis, in: *Numerical Ecology 2<sup>nd</sup> English edition*, Elsevier Science B. V., Amsterdam, 1998, pp. 779-783.
- [10] Lepais O., Léger V., Gerber S., High throughput microsatellite genotyping in oak species, *Silvae Genetica* (2006): sous presse.
- [11] Nei M., Estimation of average heterozygosity and genetic distance from a small number of individuals, *Genetics* 89 (1978) 583-590.
- [12] Nei M., Chesser R.K., Estimation of fixation indices and gene diversities, *Ann Hum Genet* 47 (1983) 253-259.
- [13] Olalde M., Herran A., Espinel S., Goicoechea P.G., White oaks phylogeography in the Iberian Peninsula, *Forest Ecology and Management* 156 (2002) 89-102.

## Diversité génétique des chênes

- [14] Petit R.J., Pineau E., Demesure B., Bacilieri R., Ducouso A., Kremer A., Chloroplast DNA footprints of postglacial recolonization by oaks, *Proceedings of the National Academy of Sciences of the United States of America* 94 (1997) 9996-10001.
- [15] Petit R.J., Brewer S., Bordács S., et al., Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence, *Forest Ecology and Management* 156 (2002b) 49-74.
- [16] Petit R.J., Csaihl U.M., Bordács S., et al., Chloroplast DNA variation in European white oaks, phylogeography and patterns of diversity based on data from over 2600 populations, *Forest Ecology and Management* 156 (2002a) 5-26.
- [17] Petit R.J., Latouche-Hallé C., Pemonge M.H., Kremer A., Chloroplast DNA variation of oaks in France and the influence of forest fragmentation on genetic diversity, *Forest Ecology and Management* 156 (2002c) 115-129.
- [18] Petit R.J., Bodénès C., Ducouso A., Roussel G., Kremer A., Hybridization as a mechanism of invasion in oaks, *New Phytologist* 161 (2004) 151-164.
- [19] Potts B.M., Reid J.B., Hybridization as dispersal mechanism, *Evolution* 42 (1988) 1245-1255.
- [20] Steinkellner H., Fluch S., Turetschek E., Lexer C., Streiff R., Kremer A., Burg K., Glössl J., Identification and characterization of (GA/CT)<sub>n</sub> – microsatellite locus from *Quercus petraea*, *Plant Molecular Biology* 33 (1997) 1093-1096.
- [21] Streiff R., Labbe T., Bacilieri R., Steinkellner H., Glössl J., Kremer A., Within population genetic structure in *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. assessed with isozymes and microsatellites, *Molecular Ecology* 7 (1998) 317-328.
- [22] Weir B.S., Cockerham C., Estimating F-Statistics for the analyses of population structure, *Evolution* 6 (1984) 1358-1370.



**Leaf morphological analyses in four European oak species (*Quercus*) and their hybrids: a comparison of traditional and geometric morphometric methods.**

VINCENZO VISCOSI<sup>1\*</sup>, OLIVIER LEPAIS<sup>2,3</sup>, SOPHIE GERBER<sup>2,3</sup> AND PAOLA FORTINI<sup>1</sup>

<sup>1</sup>Dipartimento di Scienze e Tecnologie per l'Ambiente e il Territorio, Università del Molise, Contrada Fonte Lappone, Pesche, I-86090, Italy.

<sup>2</sup>INRA, UMR1202 BIOGECO, 69 route d'Arcachon, F-33612 Cestas, France

<sup>3</sup>Université de Bordeaux 1, UMR1202 BIOGECO, 69 route d'Arcachon, F-33612 Cestas, France.

\* Corresponding author : vincenzoviscosi@yahoo.it ; Tel. and Fax: +39 0874404174

**ABSTRACT**

In this study, leaf morphology was assessed in a mixed oak stand (western France) using two geometric morphometrics (landmarks and outlines) a data set of 19 leaf measures.

All adult oaks (817 oaks) were sampled, comprising four white oak species (*Quercus petraea*, *Q. robur*, *Q. pubescens* and *Q. pyrenaica*) for DNA extraction and genetic analysis. Leaf morphology was assessed on a subset of 336 oaks, comprising pure species and hybrids as determined by genetic assignment.

This comparative study of oak leaf morphology, based on the use of two free size geometric morphometric methods (landmark and outline) and of a set of leaf measurements, combined with the genetic assignment of individuals to pure species or hybrids, provided information about the differences among species and the intermediate leaf morphology of their hybrids.

**Key words:** genetic assignment, hybridization, traditional morphometrics, geometric morphometrics, leaf morphology, nuclear microsatellites, *Quercus*.

**Introduction**

The genus *Quercus* (*Fagaceae*) is widespread in the northern hemisphere. Deciduous oaks are significant components of temperate forests of North America and Eurasia (Camus, 1936-54; Manos, 1999). In Europe, subgenus *Quercus* spread quickly during the second half of the Tertiary era (Axelrod, 1993) and during the ice ages of the Quaternary period (120,000 – 18,000 years ago). The southern areas of the Iberian, Italian and Balkan peninsulas were important refuges for numerous oak populations (Brewer et al., 2002).

Several studies have investigated post-glacial routes of oaks using maternal inherited chloroplast genome (cpDNA) and palynological technics (Brewer et al., 2002; Petit et al., 2002a) and have established that dynamics involved movements from refuge areas towards the north. In France, oak populations originated from all three primary refuge areas, located in Italian, Balkan and Iberian peninsulas (Petit et. al., 2002b). Oak species were not correlated with a particular refugia, however, spatial analysis of chloroplast DNA at a regional scale indicated that species occurring in sympatry tend to share the same maternal



genome indicating a potential role of hybridization during postglacial migration (Petit et al., 1997; Petit et al., 2004).

The interspecific gene flow is a common phenomenon in genus *Quercus*, owing to the absence of real reproductive barriers within the genus (Burger, 1975; Spellenberg, 1995; Gonzalez-Rodriguez et al., 2004); moreover, natural hybridization was observed by mating system analysis with gene markers in natural mixed populations (Bacilieri et al., 1996; Streiff et al., 1999). As a consequence introgression has led to a proliferation of fertile hybrids displaying a very wide range of intermediate morphological features (Gellini & Grossoni, 1997).

The hybridization has been attributed to both genetical and ecological effects (Boavida et al., 2001; Belahbib et al., 2001; Petit et al., 2003). When the ranges of two or more species overlap, hybrids occur more frequently at their geographical or ecological margins (Valbuena-Carabana et al., 2005), where environmental conditions might be more suitable for hybrids with intermediate characters (Williams et al., 2001).

Despite this, the existence of interspecific differences has been shown in some studies based on genetic markers. Bordàcs & Burg (1997) demonstrated the existence of two species-specific RAPD markers differentiating *Q. robur* from *Q. petraea*, while more recently Scotti-Saintagne et al. (2004) have shown that interspecific differentiation was heterogeneous and spatially autocorrelated across the genomes using a variety of molecular markers (isozymes, AFLPs, SCARs, microsatellites and SNPs). Independently, natural oak population studies based on microsatellites and individual-based genetic assignment methods (Muir, 2000; Gugerli et al., 2007; Valbuena-Carabana et al., 2005, 2007; Muir et al., 2005; Curtu et al., 2007) provided multiple evidences that oaks species represent separate taxonomical units.

Morphological analyses have also been assessed for species differentiation in European white oaks (Dupouey & Badaeu, 1993; Bacilieri et al., 1995; Kleinschmit et al., 1995; Bruschi et al., 2000; Kremer et al., 2002; Borazan & Babac, 2003), but, as hybrid identification method was not available, information on hybrid morphology remained poor. However, two recent studies combining genetic assignment and morphological analysis (Gugerli et al., 2007; Curtu et al., 2007) demonstrated a general congruence of both species determination. Hybrids, as defined by genetic assignment, did not necessarily present intermediate morphology, so that intermediate morphology does not systematically mean hybridization. These results confirm the polygenic nature of morphological species differences in oak species demonstrated by quantitative traits loci analyses (QTL; S et al., 2004). In addition, complexity, morphological variability and plasticity of the phenotype could render morphological features inadequate for quantifying the degree of hybridization. In this study, the morphological analysis of leaves was undertaken to investigate the interspecific relationships among four sympatric oak species (*Quercus pubescens* Will., *Q. robur* L., *Q. petraea* (Matt.) Liebl. and *Q. pyrenaica* Will.) and their hybrids, in a mixed stand of western France. Using a Bayesian approach, genetic markers (nuclear microsatellites) were used to assign individuals to species or hybrid class without *a priori* classification. In the present article we will refer to species determination as a result of genetic assignment only, following the blind approach (D et al., 2006). This method provides an independent classification of individual trees that can be compared to morphological analysis. Leaf morphology was assessed by means of morphological measures (Kremer et al., 2002), well suited to discriminate two additional species, and two geometric morphometric methods (landmark and outline) described in Viscosi et al. (2008). The aims of the present study are, firstly, to investigate the morphological relationships among the four oak species and their genetic hybrids in a mixed stand, secondly to analyze the correlation among the three morphological datasets used in relation to genetic assignment of individuals to species or hybrid groups. Also, this work points out the functional capacity of geometric morphometrics methods, (Viscosi et al., 2008), to analyze

the leaf shape morphology of European white oaks, when molecular species assignment of individuals, or more generally when molecular markers are available. In addition to population genomics tools, this new method could be very efficient to give new insights about morphological species differences with a more functional perspective.

### Material and Methods

All adult oaks (817 oaks) were sampled in a mixed oak stand (Briouant) located in southwestern France and comprising four white oak species (*Quercus petraea* [Matt.] Liebl., *Q. robur* L., *Q. pubescens* Willd. and *Q. pyrenaica* Willd.) for DNA extraction and genetic analysis. Leaf morphology was assessed on a subset of 336 oaks, comprising pure species and hybrids as determined by genetic assignment. For all selected trees, 10 leaves were sampled in the highest part of the tree, to perform morphological analyses.

#### Genetic analysis

Genotyping of 10 microsatellites loci (Kampfer et al., 1997; Steinkellner et al., 1997) was done on all adult trees according to a multiplex protocol described in Lepais et al. (2006). A Bayesian assignment method implement in the software Structure version 2.1 (Pritchard et al., 2000; Falush et al., 2003) was carried out to assign individuals to species without any other information than multilocus genetic data (field species determination was not used as a prior for the analyse). Because such a method need that all parental populations have been sampled to recover the genetic structure, and given the very small number of oaks that looks like *Q. petraea*, genetic assignment analysis were performed with several others oak populations with a total of 2163 individuals (Lepais et al., in prep.), providing a sufficient number of individuals per species.

In short, Structure was run with the admixture model, 100000 Burnin periods and 1000000 MCMC repeat after burnin. The number of clusters was set to  $K = 4$ , each cluster corresponding to one oak species, providing the best partition of the genetic data. Genetic assignment of individuals to species was obtained by probability to belong to each of the four clusters computed by Structure. A specimen was assigned to a species, when its propability for one cluster was  $> 0.90$ . Individuals that could not be assigned to a pure species were considered hybrid originated from a cross between the two species showing the higher probabilities. This assignment procedure, tested by simulations, was found (1) to slightly underestimate hybrid proportions (Lepais et al., in prep) and (2) to assign correctly up to the second hybrid generation. So we defined hybrids as groups of first (F1) and second generation hybrids (backcrosses and F2).

All specimens were assigned to four species and six hybrids classes. The 336 trees analysed for leaf morphology were assign as follows: *Q. robur* (77), *Q. petraea* (6), *Q. pubescens* (54) and *Q. pyrenaica* (67) for pure species; *robur* X *petraea* (10), *robur* X *pubescens* (21), *robur* X *pyrenaica* (29), *pubescens* X *pyrenaica* (41), *petraea* X *pubescens* (8), *petraea* X *pyrenaica* (23) for hybrids.

#### Morphological analysis

Leaf morphological analyses were performed by means of three different methods: a morphological analysis of leaf variables as described in Kremer et al. (2002) and two geometric mophometric analyses of leaf shape (landmark and outline), performed following Viscosi et al. (2008).

First, leaf morphology was assessed by means of 19 variables (Kremer et al., 2002): petiole length (pl), lamina length (ll), lobe width (lw), sinus width (sw) and length of the lamina at largest width (wp) were measured on leave pictures and converted in millimetre according to the scale of each picture. Number of lobes (nl) and number of veins (nv) were counted for each leaf. Basal shape of the lamina (bs) was scored from 1 to 9 and pubescence density was estimated from 0 (no pubescence) to 5 (dense pubescence) for abaxial lamina

(hr), adaxial lamina (sr), central nervure (nr) and petiole (ptr). We used five transformed variables: lamina shape (ob), petiole ratio (pr), lobe depth ration (ldr), percentage of venation (pv) and lobe width ratio (lwr). To take into account the general lamina morphology, we used the software ImageJ (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, <http://rsb.info.nih.gov/ij/>, 1997-2007.) to measure leaf perimeter and circularity calculated as  $c = 4\pi(\text{area}/\text{perimeter}^2)$  which takes a value of 0 for a perfect circle and 1 for a increasingly elongated polygon. Second, for geometric morphometric analyses scanned images were used to record 11 landmarks on the right half of each leaf and the outlines of leaves were automatically recorded and reduced to Fourier coefficients for 50 harmonics (Viscosi et al., 2008).

### *Statistical analysis*

For morphological measurements we computed the mean of 19 leaf variables for all trees, and Principal Component Analysis (PCA) was performed on this dataset shifting variables to zero centered and scaling them to have a unit variance. For landmark data, Relative Warp Analysis (Rohlf & Slice, 1990; see Viscosi et al., 2008) was performed on standardized matrix of 336 consensus extracted from all individual trees, after multivariate analysis of outliers; all configurations were optimally rotated to minimize the squared differences between corresponding landmarks by using the Generalized Procrustes Analysis (Rohlf & Slice, 1990). For outline data, PCA was performed on variance-covariance matrix extracted from coefficients of 50 harmonics and the size and orientation of each contour was standardized by the longest radius method (Kuhl & Giardina, 1982; Iwata & Ukai, 2002).

Both, landmark and outline data sets were implemented by means of multivariate analysis of the outliers (Gschwandtner & Filzmoser, 2004; Filzmoser, 2005) All outliers identified through this procedure were eliminated from the data matrices of landmarks and outlines for each individual tree. A consensus configuration (landmark data) and the mean of EFDs (outline data) were extracted for each tree (see Viscosi et al., 2008).

Afterwards, to realize a model for each data set, all pure species trees were subjected to canonical discriminant analysis (CVA) using principal component scores (for morphological data and for outline data) and relative warp scores (for landmark data). The discriminant functions extracted were used to compute the pairwise group comparisons (step by step method) to verify discrimination degree among species. Moreover, because in discriminant function analysis the solutions are often unreliable, the test of cross validation (the process of testing a model on more than one sample) was performed to assess the reliability and generalisability of the findings for each data set.

Then, to verify if hybrids had an intermediate leaf morphology between the two parental species, three new matrices containing hybrids were obtained and analyzed by means of PCA. For each pure species and hybrid groups we computed the mean of leaf variables (morphological measures), the consensus configuration (landmarks) and the mean EFDs (outline).

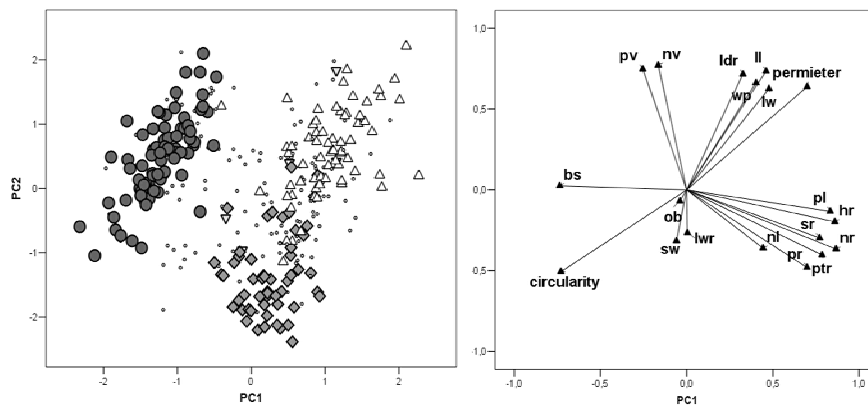
Finally, three Mantel test (Mantel, 1967) were run among the euclidean distance matrices of three data sets, that were computed on the average canonical scores of species without hybrids (using NTSYSpc 2.2)

## **Results**

### *Morphological leaf variables*

From principal component analysis, the two first axes explained 58.50% of the variance with 33.89% for PC1 and 24.61% for PC2. The ordering of individuals and variable vectors on the first two PCs (Fig. 1), highlighted that *Q. robur* individuals were clearly separated from the others mainly due to the effect of basal shape, circularity and other variables (pl, hr, sr, nr, nl, pr and ptr). *Q. pyrenaica* and *Q. pubescens* were partially overlapped, but

distinguishable by perimeter and circularity variables. For *Q. petraea* clear results were not found, because of low number of individuals in the dataset.



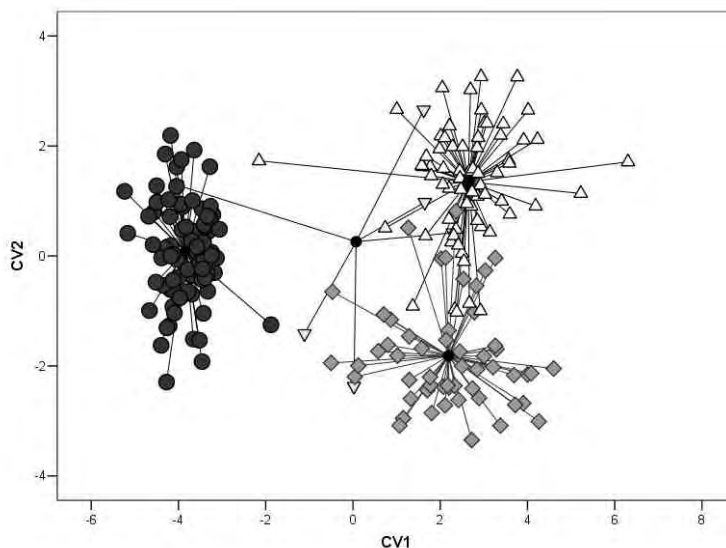
**Figure 1.** Ordination diagrams (PCA) computed on 19 leaf morphological variables and 336 trees. On the left was reported the distribution of trees on PC1 and PC2 axes; on the right was reported the vector plot of variables on PC1 and PC2 axes. Filled circles (*Q. robur*); filled diamonds (*Q. pubescens*); open triangles (*Q. pyrenaica*); open upturned triangles (*Q. petraea*); small point (hybrids).

Afterwards, from canonical discriminant analysis (CVA) the greater part of total variance (99.39%) was explained along the first two canonical variates (CVs): respectively, CV1 (Wilks'  $\lambda = 0.036$ ;  $df = 15$ ;  $p < 0.001$  ) and CV2 (Wilks'  $\lambda = 0.375$ ;  $df = 8$ ;  $p < 0.001$  ) explained 85.58% and 13.81% of the total variance; while CV3 (Wilks'  $\lambda = 0.939$ ;  $df = 3$ ;  $p = 0.006$  ) explained only 0.61% of total variance.

Scatter plot of CV1 and CV2 (Fig. 2) showed that along CV1, *Q. robur* was separated from all other three species, while CV2 explained differences between *Q. pyrenaica* and *Q. pubescens*.

From pairwise group comparisons the differences among the four species appeared significant the greater differences were observed between *Q. robur*, *Q. pubescens* and *Q. pyrenaica*, but also *Q. petraea* was significantly discriminated from all other species (Table I).

From test of cross validation, 90.7% of oaks were correctly classified: all specimens for *Q. robur*, the greater part for *Q. pubescens* and *Q. pyrenaica*, and no trees for *Q. petraea* (Table II).



**Figure 2.** Scatter plot of the first two canonical variate scores for 19 morphological variables. Filled circles (*Q. robur*); filled diamonds (*Q. pubescens*); open triangles (*Q. pyrenaica*); open upturned triangles (*Q. petraea*); hybrids were not projected.

**TABLE I.** Pairwise Group Comparisons. F-values (below) and *p*-values (above) were indicated for each species comparison at step 5; degree of freedom (df) were 5.196, 6.195 and 10.191 for 19 leaf variables landmark data outline data, respectively. F-values significant at  $p < 0.001$  are in bold type.

<u>19 Leaf Variables</u>				
Species	<i>Q.pet</i>	<i>Q.pub</i>	<i>Q.pyr</i>	<i>Q.rob</i>
<i>Q.petraea</i>	-	< 0.001	< 0.001	< 0.001
<i>Q.pubescens</i>	<b>11.62</b>	-	< 0.001	< 0.001
<i>Q.pyrenaica</i>	<b>10.85</b>	<b>59.93</b>	-	< 0.001
<i>Q.robur</i>	<b>19.17</b>	<b>248.89</b>	<b>306.75</b>	-
<u>Landmark Data</u>				
Species	<i>Q.pet</i>	<i>Q.pub</i>	<i>Q.pyr</i>	<i>Q.rob</i>
<i>Q.petraea</i>	-	0.002	< 0.001	< 0.001
<i>Q.pubescens</i>	3.61	-	< 0.001	< 0.001
<i>Q.pyrenaica</i>	<b>4.65</b>	<b>43.59</b>	-	< 0.001
<i>Q.robur</i>	<b>8.88</b>	<b>94.30</b>	<b>91.94</b>	-
<u>Outline Data</u>				
Species	<i>Q.pet</i>	<i>Q.pub</i>	<i>Q.pyr</i>	<i>Q.rob</i>
<i>Q.petraea</i>	-	0.064	< 0.001	0.002
<i>Q.pubescens</i>	1.79	-	< 0.001	< 0.001
<i>Q.pyrenaica</i>	<b>3.74</b>	<b>30.64</b>	-	< 0.001
<i>Q.robur</i>	2.91	<b>35.08</b>	<b>52.65</b>	-

**TABLE II.** Results of test of cross-validation of the three data sets. 90.7%, 87.7% and 81.9% of cross-validated grouped cases were correctly classified for 19 leaf variables, landmark and outline data, respectively.

<u>19 leaf variables</u>					
Cross-valid.(%)	<i>Q.pet</i>	<i>Q.pub</i>	<i>Q.pyr</i>	<i>Q.rob</i>	Total
<i>Q.petraea</i>	<b>0.0</b>	16.7	50	33.30	100.0
<i>Q.pubescens</i>	1.9	<b>87.0</b>	11.1	0.0	100.0
<i>Q.pyrenaica</i>	0.0	7.5	<b>91.0</b>	1.5	100.0
<i>Q.robur</i>	0.0	0.0	0.0	<b>100.0</b>	100.0
<u>landmark data</u>					
<i>Q.petraea</i>	<b>0.0</b>	50.0	33.3	16.7	100.0
<i>Q.pubescens</i>	1.9	<b>79.5</b>	13.0	5.6	100.0
<i>Q.pyrenaica</i>	0.0	9.0	<b>89.5</b>	1.5	100.0
<i>Q.robur</i>	0.0	0.0	1.3	<b>98.7</b>	100.0
<u>outline data</u>					
<i>Q.petraea</i>	<b>16.7</b>	33.3	16.7	33.3	100.0
<i>Q.pubescens</i>	9.3	<b>74.1</b>	9.3	7.3	100.0
<i>Q.pyrenaica</i>	3.0	6.0	<b>86.5</b>	4.5	100.0
<i>Q.robur</i>	7.8	3.9	0.0	<b>88.3</b>	100.0

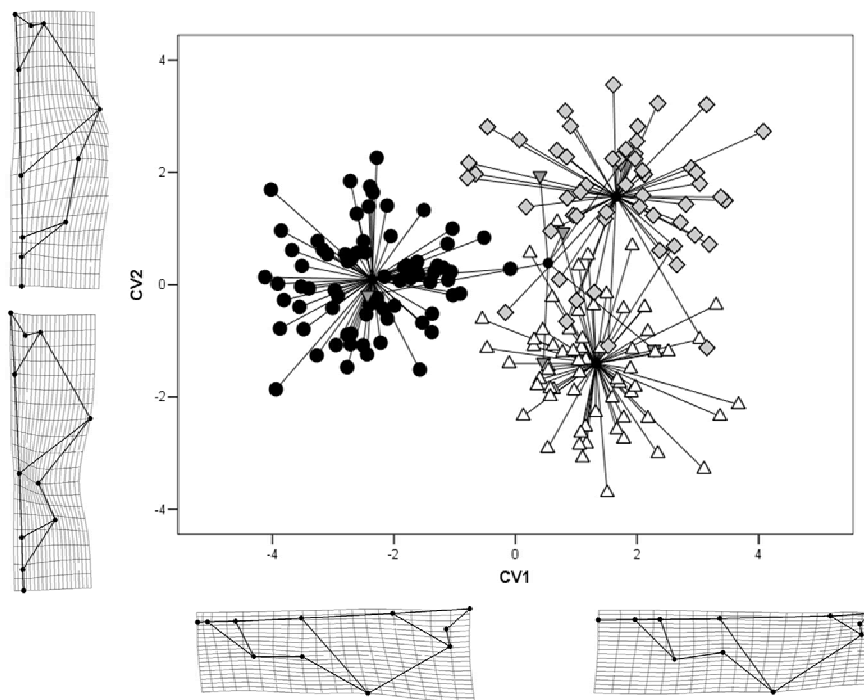
#### Landmark data

Relative Warp Analysis showed that the greater part of total variance (83.72%) was explained along the first four RWs: first two RWs explained only 55.31% of total variance,

with 32.85% and 22.49%, respectively for RW1 and RW2; while RW3 and RW4 explained 16.75% and 11.66% of total variance, respectively.

The pattern of shape variation was clarified in the canonical variate space (Fig. 3), where species were significantly discriminated along CV1 (Wilks'  $\lambda = 0.092$ ;  $df = 18$ ;  $p < 0.001$ ) and CV2 (Wilks'  $\lambda = 0.411$ ;  $df = 10$ ;  $p < 0.001$ ), that explained 99.20% of the total variance. CV3 was not significant (Wilks'  $\lambda = 0.961$ ;  $df = 4$ ;  $p = 0.095$ ). Specifically, CV1 (71.60% of total variance) explained differences the length of the petiole and the shapes of the basal and apical regions and clearly distinguished *Q. robur* from *Q. pubescens* and *Q. pyrenaica*; CV2 (27.50%), explaining the leaf variability about deepness of lobes and shape of apical region, separated *Q. pubescens* from *Q. pyrenaica*. Pairwise group comparisons were computed among species and the differences among *Q. robur*, *Q. pubescens* and *Q. pyrenaica* were meaningful, also *Q. petraea* was discriminated from other species (Table I).

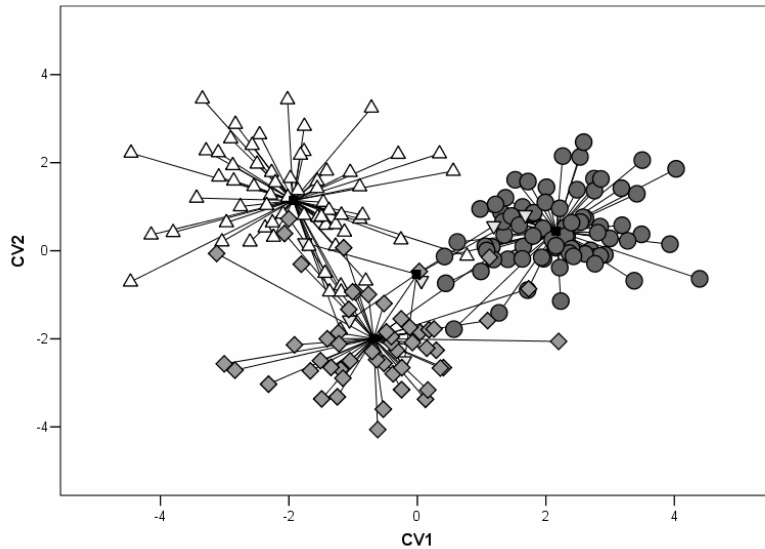
The test of cross-validation showed high percentage of cases correctly classified (87.7%). For *Q. robur*, *Q. pubescens* and *Q. pyrenaica*, greater part of specimens were correctly classified (Table II).



**Figure 3.** Scatter plot of the first and the second canonical variate scores derived from landmark data. Deformation grids showing major shape features correlated with the extremes of variation along the axes and were obtained through regression of shape variables on discriminant scores. Filled circles (*Q. robur*); filled diamonds (*Q. pubescens*); open triangles (*Q. pyrenaica*); open upright triangles (*Q. petraea*); hybrids were not projected.

#### Outlines data

From PCA resulted that P1 and PC2 explained 67.15% of total variance, with 55.70% and 11.45%, respectively. Discriminant analysis carried out similar results to other data sets: three species were significantly discriminated along first two CVs (Fig. 4); from pairwise group comparisons resulted that only the differences between *Q. petraea* and *Q. pubescens* resulted not significant (Table I); the test of cross-validation (leave-one-out method) showed high percentage of correct classification (Table II).



**Figure 4.** Scatter plot of the first and the second canonical variate scores derived from outline data. Filled circles (*Q. robur*); filled diamonds (*Q. pubescens*); open triangles (*Q. pyrenaica*); open upright triangles (*Q. petraea*); hybrids were not projected.

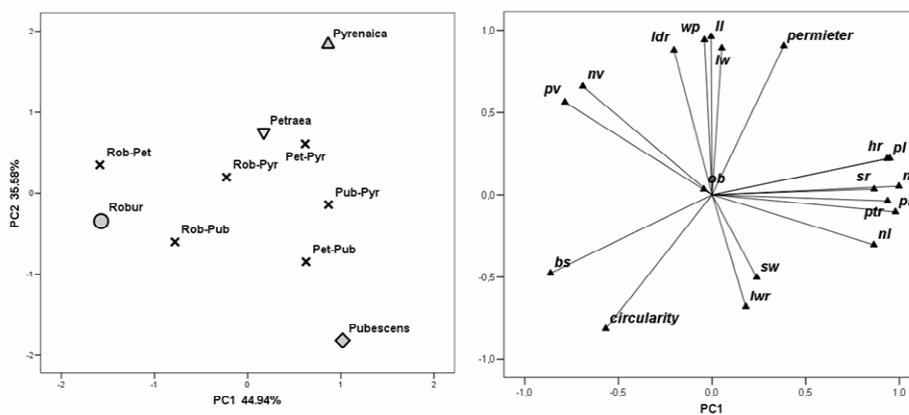
*Analyses of hybrids*

For each data set, the mean leaves of the ten groups, comprising four pure species and six hybrid groups, were analyzed by means of PCA.

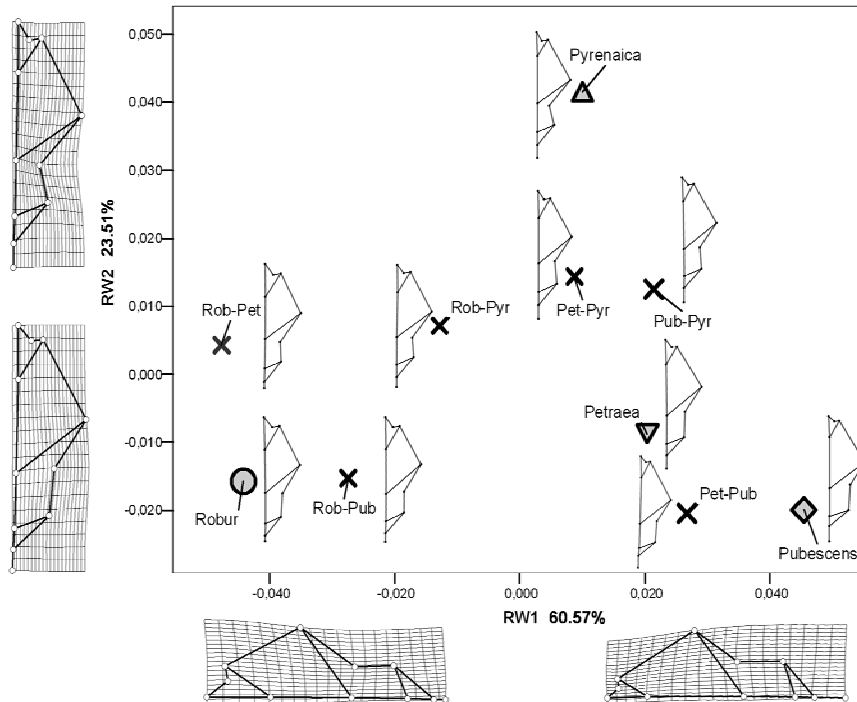
For morphological leaf variables, PC1 and PC2 explained 80.52% of total variance, it appeared that the projections of hybrid groups of *robur X pubescens*, *robur X pyrenaica*, *petraea X pubescens*, *pubescens X pyrenaica*, were located between the parental species, for each PC (Fig. 5).

RWA was performed for landmark data: the first two RWs explained 84.09% of the total variance, resulting an intermediate leaf shape for the hybrid groups, in respect to parental species, except for *robur X petraea* (Fig. 6).

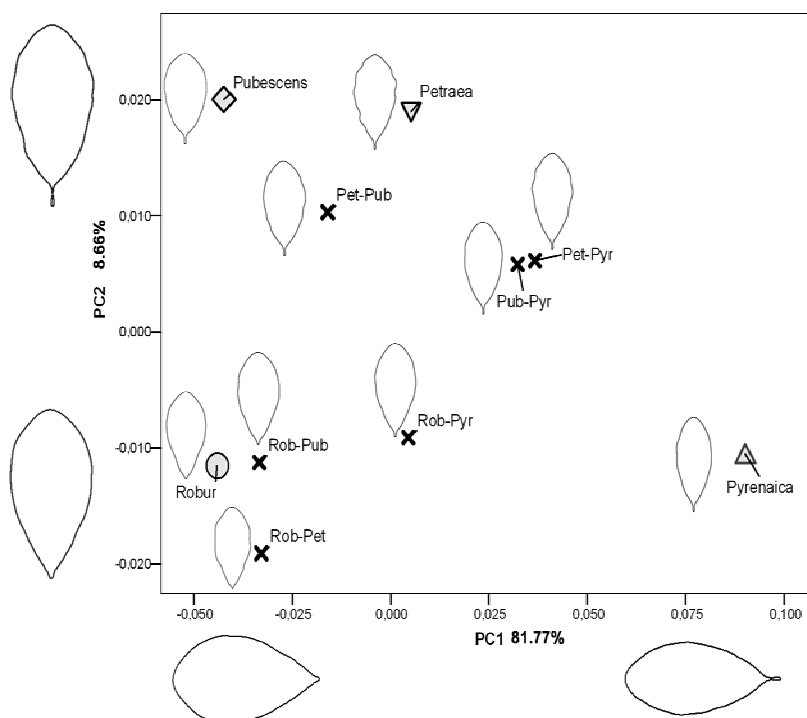
For outline data, PC1 and PC2 explained 90.43% of total variance and all hybrid groups, in respect to parental species, were intermediate located at first one or both PCs (Fig. 7).



**Figure 5.** Ordination diagrams (PCA) computed on 19 leaf morphological variables and 10 average group (4 pure species and 6 hybrid group). On the left was reported the distribution of average leaf on PC1 and PC2 axes; on the right was reported the vector plot of variables on PC1 and PC2 axes. Filled circles (*Q. robur*); filled diamonds (*Q. pubescens*); open triangles (*Q. pyrenaica*); open upright triangles (*Q. petraea*); cross (hybrid groups).



**Figure 6.** Scatter plots derived from Relative Warp Analysis performed on consensus configuration of landmark computed for 10 groups (4 pure species and 6 hybrid group). Filled circles (*Q. robur*); filled diamonds (*Q. pubescens*); open triangles (*Q. pyrenaica*); open upturned triangles (*Q. petraea*); cross (hybrid groups).



**Figure 7.** Scatter plot of PC1 and PC2 scores derived from mean EFDs (outline data) of 10 groups (4 pure species and 6 hybrid group) and reconstruction of shape variation along PC1 and PC2 for d.s.  $\pm 2$  values. Filled circles (*Q. robur*); filled diamonds (*Q. pubescens*); open triangles (*Q. pyrenaica*); open upturned triangles (*Q. petraea*); cross (hybrid groups).

*Correlation between matrices*



From the three Mantel tests (Mantel, 1967) performed, all correlations resulted significant and data sets were highly correlated; in particular landmark data was more strictly related to morphological measurements than outline data (Table III).

**TABLE III.** 2-way Mantel test between data sets. For each matrices correlation were indicate *r*-values (below) and *p*-values (above) computed on 999 random permutations.

<b>r-values</b>	<b>landmark</b>	<b>outlines</b>	<b>morpho</b>
<b>landmark</b>	1	0.041	0.042
<b>outlines</b>	0.872	1	0.049
<b>morpho</b>	0.910	0.789	1

## Discussion

In this paper, an integrated analysis of leaf morphology of some European white oaks species was proposed. The species identification was performed by means of genetic markers and assignment of individuals to species or hybrids using unsupervised Bayesian analyses. This approach assumes that specimens from the same species will have a more similar genotype than that of individuals from a different species (Baudouin et al., 2004; Congdon, 2001).

Many reports about intermediate phenotypes have received special attention, first by delineating a specific class of intermediate phenotypes in comparison to reference populations (Elsner, 1993; Rushton, 1983; Ietswaart & Feij, 1989) and then by designating these trees as introgressed forms. Other papers indicate that the within species variation encompassed these intermediate forms, despite the interfertility between the two species (Kremer et al., 2002). Natural hybridization has been shown by mating system analysis with gene markers in natural mixed populations (Bacilieri et al., 1996 ; Sreiff et al., 1999) and interspecific hybridization was also demonstrated by controlled crossings (Kleinschmit & Kleinschmit, 2000).

This comparative study of oak leaf morphology, based on the use of two free size geometric morphometric methods (landmark and outline) and of a set of leaf measurements, combined with the genetic assignment of individuals to pure species or hybrids, provided information about the differences among species and the intermediate leaf morphology of their hybrids.

Genetic assignment of individuals clearly demonstrated that the analyzed mixed oak stand consisted of four species (*Q. robur*, *Q. pubescens*, *Q. petraea* and *Q. pyrenaica*) and their genetic hybrids.

The multivariate statistical analyses (PCA, CVA) were performed on all three morphological data sets and similar results were found about the leaf morphology of species and the relationships with their hybrids.

First, for all morphological data sets, the original variables were summarized in several synthetic variables (PCs) that were subjected to discriminant analysis (CVA). It resulted that all species were discriminated by means of all three morphological methods: the greater part of differences among species were explained among *Q. robur*, *Q. pubescens* and *Q. pyrenaica*, while *Q. petraea* was less clearly differentiated, because of low number of individuals sampled.

Second, for each data set a mean was computed for all defined groups: four pure species and six hybrid groups. The three new matrices obtained, were subjected to PCA and the mean of each group was projected on PC1 and PC2 axes, to visualize the location of hybrid groups in respect to their parental species. For all three data sets, hybrids were located

between the two parental species on the PC1 and PC2 axes, indicating that hybrids are globally characterized by an intermediate but highly variable leaf morphology.

Like other papers that reported the use of the same methods on European white oaks (Dupouey & Badaeu, 1993; Kremer et al., 2002; Curtu et al., 2007; Viscosi et al., 2008), our results confirmed that oak species were distinguishable by means of leaf morphology.

Morphological data analysis presented in Kremer et al. (2002) were established to discriminate *Q. petraea* from *Q. robur*. In the present study, three additional variables (pubescens of the adaxial area, perimeter and roundness) were added to improve the discrimination power for *Q. pyrenaica* and *Q. pubescens* species. To discriminate *Q. petraea* from *Q. robur*, Kremer et al. (2002) reported that petiole length (pl and pr), intercalary venation (nv and pv), pubescence (pu) and sinus width (sw and ldr) were the original variables that exhibited the highest weight in species differentiation, while the basal shape of the lamina (bs) and number of lobes (nl) contributed moderately. Our results confirmed these discriminant variables and identified others that were discriminating for other species. While pubescence variables differentiate *Q. robur* from *Q. pubescens* and *Q. pyrenaica*, the leaf circularity and some dimensional variables as lobe depth ration (ldr), lamina length (ll), perimeter, length of the lamina at largest width (wp), lobe width (lw), distinguished *Q. pyrenaica* from other species. *Q. pubescens* was principally distinguished by means of percentage venation (pv) and number of veins (nv).

Leaf shape analyses (landmark and outline methods) confirmed as explained in Viscosi et al. (2008): *Q. robur* was characterized by short petiole and obovate leaf blade with a narrow and auriculate basal shape; *Q. petraea* was distinguished by elliptical leaf blade with a sub-acute apical lobe and for *Q. pubescens* the leaf blade was elliptic-obovate with cordate basal region and obtuse apex. Finally, *Q. pyrenaica* was characterized by an elliptical and deeply lobate leaf blade.

In this paper, we tested the discrimination power of the geometric morphometric approaches in comparison with other morphological measurements and analyzed morphological relationships among species and their putative hybrids, identified using molecular markers (microsatellites). We discovered that genetic assignment using 10 microsatellites is able to discriminate hybrids up to the second generation. This is clearly a limitation since species group will contain some third or latter generation hybrids, which can partially explain the high morphological variability observed in pure species. A more precise vision of hybrid morphology will be gained by adding additional genetic markers to get a deeper and more complete view of morphological consequences of hybridization. Due to the increasing availability of Expressed Sequence Tags (EST) in many species including oaks (e.g. European "Network of Excellence" EVOLTREE: Evolution of Trees as Drivers of Terrestrial Biodiversity), rapid development of numerous new molecular markers, in particular microsatellites located in transcribed region of the genome (EST-SSR), will lead to an increasing number of loci and to fast genotyping as well. This kind of loci could show a much higher genetic differentiation between species because they are located in transcribed region (Pashley et al., 2006; Ellis & Burke, 2007). As a result, it will soon be possible to genotype oaks with a much higher number of loci showing high genetic differentiation, a technological advance that will allow to increase resolution for latter generation hybrids identification. In addition, this will offer a more functional and evolutionary perspective in the field of morphological differences between species (see Kashi & King, 2006 for a recent review), especially when coupled with population genomics approaches (i. e., using a high number of genetic markers to saturate the whole genome; Scotti-Saintagne et al., 2004). The new oak geometrics morphometrics analysis presented in this issue (Viscosi et al., 2008) compared in the present study with a classical morphological analysis method, provides a fast and efficient tool for phenotyping oak leaves morphology, in conjunction with the upcoming new powerfull genotyping technics. Finally, we observed that hybrids were a mosaic of phenotypes with parental and

intermediate characters. This result could be in part due to the limitation of hybrid identification by our genetic markers as discussed above, but it also highlights the fact that hybrid groups are not homogeneous but instead a mixture of heterogeneous genotypes (Anderson & Jr, 1954; Arnold, 1997). Nevertheless when mean leaves were analyzed for hybrid groups, they resulted in intermediate leaf morphology, in respect to parental species. This confirms the difficulties to use morphology features for hybrid identification, even if this issue will probably be clarified to some extent by analysing a larger sample of species and locations, to understand the taxonomic and geographical distribution of species hybrids in European white oaks.

## ACKNOWLEDGEMENTS

We thank Jérôme Willm and Guy Roussel for helping in collecting leaves and Florian Alberto for leaves measurements. Genotyping was performed at the Genotyping and Sequencing facility of Bordeaux (grants from the Conseil Régional d'Aquitaine n° 20030304002FA and 20040305003FA and from the European Union, FEDER n° 2003227).

## REFERENCES

- Anderson E, Stebbins J. 1954. Hybridization as an evolutionary stimulus. *Evolution* 8: 378-388.
- Arnold ML 1997. *Natural Hybridization and Evolution*. Oxford University Press, New York
- Axelrod DJ, 1983. Biogeography of oaks in the Arcto-Tertiary Province. *Ann. Missouri Bot. Gard.* 70: 629-657.
- Bacilieri R, Ducouso A, Kremer A, 1995. Genetical, morphological, ecological and phenological differentiation between *Quercus petraea* (Matt.) Liebl. and *Q. robur* L. in a mixed stand of northwest of France, *Silvae genet.* 44: 1-10.
- Bacilieri R, Ducouso A, Petit RJ, Kremer A, 1996. Mating System and Asymmetric Hybridization in a mixed stand of European oaks. *Evolution* 50: 900-908.
- Baudouin L, Piry S, Cornuet JM, 2004. Analytical Bayesian Approach for Assigning Individuals to Populations. *Journal of Heredity* 95: 217-224.
- Belahbib N, Pemonge MH, Ouassou A, Sbay H, Kremer A, Petit RJ. 2001. Frequent cytoplasmic exchanges between oak species that are not closely related: *Quercus suber* and *Q. ilex* in Morocco. *Mol Ecol* 10: 2003-2012.
- Boavida LC, Silva JP, Feijó JA. 2001. Sexual reproduction in the cork oak (*Quercus suber* L). II Crossing intra- and interspecific barriers. *Sexual Plant Reprod* 14: 143-152.
- Borazan A, Babac MT, 2003. Morphometric leaf variation in oaks (*Quercus*) of Bulu, Turkey. *Ann. Bot. Fennici* 40: 233-242.
- Bordacs S, Burg K. 1997. Genetic differentiation by RAPD-markers of oak species in Hungary. In: Steiner, K. C. (ed.) *Diversity and Adaptation in Oak Species*, pp. 121-131. The Pennsylvania State University, Pennsylvania.
- Brewer S, Cheddadi R, Beaulieu JL, Reille M, 2002. The spread of deciduous *Quercus* throughout Europe since the last glacial period. *Forest Ecology and Management* 156: 27-48.
- Bruschi P, Vendramin GG, Bussotti F, Grossoni P. 2000. Morphological and Molecular Differentiation between *Quercus petraea* (Matt.) and *Quercus pubescens* Willd. (Fagaceae) in Northern and Central Italy. *Annals of Botany* 85: 325-333.
- Burger WC, 1975. The Species Concept in *Quercus*. *Taxon* 24 : 45-50.

- Camus A, 1936-54. Le Chénes: Monographie du genre *Quercus*. P. Lechevalier, Paris.
- Congdon P, 2001. Bayesian Statistical Modeling. New York: Wiley. Science 298: 2381-2385.
- Curtu AL, Gailing O, Leinemann L, Finkeldey R. 2007. Genetic Variation and Differentiation Within a Natural Community of Five Oak Species (*Quercus* spp.). *Plant Biology* 9: 116 – 126.
- 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.
- Dupouey JL, Badeau V, 1993. Morphological variability of oaks (*Q. robur* L., *Q. petraea* (Matt) Liebl., *Q. pubescens* Willd.) in northeastern France: preliminary results. *Annales des Sciences Forestières* 50(1): 35-40.
- Ellis JR, Burke JM. 2007. EST-SSRs as a resource for population genetic analyses. *Heredity* 99, 125–132.
- Elsner G, 1993. Preliminary variability of oak stands (*Quercus petraea* and *Quercus robur*) in northern Germany. *Ann. Sci. For.* 50: 228-232.
- Falush D, Stephens M, Pritchard JK, 2003. Inference of population structure II. Linked loci and correlated allele frequencies. *Genetics* 164: 1567-1587.
- Filzmoser P. 2005. Identification of multivariate outliers: a performance study. *Austrian Journal of Statistics* 34(2): 127–138.
- Gellini R, Grossoni P, 1997. *Botanica Forestale* 2: 92-162. CEDAM, Milano.
- González-Rodríguez A, Arias DM, Oyama K. 2004. Morphological and RAPD analysis of hybridization between *Quercus affinis* and *Q. laurina* (Fagaceae), two Mexican red oaks *Amer. J. Bot.* 91(3): 499–509.
- Gschwandtner M, Filzmoser P. 2004. Multivariate outlier detection based on robust methods. *mvoutliers* package for R 2.2.0. URL <http://www.statistik.tuwien.ac.at/public/filz/> (27 Dic. 2006).
- Gugerli F, Walser JC, Dounavi K, Holderegger R, Finkeldey R. 2007. Coincidence of Small-scale Spatial Discontinuities in Leaf Morphology and Nuclear Microsatellite Variation of *Quercus petraea* and *Q. robur* in a Mixed Forest. *Annals of Botany* 99: 713–722.
- Ietswaart JH, Feij AE. 1989. A multivariate analysis of introgression between *Quercus robur* and *Q. petraea* in The Netherlands. *Acta Bot. Neerl.* 38(3) : 313-325.
- Iwata H, Ukai Y. 2002. SHAPE ver. 1.2. A computer program package for quantitative evaluation of biological shape based on elliptic Fourier descriptors. *Journal of Heredity* 93(5): 384-385.
- Jensen JS. 2000. Provenance variation in phenotypic traits in *Quercus robur* and *Quercus petraea* in Danish provenance trials. *Scandinavian Journal of Forest Research* 15: 297–308.
- Kampfer S, Lexer C, Glössl J, Steinkellner H. 1998. Characterization of (GA)(n) microsatellite loci from *Quercus robur*. *Hereditas* 129: 183–186.
- Kashi Y, King DG (2006). Simple sequence repeats as advantageous mutators in evolution. *Trends Genet* 22: 253–259.
- Kleinschmit JRG, Bacilieri R, Kremer A, Roloff A. 1995. Comparison of morphological traits of pedunculate oak (*Q. robur* L.) and sessile oak (*Q. petraea* (Matt.) Liebl., *Silvae Genet.* 44: 256–269.
- Kleinschmit J, Kleinschmit JGR. 2000. *Quercus robur* – *Quercus petraea*: a critical review of the species concept. *Glasnik Za šumske Pokuse* 37: 441– 452.
- Kremer A, Dupouey LJ, Deans JD, Cottrell J, Csaikl U, Finkeldey R, Espinel S, Jensen J, Kleinschmit J, Van Dam B, Ducouso A, Forrest I, De Heredia UL, Lowe AJ, Tutkova M, Munro RC, Steinhoff S, Badeau V. 2002. Morphological variation in mixed oak stands (*Quercus robur* and *Quercus petraea*) is stable western European population). *Ann. For. Sci.* 59: 777–787.

- Kuhl FP, Giardina CR. 1982. Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing* 18: 236–258.
- Lepais O, Leger V, Gerber S. 2006. Short note: High throughput microsatellite genotyping in oak species. *Silvae Genetica*. 2006, 55 (4-5): 238-240.
- Manos PS, Dojle JJ, Nixon KC. 1999. Phylogeny, biogeography and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). *Mol. Phylogenet Evol.* 12: 333-349.
- Mantel NA. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.*, 27:209-220.
- Muir G, Fleming CC, Schlötterer C. 2000. Species status of hybridizing oaks. *Nature*, 405, 1016.
- 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
- Pashley CH, Ellis JR, McCauley DE, Burke JM. 2006. EST databases as a source for molecular markers: Lessons from *Helianthus*. *Journal of Heredity* 97: 381-388.
- Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducouso A, Kremer A. 1997. Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences of the United States of America* 94: 9996-10001.
- Petit RJ, Brewer S, Bordács S, Burg K, Cheddadi R, Coart E, Cottrell J, Csaikl UMA, Van Dam B, Deans JD, Espinel S, Fineschi S, Finkeldey R, Glaz I, Goicoechea PG, Jensen JS, König AO, Lowe AJ, Madsen SF, Mátyás G, Munro RC, Popescu F, Slade D, Tabbener H, De Vries SGM, Ziegenhagen B, De Beaulieu JL, Kremer A. 2002a. Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecol. and Manag.* 156: 49-74.
- Petit RJ, Latouche-Hallé C, Pemonge MH, Kremer A. 2002b. Chloroplast DNA variation of oaks in France and the influence of forest fragmentation on genetic diversity *Forest Ecology and Management* 156: 115–129
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Petit RJ, Bodénès C, Ducouso A, Roussel G, Kremer A. 2004. Hybridization as a mechanism of invasion in oaks. *New Phytologist* 161: 151–164.
- Rohlf FJ, Slice D. 1990. Extension of the Procrustes method for the optimal superimposition of landmark. *Systematic Zoology* 39(1): 40–59.
- Rushton BS, 1983. An analysis of variation of leaf characters in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. population samples from Northern Ireland. *Ir For.* 40:52–77
- Saintagne C, Bodenes C, Barreneche T, Pot D, Plomion C, Kremer A.. 2004. Distribution of genomic regions differentiating oak species assessed by QTL detection. *Heredity* 92: 20–30
- Scotti-Saintagne C, Mariette S, Porth I, Goicoechea PG, Barreneche T, Bodenes C, 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.
- Spellenberg R. 1995. On the hybrid nature of *Quercus basaseachicensis* (Fagaceae, sect. *Quercus*). *Sida* 16(3): 427-434.
- Steinkellner H, Fluch S, Turetschek E, Lexer C, Streiff R, Kremer A, Burg K And Glössl J. 1997. Identification and characterization of (GA/CT)<sub>n</sub> -microsatellite loci from *Quercus petraea*. *Plant Molecular Biology*, 33: 1093-1096.

- Streiff R, Ducouso A, Lexer C, Steinkellner H, Glössl J, Kremer A. 1999. Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. *Molecular Ecology* 8: 831-841.
- Valbuena-Carabaña M, González-Martínez SC, Hardy OJ, Gil L. 2007. Fine-scale spatial genetic structure in mixed oak stands with different levels of hybridization. *Molecular Ecology* 16(6):1207-19.
- Valbuena-Carabana M, González-Martínez SC, Sork VL, Collada C, Soto A, Goicoechea PG, Gil L. 2005. Gene flow and hybridisation in a mixed oak forest (*Quercus pyrenaica* Willd. and *Quercus petraea* (Matts.) Liebl.) in central Spain. *Heredity* 95, 457–465
- Viscosi V, Fortini P, Slice DE, Loy A, Blasi C. 2008. Geometric morphometric analyses of leaf variation in four oak species of subgenus *Quercus* (Fagaceae). *Plant Biosystems* (same volume).
- Williams JH, Williams JB, Howard DJ. 2001. Reproductive processes in two oak (*Quercus*) contact zones with different levels of hybridisation. *Heredity* 87: 680–690.



HortScience 43(5):1371–1376. 2008.

## Characterization and Identification of Tunisian Olive Tree Varieties by Microsatellite Markers

**Imen Rekik**

*Institut de l'Olivier, P.O. Box 1087, 3018 Sfax, Tunisia; and Centre de Biotechnologie de Sfax, P.O. Box 'K', 3038 Sfax, Tunisia*

**Amelia Salimonti**

*Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Centro di Ricerca per l'olivicoltura e l'industria olearia (CRA-OLI), C.da Li Rocchi, 87036 Rende (CS), Italy*

**Naziha Grati Kamoun**

*Institut de l'Olivier, P.O. Box 1087, 3018 Sfax, Tunisia*

**Innocenzo Muzzalupo<sup>1</sup>**

*Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Centro di Ricerca per l'olivicoltura e l'industria olearia (CRA-OLI), C.da Li Rocchi, 87036 Rende (CS), Italy*

**Oliver Lepais and Sophie Gerber**

*Institut National de la Recherche Agronomique (INRA), UMR 1202 Biodiversity Genes and Communities (BIOGECO), 69 route d'Arcachon, F33612 Cestas cedex, France*

**Enzo Perri**

*Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Centro di Ricerca per l'olivicoltura e l'industria olearia (CRA-OLI), C.da Li Rocchi, 87036 Rende (CS), Italy*

**Ahmed Rebai**

*Centre de Biotechnologie de Sfax, P.O. Box 'K', 3038 Sfax, Tunisia*

*Additional index words.* cultivar genotyping, SSR markers, *Olea europaea*, genetic diversity

**Abstract.** In the Mediterranean basin, a large number of olive varieties are present. This poses a series of problems concerning germplasm characterization and management. In addition, there is a problem arising from the existence of homonyms and synonyms. This makes cultivar identification very difficult and complex. Microsatellites or simple sequence repeat (SSR) are locus-specific codominant markers showing a high degree of polymorphism and multiple alleles per locus. Their high informativeness makes them the markers of choice in genetic diversity studies. This work presents the results of molecular characterization and identification of 20 Tunisian olive varieties using 10 SSR markers. All the SSR amplification products were sequenced to determine the number of repeats and the range of allele size. The number of alleles per SSR varied from three to six and the average heterozygosity rate ranged from 30% to 95%. Hierarchical classification of varieties base on similarity measures and clustering was globally consistent with the grouping of varieties by end use and phenotypic characteristics. The result is that varieties having the same name were found to have a clonal relationship. Paternity analysis showed also clone relationships between varieties not known to be related.

Received for publication 30 Jan. 2008. Accepted for publication 23 Apr. 2008.

This research was supported by the Italian Ministry of Agriculture, Food and Forestry Policies and Consiglio per la Ricerca e la Sperimentazione in Agricoltura and by the Tunisian Ministry of Agriculture and the Tunisian Ministry of Higher Education, Research and Technology.

<sup>1</sup>To whom reprint requests should be addressed, e-mail innocenzo.muzzalupo@entecra.it

In Tunisia, olive (*Olea europaea* L.) is cultivated throughout the country. Olive farms cover more than one-third of arable land producing 6.0% of the world's olive production and contributing 45.0% of food export receipts, 4.5% of total exports, and 10.0% of the total agricultural production value. The Tunisian olive grove lands, although dominated by two major varieties, 'Chemlali' in the South and the Center and 'Chetoui' in the North, is rich in cultivars.

Many of these cultivars have been characterized by pomological and chemical traits (Grati Kamoun et al., 2000; Grati Kamoun and Khlif, 2001; Mehri et al., 1997) with isozyme markers (Grati Kamoun et al., 1992, 1999, 2002) and more recently with amplified fragment length polymorphism markers (Grati Kamoun et al., 2006; Taamalli et al., 2006) and simple sequence repeat (SSR) markers (Taamalli et al., 2006, 2007). However, there is still a need for better genetic diversity assessment and varietal identification using high throughput marker technologies such as SSR markers.

Morphological and biometrical characters have been widely used to describe olive germplasm (Grati-Kamoun et al., 1999, 2000; Lombardo et al., 2003, 2004; Ruby, 1918). Recently, biochemical and molecular markers have been used to get better insight into the diversity of olive genetic resources. The first technique to be used was isoenzymes (Grati Kamoun et al., 1992, 2002; Lumaret et al., 2004; Perri et al., 1995; Trujillo and Rallo, 1995), which were shown to be very useful for varietal identification of olive. However, the lack of specificity of isoenzymes and their sensitivity to environmental and developmental factors have limited the widespread use of this technique for routine genotypic identification (Zhang et al., 1999). With the advent of molecular techniques, several types of DNA markers have been used in genetic diversity assessment of olive cultivars. The first technique to be used was randomly amplified polymorphic DNA because of its simplicity and low cost (Besnard et al., 2001; Belaj et al., 2002; Bogani et al., 1994; Bronzini de Caraffa et al., 2002; Fabbri et al., 1995; Muzzalupo et al., 2007a; Perri et al., 2002; Wiesman et al., 1998). The second type of DNA markers to be used was amplified fragment length polymorphism (AFLP) (Angiolillo et al., 1999; Bandelj et al., 2004; Belaj et al., 2003; Grati Kamoun et al., 2006; Montemurro et al., 2005; Sensi et al., 2003). More recently, several microsatellites have been isolated from the olive tree. Currently, ~106 SSR markers are available and have been characterized on many olive cultivars worldwide (Carriero et al., 2002; Diaz et al., 2006; Rallo et al., 2000; Sarri et al., 2006; Sefc et al., 2000). This technique has gained interest and has become the standard approach to study genetic diversity of olive germplasm and is also used for the development of linkage maps (Wu et al., 2004).

SSR markers are easily amplified by polymerase chain reaction (PCR) and are highly reproducible among different laboratories. Almost all reports of cultivar identification using SSRs assessed differences in lengths of amplified alleles (Bandelj et al., 2002, 2004; Belaj et al., 2004; Breton et al., 2006; De La Rosa et al., 2004; Hess et al., 2000; Khadari et al., 2003; Muzzalupo et al., 2006a, 2008a, 2008b; Sajjad et al., 2006; Taamalli et al., 2006, 2007). More recently, it was reported that DNA sequence analysis of microsatellites enhances the





**PHYLOGEOGRAPHY OF NORTH AFRICAN ATLAS CEDAR  
(*CEDRUS ATLANTICA*, PINACEAE): COMBINED MOLECULAR AND  
FOSSIL DATA REVEAL A COMPLEX QUATERNARY HISTORY<sup>1</sup>**

ANASS TERRAB,<sup>2,3,6</sup> ARNDT HAMPE,<sup>4</sup> OLIVIER LEPAIS,<sup>4</sup> SALVADOR TALAVERA,<sup>3</sup>  
ERROL VELA,<sup>5</sup> AND TOD F. STUESSY<sup>2</sup>

<sup>2</sup>Department of Systematic and Evolutionary Botany, Institute of Botany, University of Vienna, Rennweg 14, A-1030 Vienna, Austria; <sup>3</sup>Departamento de Biología Vegetal y Ecología, Facultad de Biología, Universidad de Sevilla, Apartado 1095 41080-Sevilla, Spain; <sup>4</sup>INRA, UMR 1202 BIOGECO, 69, Route d'Arcachon, F-33612 Cestas Cedex, France; and <sup>5</sup>Institut Méditerranéen d'Ecologie et de Paléocologie, Faculté des Sciences de St-Jérôme, Boite 462, Av. Escadrille Normandie-Niemen, F 13397 Marseille Cedex 20, France

Northwest Africa is a major hotspot of plant biodiversity, but very little is known about the Quaternary range dynamics of plant species in this region. Here we investigate the range-wide population structure and phylogeography of Atlas cedar (*Cedrus atlantica*), an emblematic forest tree endemic to Morocco and Algeria. We genotyped 261 individuals from 11 populations using AFLP markers. Data were analyzed using both conventional  $F_{ST}$ -based techniques and Bayesian clustering. Overall population differentiation was high ( $F_{ST} = 0.25$ ). Two major groups of populations were identified, one distributed through the Rif and Middle Atlas mountains in Morocco and the other through the Algerian Tell Atlas and Aurès mountains as well as the Middle Atlas. Combined molecular and fossil data indicate that *C. atlantica* survived the Last Glacial Maximum in at least three disjunct refugia along the coast of the Mediterranean Sea, whereas the Middle Atlas, today the core of the species range, has been colonized relatively recently (<10 000 yr BP). The colonization history of individual populations has left clear imprints in their present-day diversity, which may vary greatly even between nearby stands. Our study illustrates how integrating different data sources and analytical approaches can help elucidate complex range dynamics that would otherwise remain undeciphered.

**Key words:** Algeria; AFLP; diversity; gene flow; glacial refugia; Morocco; nonmetric multidimensional scaling analysis; population divergence; range contraction; Structure program.

Phylogeographical studies have been conducted on thousands of terrestrial, freshwater, and marine organisms worldwide. The geographical coverage of the earth's surface remains highly unbalanced, however, and while some regions have been covered by very comprehensive and detailed accounts, others remain largely unexplored. A particularly sharp regional contrast exists between large parts of Europe and North Africa, which remains poorly explored in contrast to Europe. This imbalance is unfortunate, because North Africa harbors several major hotspots of plant biodiversity and endemism (Médail and Quézel, 1997) and is one of the regions of the world that will suffer the strongest negative effects from modern climate change (in particular a significant increase in aridity; IPCC, 2001). Phylogeographical patterns around the Mediterranean Basin are typically more complex and often span much longer

time-scales than those at higher latitudes (Fady-Welterlen, 2005; Petit et al., 2005a, b; Gómez and Lunt, 2007). Due to the climatic history and heterogeneous topography of the region, phylogeographical surveys typically detect a much greater allelic richness and evolutionary divergence of populations than further north (e.g., Bilton et al., 1998; Fineschi et al., 2002; Hampe et al., 2003; Petit et al., 1993), and they may trace range dynamics dating back well into the Tertiary (e.g., Lumaret et al., 2002; Caujapé-Castells and Jansen, 2003; Magri et al., 2007). In surveys of Mediterranean taxa, North Africa is often underrepresented, and detailed rangewide population surveys of exclusively North African plant species are very scarce (but see El Mousadik and Petit, 1996a; Terrab et al., 2007). Because the fossil record of this region is also fragmentary (Elenga et al., 2000; Magri and Parra, 2002), the past vegetation dynamics in the area and its consequences on gene pools remain poorly known.

Here we investigate the genetic population structure and historical dynamics of Atlas cedar, *Cedrus atlantica* (Endl.) Carrière. This ecologically and economically important forest tree is endemic to the mountain ranges of Morocco and Algeria (being the only African member of the genus). Atlas cedar forests constitute some 2.8% of the total forested area of Morocco, and the major stands occur in two widely separated geographical areas: the North Moroccan Rif mountains (total forest area: 160 km<sup>2</sup>) and the Middle and Eastern High Atlas (1160 km<sup>2</sup>). In Algeria, Atlas cedar forests occur in distant areas and cover around 300 km<sup>2</sup> in the Tell Atlas and Aurès mountains, representing some 1.3% of the total forest surface of the country (Boudy, 1950). In comparison to the estimated original forest cover, the two countries have lost around 75% of their original cedar forests between 1940 and 1982 (Benabid and Fennane,

<sup>1</sup> Manuscript received 11 January 2008; revision accepted 8 July 2008.

The author is grateful to G. Bélair, M. Haddad, A. Mahmoudi, Y. Beghami, N. Yahi, and A. Loukkas for their logistic help during the field trip across Algeria. M. Thimon provided inaccessible and unpublished fossil data. Insightful comments from R. Petit, P. Schönswetter, and two anonymous referees greatly improved a previous manuscript version. This work was funded by postdoctoral fellowships to A.T. (Proyectos de Investigación de Excelencia, Junta de Andalucía, 2005/RNM-484, and Perfeccionamiento de Doctores, Junta de Andalucía) and A.H. (Marie Curie EIF, European Union), as well as grants from the Spanish Ministerio de Educación y Ciencias (REN2002-04634-C05-03 and CGL 2006-00817/BOS to S.T. and REN2002-04354-C02-02 and CGL 2005-0195/BOS to M. Arista) and the Austrian Science Foundation (FWF P13055-BIO).

<sup>6</sup> Author for correspondence (e-mail: anass@us.es).



**Dynamique d'hybridation dans le complexe d'espèces des chênes blancs européens (chênes pédonculés - *Quercus robur* L., sessiles - *Q. petraea* (Matt.) Liebl., pubescents - *Q. pubescens* Willd. et taurins - *Q. pyrenaica* Willd.)**

**Résumé** - L'hybridation est un processus aux conséquences diverses sur l'évolution des espèces qui est difficile à étudier lorsque les espèces se distinguent mal au niveau morphologique. Afin de comprendre le rôle de l'hybridation dans l'évolution du complexe d'espèces des chênes blancs européens, nous avons utilisé des outils de la génétique des populations pour quantifier les flux de gènes interspécifiques contemporains et étudier le système de reproduction de quatre espèces. Un protocole d'analyse génétique rapide a été développé et des méthodes d'assignations génétiques, permettant de déterminer l'espèce de chaque arbre et d'identifier les hybrides, ont été testées par simulations. Cette méthode a été appliquée en populations naturelles révélant un pourcentage d'hybrides variant de 10 à 30% en fonction des populations et impliquant tous les couples d'espèces. Nous avons montré que les effectifs des espèces dans les parcelles influencent la dynamique d'hybridation et la directionnalité de l'introgression. Nous avons étudié le système de reproduction de ces espèces en croisements contrôlés et en forêt pour expliquer le maintien des espèces malgré la présence de flux de gènes interspécifiques. L'existence de plusieurs barrières reproductives contribue à un isolement partiel des espèces qui dépend principalement de barrières pré-reproductives et prézygotiques. Une analyse de paternité pratiquée sur des descendances récoltées en forêt montre que l'hybridation de première génération est rare mais que ces hybrides F1 sont fertiles et se reproduisent principalement avec l'une des espèces parentales, produisant de nombreux rétrocroisements qui expliquent le fort pourcentage d'hybrides observé dans les populations naturelles étudiées. L'hybridation et l'introgression sont donc des processus à l'œuvre chez les chênes qui contribuent à l'évolution du complexe d'espèces.

**Mots clés** : hybridation, introgression, délimitation d'espèces, assignation génétique, analyse de paternité, microsatellites, chênes, *Quercus*.

**Hybridisation dynamics in the European white oak species complex (pedunculate oak - *Quercus robur* L., sessile - *Q. petraea* (Matt.) Liebl., pubescent - *Q. pubescens* Willd. and pyrenean - *Q. pyrenaica* Willd.)**

**Abstract** - Hybridisation is a complex process with diverse consequences on species evolution. Hybridisation is difficult to study when species are not clearly morphologically distinguished. Our aim was to study the role of hybridisation in the evolution of the European white oak species complex. We used population genetic tools to quantify contemporary interspecific gene flow and to study the mating system of four oak species. A fast genetic analysis protocol was developed and genetic assignment methods were first tested by simulation and then used to determine the species of each tree and to identify hybrids. These methods revealed that hybrid percentages were between 10 to 30% depending on the natural population studied and that all species pairs were involved. We showed that the census number of species in the stands had an influence on hybridisation dynamics and on introgression direction. We studied the mating system of these species in controlled crosses and in the forest to understand the maintenance of species despite interspecific gene flow. Several reproductive barriers contribute to a partial isolation of species, mostly pre-reproductive and prezygotic. A paternity analysis of maternal progenies sampled in the forest showed that first generation hybridisation was rare but that F1 hybrids were fertile and were mating mostly with one of the two parental species, creating numerous backcrosses that explain the high percentages of hybrids observed in the natural populations studied. Hybridisation and introgression are active processes in oaks and contribute to the evolution of the species complex.

**Keywords**: hybridisation, introgression, species delineation, genetic assignment, paternity analyses, microsatellites, oaks, *Quercus*.