



**HAL**  
open science

## Distribution des régions génomiques différenciant deux espèces proches : le chêne sessile (*Quercus petraea*) et le chêne pédonculé (*Q. robur*)

Caroline Saintagne

### ► To cite this version:

Caroline Saintagne. Distribution des régions génomiques différenciant deux espèces proches : le chêne sessile (*Quercus petraea*) et le chêne pédonculé (*Q. robur*). Sylviculture, foresterie. Université Henri Poincaré - Nancy 1, 2003. Français. NNT : 2003NAN10059 . tel-01754427

**HAL Id: tel-01754427**

**<https://hal.univ-lorraine.fr/tel-01754427>**

Submitted on 30 Mar 2018

**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.



## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : [ddoc-theses-contact@univ-lorraine.fr](mailto:ddoc-theses-contact@univ-lorraine.fr)

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>



---

FACULTE DES SCIENCES & TECHNIQUES

U.F.R. Sciences et Techniques Biologiques  
Ecole Doctorale : Ingénieries des ressources, Procédés, Produits et Environnement  
Département de Formation Doctorale : Biologie Forestière

## Thèse

Présentée pour l'obtention du diplôme de

Docteur de l'Université Henri Poincaré, NANCY-I  
en Biologie Forestière

Soutenue publiquement le 7 juillet 2003 à Bordeaux

par Caroline SAINTAGNE

Distribution des régions génomiques différenciant deux espèces proches :  
le chêne sessile (*Quercus petraea*) et le chêne pédonculé (*Q. robur*)

Membres du Jury :

Président :	J.M. FAVRE	Professeur, Université H. Poincaré, Nancy-I
Rapporteurs :	J.C. GLASZMANN	Directeur de recherche CIRAD, Montpellier
	X. VEKEMANS	Professeur, Université Lille-1
Examineurs :	J.L. DUPOUEY	Directeur de recherche, INRA, Nancy
	A. KREMER	Directeur de recherche, INRA, Pierroton ( <i>Directeur de thèse</i> )
	T. ROBERT	Enseignant chercheur à l'Université Pierre et Marie Curie

---

INRA- Laboratoire de génétique et Amélioration des Arbres Forestiers  
BP 45, 33 610 Cestas, France

## ✦✦ REMERCIEMENTS ✦✦

Je remercie très vivement Antoine Kremer pour m'avoir tout d'abord proposé ce sujet de thèse, mais surtout pour avoir été toujours disponible pendant ces 3 ans de recherche, malgré ses innombrables responsabilités. En dépit de mon absence à Pierroton les 3 derniers mois, ton dynamisme, ton enthousiasme et ton effervescence sont passés au travers du mail et m'ont donné l'énergie nécessaire pour venir à bout de ce travail. Enfin, un grand merci pour avoir consacré beaucoup de ton temps, en fin de course, aux dernières analyses statistiques et à la relecture du manuscrit.

Merci Catherine d'avoir participé à l'encadrement de cette thèse. Merci pour ton aide au labo, pour tes conseils et pour ton soutien dans les moments difficiles.

Mes remerciements vont ensuite à Jean-Christophe Glaszmann et Xavier Vekemans pour avoir accepté d'être rapporteur de cette thèse.

Je remercie aussi Jean-Luc Dupouey, Jean-Michel Favre et Thierry Robert pour avoir accepté de faire partie du jury.

Cette thèse est le fruit du travail d'un grand nombre de personnes. Tout d'abord merci aux deux personnes qui s'occupent de près de nos petits chênes. Guy, Monsieur croisement contrôlé, qui a, entre autre, réalisé la population de cartographie et Mimi la reine du bouturage à laquelle aucune plante ne résiste. Merci à vous deux pour votre disponibilité lorsqu'il fallait partir faire les mesures à Bourran et merci Mimi pour les heures que tu as passées à mesurer les quelques 4000 feuilles de chênes.

Merci aussi à Alexis et Jean Marc pour avoir participé aux mesures de débourrement et de croissance et merci à Sophie, Teresa, Hakim, Jérémy, Catherine, David pour votre coup de main sur le terrain.

Merci Christophe (Plomion) pour m'avoir initié à l'art de la cartographie génétique, mais aussi pour m'avoir fait part de tes idées et de tes critiques.

Merci David (Pot) pour toutes ces heures passées à me déboguer sous S+, et pour nos discussions très QTLeuses. Mais surtout merci pour ta bonne humeur, ta disponibilité et pour tes cours de planche à voile, même s'ils peuvent parfois devenir très dangereux...

Merci Teresa, Agnès, Grégoire, Céline Lalanne pour votre aide au labo et merci Pauline pour tes conseils et ton aide dans les analyses.

Merci Manuela pour ton soutien en fin de course, j'espère que l'on continuera à travailler ensemble par la suite.

Merci David (Chagné) pour ta bonne humeur constante, tes cours de rollers et tes couleurs fleuries qui égailent le labo.

Merci Marie et Jean Marc pour avoir été là dans les moments difficiles, et pour les bons moments passés ensemble.

Un grand merci à tous les thésards et stagiaires de passage qui rendent le labo si dynamique. Céline Latouche, Carlos, Jérémy, Jérôme, Hakim, Catherine, Laetitia, Jorge, Antoine (Leveau), Guylaine, Gabriel, Olivier, Marie-Hélène, Marta, Josquin, Patricia, Manu, Alexandre.

Enfin, merci à Jeannot et Erminia pour m'avoir gentiment hébergé en fin de course.

Mes remerciements vont maintenant pour ma petite famille Anduzienne. Merci Maman, pour la joie que tu sèmes autour de toi, merci d'avoir toujours été présente lorsque le moral était au plus bas. Merci à Rémi, Marie, Jean et Louise. Et merci à Anne Lise, Claire, Coralie, Mathilde, Julia, Julie, Laurence, Florian, Jérôme, Jérémy, Pierre, Renaud, Yvan, mes amis de toujours.

Mes remerciements vont enfin à ma petite famille Montmeyrannaise. Merci Papa et Faby pour vos mails qui m'ont souvent remonté le moral. Merci Faby pour le temps que tu as passé à relire l'ensemble du manuscrit. Merci à mes sœurs et frère : Lisa, Nelly et Tom pour avoir aussi aidé à la relecture du manuscrit.

Infine grazie a te Ivan. Grazie di essere venuto a Pierroton durante quello straordinario mese di giugno 2002 e di avermi portata via sul tuo bel cavallo blu. La stesura di questa tesi resterà un meraviglioso ricordo grazie a te, il ricordo delle lunghe discussioni scientifiche, così come quello dei mazzi di fiori di prato. Questa tesi è la fine dei miei studi ma soprattutto la porta del nostro avvenire.



S.C.D. - U.H.P. NANCY 1  
BIBLIOTHEQUE DES SCIENCES  
Rue du Jardin Botanique - BP 11  
54601 VILLERS-LES-NANCY Cédex

## TABLE DES MATIERES

<b>REMERCIEMENTS</b>	<b>2</b>
<b>PREAMBULE</b>	<b>7</b>
<b>CHAPITRE A SYNTHÈSE BIBLIOGRAPHIQUE</b>	<b>10</b>
<b>A.1 ESPECE ET SPECIATION</b>	<b>10</b>
A.1.1 <i>Notion d'espèce</i>	10
A.1.1.1 Introduction	10
A.1.1.2 Principaux concepts d'espèce	10
A.1.2 <i>Les modes de spéciation</i>	14
A.1.2.1 La spéciation allopatrique	14
A.1.2.2 La spéciation sympatrique	15
A.1.2.3 La spéciation parapatricque	17
A.1.2.4 La spéciation par réarrangement chromosomique	17
A.1.2.5 La spéciation par changement du système de reproduction	19
A.1.2.6 La spéciation par dérive génétique	19
A.1.2.7 La spéciation par sélection naturelle	20
A.1.2.8 Conclusion du chapitre	22
<b>A.2 GENETIQUE DE LA DIFFERENCIATION INTERSPECIFIQUE</b>	<b>23</b>
A.2.1 <i>Historique</i>	23
A.2.2 <i>Architecture génétique de la différenciation interspécifique</i>	24
A.2.2.1 Nombre de facteurs impliqués dans la différenciation interspécifique	25
A.2.2.2 Effet des facteurs impliqués dans la différenciation interspécifique	25
A.2.2.3 Interactions entre les gènes impliqués dans la différenciation interspécifique	28
A.2.3 <i>Mécanismes conduisant à la différenciation interspécifique</i>	30
A.2.3.1 Facteurs agissant sur la diminution de la recombinaison au niveau des gènes de spéciation	30
A.2.3.2 Forces évolutives impliquées dans la différenciation interspécifique	32
A.2.3.3 Conclusion du chapitre	34
<b>A.3 PRESENTATION DU CHENE</b>	<b>35</b>
A.3.1 <i>Le genre Quercus</i>	35
A.3.1.1 Taxonomie	35
A.3.1.2 Histoire des chênes	35
A.3.2 <i>Caractéristiques des chênes sessiles et pédonculés</i>	36
A.3.2.1 Différenciation phénotypique	36
A.3.2.2 Hybridation	38
A.3.2.3 Différenciation moléculaire	39
A.3.2.4 Où se situent les chênes sessile et pédonculé dans le processus de spéciation ?	40
<b>CHAPITRE B PROBLEMATIQUE</b>	<b>44</b>
<b>B.1 INTRODUCTION ET RAPPELS</b>	<b>44</b>
<b>B.2 STRATEGIES DEVELOPPEES</b>	<b>45</b>
B.2.1 <i>La stratégie descendante : du phénotype au génome</i>	45
B.2.2 <i>La stratégie ascendante : du génome aux régions discriminantes</i>	45

<b>C.1 MATERIEL ET METHODES</b>	<b>49</b>
C.1.1 <i>Construction d'une carte génétique</i>	49
C.1.1.1 Principe	49
C.1.1.2 Cas particulier des arbres forestiers	50
C.1.1.3 Caractéristiques de la carte génétique du chêne pédonculé	51
C.1.1.4 Construction de la carte jalon du chêne pédonculé, support de la détection de QTL.	51
C.1.2 <i>Détection de QTL</i>	52
C.1.2.1 Principales méthodes de détection de QTL.	52
C.1.2.2 Cas particulier des arbres forestiers	54
C.1.2.3 Détection de QTL chez le chêne pédonculé	55
C.1.3 <i>Distribution des QTL le long des groupes de liaison</i>	64
C.1.3.1 Hypothèse de distribution totalement aléatoire (Processus poissonien)	64
C.1.3.2 Comparaison de la distribution observée à l'Hypothèse nulle de distribution aléatoire	64
C.1.3.3 Colocalisations entre les QTL détectés	65
<b>C.2 RESULTATS</b>	<b>66</b>
C.2.1 <i>Détection de QTL impliqués dans les caractères qui différencient le chêne sessile et le chêne pédonculé</i>	66
C.2.1.1 QTL impliqués dans la morphologie foliaire	66
C.2.1.2 QTL impliqués dans l'efficacité d'utilisation de l'eau	72
C.2.1.3 QTL impliqués dans la croissance	75
C.2.1.4 Conclusion	78
C.2.2 <i>Détection de QTL impliqués dans les caractères qui ne différencient pas le chêne sessile et le chêne pédonculé</i>	79
C.2.2.1 QTL impliqués dans la phénologie du débourrement	79
C.2.2.2 QTL impliqués dans l'aptitude au bouturage	86
C.2.3 <i>Comparaison des QTL de caractères discriminants et non discriminants.</i>	88
C.2.3.1 Nombre de QTL détectés et distribution de la variance expliquée par les QTL	89
C.2.3.2 Répartition spatiale dans le génome	90
C.2.4 <i>Conclusion générale</i>	91

**CHAPITRE D DISTRIBUTION DES INDICES DE DIFFERENCIATION GENETIQUE SUR LE GENOME**

<b>D.1 MATERIEL ET METHODES</b>	<b>94</b>
D.1.1 <i>Echantillonnage des populations naturelles</i>	94
D.1.2 <i>Types de marqueurs utilisés pour l'évaluation de la différenciation génétique</i>	95
D.1.2.1 Marqueurs isozymes	96
D.1.2.2 Marqueurs RAPD transformés en SCAR	96
D.1.2.3 Marqueurs microsatellites	97
D.1.2.4 Marqueurs SNP	98
D.1.2.5 Marqueurs AFLP	98
D.1.3 <i>Estimation de la différenciation génétique interspécifique</i>	100
D.1.3.1 Indices de différenciation : $G_{ST}$ de Nei ; $F_{ST}$ de Weir et Cockerham	100
D.1.3.2 Tests des indices de différenciation	104
D.1.3.3 Comparaison du niveau de différenciation estimé avec différents types de marqueurs	105
D.1.4 <i>Distribution des valeurs de <math>G_{ST}</math></i>	106
D.1.5 <i>Construction de la carte génétique et positionnement des QTL</i>	107

D.1.6	<i>Structuration le long des groupes de liaison des valeurs de différenciation interspécifique</i>	108
D.1.6.1	Comparaison à la distribution aléatoire	109
D.1.6.2	Analyse de la distribution spatiale par autocorrélation.	110
D.1.7	<i>Associations entre marqueurs et caractères quantitatifs</i>	111
<b>D.2</b>	<b>RESULTATS</b>	<b>112</b>
D.2.1	<i>Distribution des valeurs de <math>G_{ST}</math> selon les marqueurs</i>	112
D.2.2	<i>Construction de la carte génétique consensus</i>	113
D.2.3	<i>Distribution des <math>G_{ST}</math> le long des groupes de liaison</i>	115
D.2.4	<i>Etude d'association entre marqueurs et caractères</i>	116
<b>CHAPITRE E COLOCALISATIONS ENTRE <math>G_{ST}</math> ET QTL</b>		<b>118</b>
E.1.1	<i>Colocalisations entre QTL et marqueurs impliqués dans la différenciation génétique interspécifique</i>	119
<b>CHAPITRE F CONCLUSION GENERALE</b>		<b>121</b>
F.1	<b>RAPPELS DES PRINCIPAUX RESULTATS</b>	<b>121</b>
F.2	<b>PROBLEMES RENCONTRES ET REPOSES PROPOSEES</b>	<b>123</b>
F.2.1	<i>Puissance de détection des QTL</i>	123
F.2.2	<i>Recherche de nouveaux QTL et validation des QTL déjà détectés</i>	123
F.2.3	<i>L'approche gènes candidats : une nouvelle strategie pour détecter des régions discriminantes</i>	124
F.2.4	<i>Augmenter la densité des <math>G_{ST}</math> sur la carte génétique</i>	125
F.2.5	<i>Recherche de gènes létaux</i>	126
F.2.6	<i>Le paradoxe différenciation moléculaire/ différenciation phénotypique des deux espèces</i>	126
<b>BIBLIOGRAPHIE</b>		<b>129</b>
<b>ANNEXE</b>		<b>143</b>

**S.C.D. - U.H.P. NANCY 1**  
**BIBLIOTHÈQUE DES SCIENCES**  
Rue du Jardin Botanique - BP 11  
54601 VILLERS-LES-NANCY Cédex

# PREAMBULE

La notion d'espèce constitue un lieu commun de la biologie depuis que l'homme a cherché à décrire et à classifier le vivant. Le débat autour de la définition de l'espèce mais aussi de la spéciation a connu un regain dans l'actualité de la biologie évolutive. En effet deux numéros spéciaux de revues en Biologie Evolutive ont été consacrés à ce débat en 2001 (*Trends in Ecology & Evolution* et *Journal of Evolutionary Biology*), en réponse aux perspectives que laissaient présager les nouveaux acquis en phylogénie et en génomique. Au centre des préoccupations se dégage aujourd'hui la définition de l'espèce au niveau du génome. Deux articles tirés de ces revues illustrent ces nouvelles orientations. Orr, (2001) en faisant un bilan des QTL impliqués dans des caractères morphologiques ou comportementaux qui différencient des espèces proches, s'intéresse à la distribution, dans le génome, des régions impliquées dans la différenciation interspécifique. De même, en considérant la spéciation avant tout comme une différenciation adaptative par évolution de quelques gènes, Wu (2001) définit l'espèce en fonction de ses gènes. Il remet alors en cause le concept biologique de l'espèce, dans lequel l'espèce résulte d'abord de l'isolement reproductif (Mayr, 1963) et dont les hypothèses sous-jacentes sont d'une part, l'implication d'un grand nombre de gènes dans la différenciation interspécifique et d'autre part, que le génome entier est une unité génétique coadaptée. Selon cette définition, l'échange de gènes entre deux groupes divergents détruirait leur intégrité. Par opposition, la vision Néo-Darwinienne de la spéciation met en avant la différenciation adaptative. L'isolement reproductif est alors vu comme une des "résultantes" de l'adaptation différentielle mais n'est plus considéré comme le processus déclenchant la spéciation. Ces deux articles illustrent aussi les deux volets du débat autour de la notion d'espèce. D'une part, le volet purement descriptif qui s'attache à identifier et mesurer les différences interspécifiques et d'autre part, le volet plus explicatif cherchant à comprendre les processus qui conduisent à la spéciation des espèces.

Les chênes sont un cas classique de groupe taxonomique pour lequel le concept biologique des espèces est discuté de longue date (voir Kleinschmit et Kleinschmit, 2000 pour une revue récente). En Europe, ces débats se sont principalement focalisés autour des deux "espèces"

majoritaires de chênes : le chêne pédonculé (*Quercus robur*) et le chêne sessile (*Quercus petraea*), qui sont sympatriques et inter-fertiles. Mais il ne s'agit pas là d'un cas particulier : la définition de l'espèce est un lieu commun au sein du genre sur les trois continents où le genre *Quercus* est représenté. Les causes du débat sont le maintien de différences écologiques, morphologiques et écophysiologiques, en dépit d'un fort flux de gènes entre les deux espèces, associées à une très faible différenciation moléculaire, guère plus élevée que la diversité intraspécifique. Cette réalité soulève deux types de questions : quelle est l'origine génétique et moléculaire des différences phénotypiques observées ? Comment peuvent se maintenir les deux espèces alors qu'elles s'hybrident facilement ? Dans cette thèse nous allons tenter de répondre à la première question et nous rejoignons de ce fait le débat actuel évoqué au tout début de cette introduction sur l'organisation moléculaire des différences interspécifiques. Comment sont distribuées dans le génome les régions qui différencient le chêne sessile et le chêne pédonculé ?

Dans le premier chapitre, une synthèse bibliographique a été réalisée qui regroupe les différentes notions de l'espèce ainsi que les différents mécanismes qui mènent à la spéciation, en se focalisant tout particulièrement sur la notion de différenciation génétique interspécifique. Ce premier chapitre comprend également une synthèse sur les connaissances en génétique du chêne sessile et du chêne pédonculé. Seront tout d'abord précisés, les caractères phénotypiques qui les distinguent, seront ensuite présentés les principaux résultats obtenus en terme d'hybridation interspécifique pour finir sur les résultats obtenus à ce jour en terme de différenciation génétique interspécifique. Nous verrons entre-autre, que ces études n'ont pas permis de répondre à la question fondamentale de l'étude de la différenciation génétique interspécifique : comment expliquer au niveau génétique le maintien des différences phénotypiques en dépit d'un flux de gènes important ? Est ce que, par exemple, ces différences sont le résultat de l'effet de plusieurs gènes à effets mineurs ou de quelques gènes à effets majeurs ? La distribution de la différenciation génétique interspécifique est elle concentrée dans des zones du génome ou plutôt dispersée ?

Pour apporter des réponses à ces questions nous avons adopté deux stratégies différentes mais complémentaires qui sont présentées dans le chapitre B. La première approche, détaillée dans le chapitre C, consiste à rechercher dans le génome du chêne, des régions qui sont en déséquilibre de liaison physique avec des caractères qui différencient les deux espèces. Cette approche consistera donc à rechercher des QTL (Quantitative Trait Loci). La deuxième approche, détaillée dans le chapitre D, va consister à prospector le génome de manière systématique à l'aide de différentes techniques moléculaires et d'estimer pour chacun de ces marqueurs un coefficient de différenciation interspécifique ( $G_{ST}$ ). Le but final, détaillé dans le chapitre E, a enfin été d'étudier les colocalisations sur le génome des QTL et des marqueurs impliqués dans la différenciation interspécifique. Enfin, les principaux résultats de la thèse sont rappelés en conclusion en appuyant sur les problèmes rencontrés et les solutions proposées. L'originalité de ce travail repose donc dans l'utilisation conjointe d'outils empruntés à la génétique quantitative, à la cartographie génétique et à la génétique des populations.

## Chapitre A

# SYNTHESE BIBLIOGRAPHIQUE

## CHAPITRE A SYNTHÈSE BIBLIOGRAPHIQUE

### A.1 ESPECE ET SPECIATION

#### A.1.1 NOTION D'ESPECE

##### A.1.1.1 Introduction

Jusqu'au début du XVIIIème siècle, la vision du monde était statique. La création avait eu lieu qu'une seule fois et les êtres restaient semblables à leurs premiers ancêtres. Le concept d'espèce était alors typologique et morphologique : tous les membres d'une même espèce avaient une caractéristique commune qui ne variait pas. Au XIX siècle un nouveau courant de pensées émergea avec Darwin (1809-1882). L'espèce n'était plus considérée comme fixée mais au contraire une entité qui évolue en continu dans l'espace et dans le temps. Cette nouvelle vision a, par la suite et jusqu'à notre époque, suscité des débats très vifs sur le concept même d'une espèce comme en témoigne la mise au point récente de Mayden (1997) qui évoque jusqu'à 24 définitions différentes de l'espèce.

Je n'aurai pas la possibilité ni la prétention de vous décrire chacun de ces concepts mais il me semble intéressant de donner la définition des plus communs, qui permettra de discuter par la suite la position du chêne sessile et du chêne pédonculé.

##### A.1.1.2 Principaux concepts d'espèce

###### A.1.1.2.1 Le concept d'espèce Typologique (Linné)

L'espèce est définie comme un groupe d'individus qui diffère des autres groupes au niveau de caractères diagnostiques constants (critères anatomiques et physiologiques). C'est sur ces critères que le Suédois Carl Von Linne (1707-1778) élaborera la nomenclature binomiale (genre - espèce) qui

**Tableau 1** : Différents types d'isolements reproductifs.

Isolement pré-zygotique	Isolement gamétique	Incompatibilité sperme /ovule
	Isolement temporel	Période de reproduction décalée
	Isolement comportementaux	Différentes stratégies pour attirer les partenaires : cours, parade, émission de signaux
	Isolement morphologique	Incompatibilité des organes reproducteurs
	Isolement géographique	Barrière physique
	Isolement de l'habitat	Niches écologiques différentes
Isolement post-zygotique	Mortalité des hybrides	Des incompatibilités génétiques peuvent causer la mort de l'embryon
	Stérilité des hybrides	Les descendants sont vigoureux mais ne peuvent pas se reproduire
	Déchéance des hybrides (Hybrid breakdown)	Des hybrides vigoureux s'accouplent pour produire une progéniture faible et stérile

permet encore aujourd'hui le classement de tous les végétaux connus. Ce concept présente néanmoins certains problèmes de classification lorsque des phénotypes intermédiaires apparaissent.

#### **A.1.1.2.2 Le concept d'espèces évolutives (Simpson, 1951)**

Dans ce concept, développé pour l'étude d'espèces fossiles, une espèce regroupe les individus qui partagent une histoire évolutive commune. Ce concept a l'avantage d'être applicable sur les êtres vivants ou fossiles ainsi que sur des organismes sexuels et asexués. Néanmoins il est limité par la difficulté d'appréhender ce qui constitue une histoire évolutive commune. Il faut définir un trait pertinent pour catégoriser l'espèce (souvent morphologique), tout en tenant compte de sa variabilité. Se pose alors encore le problème de définir les limites de la variation phénotypique autorisée au sein d'une même espèce.

#### **A.1.1.2.3 Le concept biologique des espèces (Dobzhansky, 1937; Mayr, 1942)**

C'est le concept des espèces qui domine dans la littérature évolutionniste. L'espèce biologique est définie comme un groupe d'individus ou de populations d'individus qui se reproduisent entre eux et occupent une niche écologique spécifique. C'est donc la capacité de se reproduire qui détermine les limites de l'espèce, et non la présence ou l'absence de certaines caractéristiques morphologiques ou physiologiques communes. Les barrières d'isolement qui peuvent être pré- ou post-zygotiques (Tableau 1) définissent les frontières de la communauté reproductrice et préservent l'intégrité du système génétique vis à vis des autres espèces. Bien que cette définition ait l'avantage de fournir un critère simple pour caractériser l'appartenance de deux individus à des espèces différentes, elle présente cependant certaines limites. Comment classer, par exemple, les espèces jumelles interstériles ou les espèces fossiles ? et, que faire des groupes de populations isolées qui paraissent différentes mais qui peuvent s'hybrider en cas de contact ? Un autre problème du concept

biologique des espèces est de considérer uniquement le produit de la spéciation (l'isolement reproductif) sans prendre en compte les mécanismes qui en sont à l'origine.

#### **A.1.1.2.4 Le concept d'espèce reconnue (Paterson 1985)**

Paterson définit l'espèce comme un groupe d'individus qui partagent un système de fertilisation commun (comportement pré-copulatoire, structure génitale, attractivité ovule/sperme). Ce concept s'oppose à celui de l'espèce biologique dans le sens où tous les mécanismes pour empêcher la reproduction interspécifique (barrière reproductrice) peuvent être vus comme des mécanismes élaborés pour faciliter la reproduction intraspécifique. Ce concept met l'accent sur les propriétés internes de l'espèce plutôt que sur les différences entre les espèces. Cette définition est cependant limitée car principalement basée sur des mécanismes de fertilisation. Elle ne peut être appliquée à des individus asexués et la présence d'hybrides crée les mêmes difficultés que pour le concept biologique des espèces.

#### **A.1.1.2.5 Le concept de cohésion des espèces (Templeton, 1989)**

A la différence du concept biologique des espèces, le concept de cohésion des espèces ne se limite pas à la reproduction mais inclut des facteurs génétiques, écologiques et historiques. Une espèce est définie comme un groupe d'organismes ayant la capacité d'échanges génétiques ou démographiques tout en maintenant leur identité. Cette définition reste néanmoins difficile à appliquer car compliquée et peu objective.

#### **A.1.1.2.6 Le concept phylogénique des espèces (Cracraft, 1983)**

L'espèce est définie comme un groupe monophylétique c'est à dire un groupe composé de l'ensemble des descendants d'un ancêtre commun. Ce concept se focalise sur l'histoire phylogénétique des populations et utilise la phylogénie moléculaire pour définir des groupes

monophylétiques distincts. Le problème reste de définir les caractères à utiliser pour la construction des arbres phylogénétiques. Doit-on identifier les espèces sur la base de mutations neutres, ou plutôt utiliser des gènes reflétant l'action de la sélection ? et comment définir les niveaux de divergence qui constituent une espèce ?

#### **A.1.1.2.7 Le concept écologique des espèces (Van Valen, 1976)**

Le concept écologique des espèces regroupe des individus ou populations qui partagent la même niche écologique et qui évoluent indépendamment des autres groupes situés à l'extérieur de la zone. Ce concept met l'accent sur l'adaptation des espèces vis à vis de leur l'environnement. Les espèces écologiques sont le produit de la spéciation écologique que nous décrirons plus loin.

#### **A.1.1.2.8 Le concept du complexe d'espèce (Pernès, 1984)**

Deux plantes appartiennent au même complexe si dans les conditions naturelles elles peuvent, avec une probabilité non nulle, échanger des gènes par hybridation, soit directement, soit par le relais de plantes intermédiaires. L'existence de plantes relais permet d'intégrer dans le même complexe, des plantes dont l'éloignement géographique empêche l'inter-pollinisation immédiate. Deux plantes d'un même complexe appartiennent à des compartiments différents s'il existe entre elles des limitations à la réussite de leur hybridation spontanée. Par exemple, deux populations mendéliennes partiellement isolées géographiquement entre lesquelles quelques échanges peuvent avoir lieu par dispersion lointaine de graines ou de pollen, constituent deux compartiments d'un complexe d'espèce.

Il est important de ne pas confondre la notion d'espèce qui est un aboutissement, avec la spéciation qui est un mécanisme. Ainsi, du point de vue du concept biologique des espèces, la spéciation est le mécanisme qui mène à l'isolement reproducteur. Nous allons voir dans le chapitre qui suit que similairement à la notion d'espèce qui

regroupe un grand nombre de concepts, plusieurs chemins peuvent mener à l'isolement reproductif.

## **A.1.2 LES MODES DE SPECIATION**

On postule généralement l'existence de trois phases dans le processus de spéciation qui peuvent se dérouler simultanément ou de manière séquentielle : l'arrêt des échanges génétiques, l'évolution des mécanismes d'isolement reproductif, l'accumulation de différences génétiques comportementales et morphologiques entre les futures espèces.

Traditionnellement, les modes de spéciation ont été classés par rapport aux conditions dans lesquelles l'isolement reproductif s'est produit. Ainsi, ils font référence à la répartition géographique des populations lors du processus de spéciation.

### **A.1.2.1 La spéciation allopatrique**

C'est le modèle de base de la spéciation décrit par Dobzhansky et Muller dans les années 1930-1940 et qui est resté le modèle dominant la vision de la spéciation de ces 60 dernières années.

Dans la spéciation allopatrique, des événements extérieurs (changements climatiques brutaux, formation des continents, ou à plus petite échelle la formation d'un lac) vont isoler géographiquement des individus appartenant à la même population mère. L'isolement géographique provoque une diminution ou un arrêt des échanges génétiques et donc l'accumulation progressive de différences entre les populations. En cas de contact secondaire (sympatrie secondaire), si le flux de gènes peut reprendre, il homogénéisera les différences naissantes et il n'y aura pas formation de nouvelles espèces. Au contraire, si la différenciation entre les deux populations a été assez forte pour modifier leur comportement sexuel, alors l'hybridation n'est plus possible, les deux populations sont isolées reproductivement, on peut parler d'espèces biologiques. Une autre alternative possible est que l'isolement reproductif n'est pas complet lors du contact secondaire

mais suffisamment avancé pour être complété par la fixation d'allèles défavorables au croisement interspécifique (Coyne et Orr, 1998). Ce mécanisme se nomme : renforcement.

Dans le modèle allopatrique, deux mécanismes principaux peuvent engendrer de la différenciation génétique et secondairement un isolement reproductif : la dérive génétique et /ou la sélection en réponse aux nouveaux environnements (Coyne, 1992 ; Rice et Hostert, 1993). Dans un article qui résume les expériences menées depuis 40 ans en laboratoire sur la spéciation, Rice et Hostert (1993) suggèrent qu'en conditions d'allopatrie, l'isolement pré-zygotique est essentiellement dépendant de l'environnement. Il serait un sous-produit et le résultat de la sélection. Ils fondent leur argumentation sur des expériences menées sur la drosophile (Killias, 1980 ; Dobb, 1989). Le résultat est moins clair pour l'isolement post-reproductif (diminution de la vigueur des hybrides, ou stérilité hybride) qui pourrait être le résultat de l'action de la sélection et/ou de la dérive génétique.

#### **A.1.2.2 La spéciation sympatrique**

Bien que la plupart des Néo-Darwiniens pensent que la spéciation est conditionnée par un isolement reproductif total engendré par l'isolement géographique (Mayr, 1963 ; Futuyma et Mayer, 1980), d'autres suggèrent que des populations qui partagent le même environnement (en sympatrie) peuvent se différencier et mener à la mise en place de nouvelles espèces sous l'action de la sélection naturelle (Via, 2001). Ainsi dans la spéciation sympatrique, la principale force en action serait la sélection disruptive (ou divergente) qui agit dans la population de façon opposée sur des caractères identiques (compétition pour les ressources, prédation, mutualisme). Elle a pu être mise en évidence par simulation (Dieckmann et Doebeli, 1999 ; Kondrashov et Kondrashov, 1999), et en conditions expérimentales contrôlées (Rice et Hostert, 1993 ; Rundle *et al.*, 2000) mais son étude en milieu naturel donne souvent des résultats ambigus. En effet, il est très difficile de prouver que la spéciation s'est réalisée exclusivement par sympatrie, il est toujours possible d'argumenter un scénario d'allopatrie suivi par la suite d'un contact secondaire. La reconstruction de l'histoire phylogénétique des taxons divergents peut être un moyen de mettre en évidence la spéciation

sympatrique (Via, 2001). C'est par ce biais que Schlieven *et al.* (2001) ont mis en évidence dans un lac du Cameroun, la spéciation sympatrique sans équivoque de l'espèce de poisson *Tilapia cf. deckerti*.

Néanmoins la question relative aux mécanismes génétiques qui permettent à des espèces de s'isoler reproductivement en présence de flux de gènes reste ouverte.

Dans un premier modèle, ce sont des gènes différents qui contrôlent les caractères soumis à la sélection disruptive (préférence de l'habitat) et les caractères impliqués dans l'assortative mating (choix du conjoint sur la base de son phénotype) (Maynard Smith, 1966). L'isolement reproductif est alors dépendant du déséquilibre de liaison qui se met en place entre les deux caractères au cours des générations successives. Le maintien du déséquilibre de liaison est fortement influencé par la fitness des hybrides d'une part, (une faible fitness favorise la sélection des déséquilibres de liaisons) et par la recombinaison d'autre part, qui va homogénéiser les effets de la sélection disruptive en mélangeant les gènes au cours de la méiose.

Un deuxième modèle beaucoup mieux accepté est celui où des gènes identiques (ou fortement liés) contrôlent les caractères adaptatifs et les caractères relatifs au choix du conjoint (pléiotropie). Dans ce modèle, l'isolement reproductif peut évoluer en sympatrie aussi bien par l'action directe de la sélection disruptive sur le choix de l'habitat que par l'action indirecte de la sélection disruptive sur d'autres caractères par effet pleiotropique (un même gène impliqué dans plusieurs caractères) ou d'entraînement (gènes fortement liés). Dans ce modèle le problème de l'antagonisme sélection - recombinaison est contourné et la seule force homogénéisatrice qui intervient est la migration par flux de gènes (Rice et Hostert, 1993). Enfin, un dernier modèle serait un contrôle polygénique des caractères adaptatifs et des caractères relatifs au croisement (Kondrashov et Kondrashov, 1999 ; Dieckmann et Doebeli, 1999). Si beaucoup de loci sont impliqués alors la recombinaison entre paires d'allèles à chaque génération ne cassera pas les associations entre les caractères.

### **A.1.2.3 La spéciation parapatrique**

Dans ce modèle, la séparation géographique n'est pas complète, deux populations partagent une même frontière perméable aux flux de gènes. La spéciation parapatrique se met en place selon le modèle de Endler (1977) lorsque des différences géographiques apparaissent entre les populations suite à la propagation d'une espèce ancestrale. Cette différenciation géographique mène à la formation d'un cline (gradation continue de caractères mesurables) souvent associé à un gradient environnemental comme le climat, ou l'altitude, qui agit comme une barrière aux flux de gènes, permettant ainsi la mise en place de divergence entre les populations situées de chaque côté du cline. La sélection est alors directionnelle (sélection pour un phénotype particulier) à l'intérieur de chaque sous-population et disruptive si les deux populations sont considérées ensemble. Depuis le centre du cline, un changement progressif de la sélection s'opère le long du gradient géographique et des phénotypes opposés sont favorisés à chaque extrême. La réduction du flux de gènes à travers la zone hybride va dépendre comme pour la spéciation sympatrique de la distance de dispersion des gènes, et de la force de sélection contre les hybrides.

Les modes de spéciation ont été classés historiquement, comme nous venons de le voir, selon la répartition géographique des populations en cours de spéciation. Ce classement met l'accent sur l'effet antagoniste du flux de gènes sur l'isolement reproductif. Un autre moyen de classer les modes de spéciation est de se concentrer sur les mécanismes à l'origine de l'évolution de l'isolement reproductif.

### **A.1.2.4 La spéciation par réarrangement chromosomique**

Les réarrangements chromosomiques provoquent presque toujours une fécondité réduite des hétérozygotes (sélection disruptive/divergente) due à des problèmes de disjonction et de ségrégation des chromosomes à la méiose. Il s'agit de barrières post-zygotiques à la reproduction.

On peut citer deux types de spéciation par réarrangements chromosomiques :

#### **A.1.2.4.1 La fixation d'une nouvelle combinaison chromosomique**

La propagation d'une nouvelle formule chromosomique, apparue dans un individu par mutation, ne pourra se propager dans une grande population qu'à la condition qu'elle dépasse une certaine fréquence d'équilibre dépendant des fitness relatives des caryotypes des hybrides (Ridley, 1993). La seule manière de pouvoir dépasser cette fréquence passe par un brusque bouleversement des fréquences alléliques dans la population, soit par dérive génétique soit par effet fondateur. Cela implique que la spéciation par des barrières chromosomiques n'est envisageable que dans une population de très faible effectif fortement consanguine. Une allopatrie prolongée favorisera ce processus et permettra l'accumulation progressive de plusieurs différences chromosomiques jusqu'à une interstérilité complète entre les espèces.

#### **A.1.2.4.2 La spéciation par polyploïdisation**

Chez les plantes, la polyploïdisation entraîne presque toujours une spéciation instantanée, et ceci même en cas de sympatrie. Le processus initial est l'hybridation entre des populations qui diffèrent dans leur composition chromosomique. La stérilité des hybrides causée par un mauvais appariement des chromosomes à la méiose, est contournée en doublant le nombre de chromosomes par autofécondation. Ce phénomène permet ainsi de produire des hybrides polyploïdes fertiles mais isolés reproductivement de leurs géniteurs. La polyploïdisation représente entre 2 et 4 % des évènements de spéciation chez les plantes à fleurs mais reste rare chez les animaux où elle est souvent associée à la parthénogenèse (Otto *et al.*, 2000). On peut citer chez les plantes l'exemple du blé : le blé dur utilisé pour "la pasta" est tétraploïde alors que le blé tendre utilisé pour le pain est hexaploïde. Dans le règne animal, un fort niveau de polyploïdisation a été observé chez le homard, l'écrevisse, le crabe ermite par contre il est très rare chez les mammifères.

### **A.1.2.5 La spéciation par changement du système de reproduction**

La plupart des plantes présentent un système d'auto-incompatibilité qui empêche le pollen de fertiliser l'ovule d'un même individu. Ce mécanisme est contrôlé par plusieurs allèles au niveau d'un locus d'auto-incompatibilité (Locus S) très polymorphe à l'intérieur de certaines espèces (11 allèles mis en évidence dans 3 espèces de la famille des Solanacées, Ioerger *et al.*, 1990). Le développement du tube pollinique est empêché lorsque le génotype au locus S du pollen est identique à celui du stigmate. Un changement dans le mode de reproduction, par exemple passer de l'auto-incompatibilité (allogamie) à l'auto-compatibilité (autogamie), peut précipiter la spéciation avec un nombre limité de changements géniques. Ce type de spéciation a par exemple été mis en évidence dans le genre *Stephanomeira* où l'espèce allogame stricte *S. exigua* est isolée reproductivement de l'espèce *S. malheurensis* qui ne présente pas de système d'auto-incompatibilité (Gottlieb, 1973).

### **A.1.2.6 La spéciation par dérive génétique**

Dans la spéciation par dérive génétique, souvent associée à l'allopatricité, l'effet de la sélection est inexistant. Les allèles sont sélectivement neutres (pour un locus donné, tous les allèles ont une valeur sélective ou adaptative identique) et les mutations apparues se fixent au hasard. Seules les mutations délétères (létales) sont éliminées.

La spéciation par dérive génétique sera d'autant plus efficace que la population est petite. En effet, une réduction de l'effectif renforce fortement la consanguinité et l'homozygotie et donc les chances de fixer un caractère nouveau. Par exemple, Dodzhansky et Pavlovsky, (1957) ont comparé l'évolution pendant 18 mois de la variabilité des fréquences génotypiques au niveau de deux inversions chromosomiques dans 10 populations de *Drosophila pseudoobscura* comprenant d'une part, 4000 individus au départ (grande population) et d'autre part, 20 individus au départ (petite population). Leurs résultats indiquent une variabilité significativement supérieure dans la petite

population. Un exemple particulier de l'influence de la dérive génétique est l'effet fondateur. L'effet fondateur est défini par Mayr (1963) comme l'établissement d'une nouvelle population à partir d'un nombre limité d'individus qui portent seulement une petite fraction de la variabilité génétique de la population parentale.

### **A.1.2.7 La spéciation par sélection naturelle**

On peut distinguer deux types de spéciation par sélection naturelle :

-La spéciation en condition de sélection naturelle uniforme où des mutations avantageuses mais incompatibles apparaissent dans des populations séparées géographiquement mais occupant des environnements similaires (Muller, 1940),

-La spéciation écologique dont le processus initial est l'action de la sélection disruptive/divergente : une population localisée dans un environnement homogène est sélectionnée simultanément pour deux phénotypes opposés.

Il me semble important de revenir sur ce dernier mécanisme car comme nous le verrons plus loin il constitue le principal moteur de la différenciation entre le chêne sessile et le chêne pédonculé.

#### **A.1.2.7.1 La spéciation écologique**

Le concept de la spéciation écologique (Schluter, 2001) est basé sur le fait que l'isolement reproductif est une conséquence de la sélection naturelle divergente (ou disruptive) sur des caractères entre différents environnements. L'environnement intègre ici les éléments biotiques et abiotiques de l'habitat (climat, ressources et structures physiques) mais aussi les interactions avec les autres espèces (compétitions pour les ressources, prédation, mutualisme). La spéciation écologique peut apparaître en allopatrie ou sympatrie. Le scénario classique serait que l'isolement reproductif entre deux populations commence en allopatrie par l'accumulation, dans les deux

populations, de différences au niveau de caractères adaptatifs. Puis, l'isolement pré-reproductif serait complété par le renforcement lors du contact secondaire entre les deux populations.

Dans le cas extrême de la spéciation sympatrique stricte où il n'y a pas de phase de différenciation phénotypique et génétique au préalable qui diminue la vigueur hybride, ce seraient les pressions de sélection écologiques qui viendraient directement réduire la fitness des formes intermédiaires. Par exemple, les phénotypes intermédiaires exploiteront moins facilement les ressources du milieu que les phénotypes extrêmes (Hatfield et Schluter 1999).

#### **A.1.2.7.2 Mise en évidence de la spéciation écologique**

La spéciation écologique peut être testée en mettant en évidence le rôle de la *sélection divergente* dans la spéciation, qui, à la différence de la dérive génétique, laisse des traces après son action (Schluter, 2001) :

##### **A.1.2.7.2.1 Réduction de la fitness des hybrides (isolement post-reproductif)**

Une des conséquences de la sélection disruptive est la réduction de la fitness des hybrides et des individus aux phénotypes intermédiaires. Ce type d'isolement est dépendant du milieu et peut disparaître en condition de laboratoire. Les démonstrations de la sélection écologique contre les hybrides restent peu nombreuses. On peut néanmoins citer les travaux de Hatfield et Schluter (1999) qui montrèrent qu'en condition de laboratoire, les hybrides F<sub>1</sub> entre les espèces limnétiques et bentiques d'épinoches avaient une forte fitness alors que leur viabilité était compromise dans leur habitat d'origine.

##### **A.1.2.7.2.2 Sélection pour des caractères impliqués dans l'isolement pré-reproductif**

La sélection disruptive peut agir sur des gènes spécifiques ou sur des caractères phénotypiques impliqués directement ou indirectement dans l'isolement reproductif. Entre *Mimulus cardinalis* et

*M. lewisii* par exemple, l'isolement pré-reproductif est associé à des morphologies florales différentes qui attirent des pollinisateurs différents (Schemske et Bradshaw, 1999).

#### **A.1.2.7.2.3 Evolution parallèle**

Si l'on écarte la spéciation par polyploidisation qui est facilement détectable, seule la sélection disruptive peut mener à l'évolution parallèle de caractères similaires dans différentes populations soumises à des environnements semblables (Johannesson, 2001). Un cas évident d'évolution parallèle a été obtenu dans des populations sympatriques d'épinoche à trois épines (*Gasterosteus aculeatus*). Dans leur étude, Rundle *et al.*, (2000) montrent que les populations d'épinoches qui ont évolué dans des conditions écologiques différentes (limnétique vs bentique) présentent un important isolement reproductif, alors que les populations qui ont évolué indépendamment dans des conditions écologiques similaires peuvent se croiser.

#### **A.1.2.8 Conclusion du chapitre**

Nous avons vu au cours de ce chapitre qu'il est difficile de s'accorder sur la définition d'espèce, beaucoup de concepts existent mais aucun n'est approprié à l'ensemble des organismes. Nous verrons dans le paragraphe A.3 que le chêne illustre bien la complexité du problème.

De la même façon, plusieurs mécanismes de spéciation existent. Certains peuvent provoquer une spéciation soudaine (réarrangement chromosomique, gènes d'incompatibilité au croisement) alors que d'autres induisent une spéciation graduelle (dérive génétique et sélection naturelle). Certains d'entre eux sont faciles à mettre en évidence comme la polyploïdisation, alors que d'autres sont beaucoup plus difficiles comme la spéciation écologique ou la dérive génétique.

## A.2 GENETIQUE DE LA DIFFERENCIATION INTERSPECIFIQUE

### A.2.1 HISTORIQUE

Un des plus gros défis de la biologie évolutive est de comprendre comment un processus continu (l'évolution) peut mener à des discontinuités morphologiques entre des groupes appelés espèces (Coyne, 1992). C'est Darwin qui commença à soulever la question avec son œuvre intitulée "On the origin of species". Puis, à partir de 1920, un certain nombre de scientifiques d'origines diverses élaborèrent la théorie de la génétique des populations qui se propose d'étudier la variation d'une génération à l'autre des fréquences génotypiques et alléliques d'une population déterminée. Il s'agit d'un travail purement théorique qui tend à traduire l'évolution des populations par la mise en équation des résultats de la génétique formelle et du Darwinisme. Ses fondateurs sont l'expérimentateur et théoricien russe S. Chetverikov, le physiologiste anglais J.B.S. Haldane, le statisticien anglais R.A. Fisher, et le biologiste américain S. Wright. Puis, à partir des années 35-40 on commença à analyser les populations naturelles. On peut citer les travaux des trois biologistes Dobzhansky (1937), J.S. Huxley (1942), E. Mayr (1942). Dans les années soixante-dix, le développement de marqueurs moléculaires de nature diverse (isozymes, métabolites secondaires, marqueurs de l'ADN) a permis pour la première fois d'estimer des fréquences alléliques entre les populations. Malheureusement, ces marqueurs étaient neutres et ne permettaient pas d'étudier l'effet de la sélection. Enfin il faudra attendre la fin des années soixante-dix pour voir émerger deux nouvelles techniques qui révolutionneront l'étude de l'évolution. La première est la cartographie dans le génome, des locus contrôlant la variation des caractères quantitatifs (QTL). Nous possédons désormais un support statistique qui permet d'utiliser le polymorphisme de marqueurs neutres pour cartographier des gènes impliqués dans les caractères qui différencient les espèces (Lander et Botstein, 1989 ; Zeng, 1994). La deuxième est le développement des outils moléculaires qui permettent d'identifier, dans les cas les plus favorables, les gènes impliqués dans la différenciation phénotypique.

Ce chapitre est consacré à l'étude de la différenciation interspécifique par la détection de QTL.

Nous essaierons de voir si cette approche permet de répondre :

- (1) aux questions relatives à l'architecture génétique de la différenciation entre les espèces :  
Combien de gènes sont impliqués ? Quelle est leur contribution dans la variation du caractère phénotypique ? Où sont ces gènes et quelle est leur fonction ? Comment expliquer les corrélations entre caractères (pléiotropie ou liaison génétique)?
- (2) aux questions relatives aux mécanismes génétiques à l'origine la différenciation entre les espèces.

## **A.2.2 ARCHITECTURE GENETIQUE DE LA DIFFERENCIATION INTERSPECIFIQUE**

Je voudrais commencer ce chapitre en rappelant que la détection de QTL ne permet pas une recherche exhaustive des gènes impliqués dans la différenciation interspécifique. Les principales limites sont des :

- contraintes expérimentales : la détection de QTL est basée sur le déséquilibre de liaison entre un locus marqueur et un caractère. Si le locus n'est pas polymorphe dans la descendance alors on ne détectera pas de QTL. Comme la plupart des pedigrees utilisés pour la détection de QTL sont issus d'un nombre très limité de parents, le risque est donc fort de sous-estimer le nombre réel de QTL présents dans des populations naturelles.
- contraintes statistiques : le pouvoir de détection de QTL (nombre et effet) est fortement influencé par la taille de la population de cartographie. Or, beaucoup d'études de détection de QTL se font sur des petites populations (entre 90 et 200 individus), surestimant ainsi l'effet de certains QTL (Beavis, 1995).

### **A.2.2.1 Nombre de facteurs impliqués dans la différenciation interspécifique**

Orr dans sa revue sur les différences génétiques entre les espèces (2001) fournit un tableau récapitulatif de 22 études menées sur 54 caractères qui différencient des espèces animales et végétales. Les résultats obtenus pour une même paire d'espèces sont très variables. Par exemple, entre *Drosophila simulans* et *D. sechella*, le contrôle est monogénique pour la morphologie de la larve (présence ou absence de poils sur la surface dorsale des segments abdominaux) (Sucena et Stern, 2000), oligogénique pour la résistance des adultes à la toxine d'une plante hôte (Jones, 1998) et polygénique pour la surface du lobe postérieur (Macdonald et Goldstein, 1999). Dans le règne végétal, Mitchell-Olds (1996) a mis en évidence chez *Arabidopsis thaliana*, deux QTL impliqués dans le contrôle génétique de la période de floraison alors que Lin et Ritland (1997) ont mis en évidence 10 QTL impliqués dans le taux d'autofécondation. Cette hétérogénéité des résultats pourrait s'expliquer par des bases génétiques différentes en réponse à la sélection naturelle ou à la sélection sexuelle (Orr et Coyne 1992). La sélection naturelle pourrait impliquer une seule période d'adaptation avant d'atteindre un optimum, alors que la sélection sexuelle impliquerait plusieurs périodes de co-évolution entre les caractères mâles et femelles menant ainsi à une base génétique polygénique.

### **A.2.2.2 Effet des facteurs impliqués dans la différenciation interspécifique**

#### **A.2.2.2.1 Résultats en faveur d'un modèle oligogénique**

La plupart des résultats indiquent que des gènes à effets majeurs jouent un rôle important dans la réponse à la sélection artificielle. Il est cependant important de noter que ces résultats souffrent de deux limites (i) ils concernent des unités génétiques qui ne sont pas encore des espèces biologiques séparées (téosinte et maïs ; tomate cultivée et sauvage), (ii) ils traduisent le résultat de la sélection

artificielle qui a pour particularité d'être forte en intensité. Ce seront donc nécessairement les gènes à effets forts qui seront sélectionnés dès les premières générations de sélection. Ainsi des résultats obtenus sur des espèces domestiquées ne sont pas le reflet de l'évolution dans la nature où l'intensité de sélection est nettement plus faible (Kingsolver *et al.* 2001). Malheureusement, les études menées sur le déterminisme génétique des caractères qui différencient des espèces sauvages sont beaucoup moins nombreuses.

#### A.2.2.2.1.1 Résultats obtenus sur des espèces domestiquées

Dans l'étude du déterminisme génétique de l'architecture de l'inflorescence qui différencie le maïs et le téosinte, Doebley et Stec (1993) ont mis en évidence, pour chaque caractère étudié, un QTL qui expliquait au minimum 22.5% de la variance phénotypique. Des QTL à effets majeurs ont aussi été mis en évidence dans le contrôle génétique de la taille du fruit qui différencie la tomate cultivée de son espèce sauvage (Grandillo et Tanksley, 1996), dans la période de floraison qui différencie les cultivars de *Brassica oleracea* (Camargo et Osborn, 1996), dans l'architecture du capitule qui différencie le séneçon cultivé *Senecio vulgaris* du séneçon sauvage *Senecio denticulatus* (Moritz et Kadereit, 2001), et dans les caractères morphologiques qui différencient le millet cultivé (*Pennisetum glaucum* ssp *glaucum*) du millet sauvage (*Pennisetum glaucum* ssp *monodii*), (Poncet *et al.*, 2000, 2002)

#### A.2.2.2.1.2 Résultats obtenus sur des espèces sauvages

Les études menées sur l'architecture génétique des caractères qui différencient des espèces sauvages sont beaucoup moins nombreuses que chez les espèces domestiquées. On peut néanmoins citer quelques études où le modèle oligogénique a été mis en évidence. Le plus bel exemple est celui de *Mimulus* (Bradshaw *et al.*, 1998) où parmi les 12 caractères de la morphologie florale qui différencient *M. lewisii* et *M. cardinalis*, 9 sont contrôlés par un QTL qui explique plus de 25% de

la variance phénotypique totale. L'effet d'un gène majeur a aussi été mis en évidence dans la différenciation du système de reproduction de deux accessions sauvages de tomate (Georgiady *et al.*, 2002). Enfin, dans le règne animal, la différenciation du taux de phéromones hydrocarbonées cuticulaires entre 43 lignées *D. melanogaster* échantillonnées de par le monde est contrôlée par l'insertion-délétion d'un nucléotide (Takahashi *et al.*, 2001).

#### A.2.2.2.2 Résultats en faveur d'un modèle polygénique

Plusieurs études sont en faveur du modèle polygénique de Fisher (1930) dans lequel les caractères quantitatifs sont contrôlés par un grand nombre de loci ayant chacun un faible effet sur l'expression du phénotype. Laurie *et al.*, (1997) ont trouvé que plusieurs QTL à effets modérés (entre 10 et 15 % de la différence interspécifique) contrôlent les différences morphologiques de l'appareil génital mâle entre *Drosophila simulans* et *D. mauritiana*. L'étude de Zeng *et al.*, (2000) conforte ces résultats en montrant que le lobe postérieur de l'organe génital mâle qui différencie les deux espèces *Drosophila simulans* et *D. mauritiana* est contrôlé par 19 QTL dont le plus gros explique 11% de la différence entre les deux espèces. Similairement, Lin et Ritland (1997), suggèrent que les différences morphologiques florales qui affectent le taux d'autofécondation qui différencie deux espèces de *Mimulus*, sont contrôlées par plusieurs gènes, la plupart ayant des effets mineurs. Chez le riz, la stérilité des hybrides issus du croisement entre deux cultivars de *Oryza sativa* L est sous le contrôle de 6 QTL qui expliquent entre 4.1 et 21.1% de la variance phénotypique (Li *et al.*, 1997). Chez *Arabidopsis thaliana*, 13 caractères du développement et de la morphologie de l'inflorescence qui différencient 21 écotypes sont essentiellement contrôlés par des QTL à effets mineurs (Ungerer *et al.*, 2002). Chez le maïs, les caractères morphologiques qui différencient 2 espèces sauvages (*Zea diploperennis* et *Z.mays ssp. parviglumis*) sont contrôlés par un grand nombre de QTL à effets mineurs (Westerbergh et Doebley, 2002). Enfin, chez le tournesol il a été mis en évidence un déterminisme polygénique pour les caractères morphologiques qui différencient les deux espèces

*H.annuus* et *H.debilis*, associé à un déterminisme oligogénique de la viabilité du pollen (Kim et Rieseberg, 1999).

### **A.2.2.3 Interactions entre les gènes impliqués dans la différenciation interspécifique**

Il est souvent observé dans les populations hybrides, une réduction de la fitness des hybrides de première génération ( $F_1$ ) et une forte mortalité ou stérilité des hybrides de deuxième génération (Hybrid breakdown). Selon la théorie, (Dobzhansky 1937 ; Muller 1942) l'isolement post-reproductif est une conséquence de la fixation dans chacune des espèces, d'allèles différents qui interagissent négativement dans les hybrides. Si les interactions se font entre allèles d'un même locus on parle d'effet de dominance et si les interactions se font entre allèles de loci différents on parle d'effet d'épistasie. L'idée la mieux acceptée pour expliquer l'augmentation de la mortalité des hybrides  $F_2$  par rapport au  $F_1$  est que les interactions entre les loci des espèces sont épistatiques et qu'elles tendent à être récessives (Muller, 1942 ; Turreli et Orr, 2000). Cependant, la plupart des études indiquent l'absence d'effet d'épistasie. Ce constat pourrait être en partie dû aux problèmes statistiques rencontrés lors des analyses (classes génotypiques non homogènes, grand nombre de tests). On peut néanmoins citer quelques récents exemples où les interactions entre gènes sont significatives. Dans une population  $F_2$  composée d'hybrides entre les insectes *Nasonia vitripennis* et *N. giraulti*, il semble que les interactions épistatiques entre gènes soient la principale cause de la stérilité ou mortalité hybride (Gadau *et al.*, 1999). Similairement, dans une population  $F_1$  issue de ce même croisement, il a été montré que la taille de l'aile (fortement discriminante) est sous le contrôle de QTL à effets majeurs qui interagissent entre eux (Gadau *et al.*, 2002). Chez la drosophile, Zeng *et al.*, (2000) suggère que les interactions épistatiques ont peu d'effet sur la morphologie du lobe postérieur qui différencie *D.simulans* et *D.mauritiana*, alors que True *et al.* (1997) mettent en évidence des effets d'épistasie sur le nombre de soies anales qui différencie ces deux mêmes

espèces. Dans le règne végétal, l'épistasie a été mise en évidence chez le riz dans une sous-espèce hybride de *Oryza sativa* L, où 21 paires de marqueurs non liés ont montré des interactions significatives avec la stérilité des épillets (Li *et al.*, 1997).

Si ces différents résultats mettent en évidence le rôle des interactions génétiques dans le processus d'isolement reproductif, il ne faut pas en faire une généralité. Par exemple, dans des hybrides naturels d'iris de Louisiane, les interactions peuvent être positives et négatives entre les génomes des espèces parentales au niveau nucléaire et cytoplasmique (Burke *et al.*, 1998). Pareillement, d'autres études ont mis en évidence des interactions favorables entre les génomes des taxons divergents qui semblent favoriser l'adaptation des hybrides (De Vicente et Tanksley, 1993 ; Li *et al.*, 1995 ; Schwarzbach *et al.*, 2001). Ces interactions favorables qui peuvent s'expliquer par des effets de sur-dominance, d'épistasie, ou la présence de QTL ayant un effet opposé dans chaque espèce parentale, peuvent conduire à de la ségrégation transgressive (Rieseberg *et al.*, 1999a). La ségrégation transgressive est observée lorsque le phénotype des hybrides excède celui des parents en positif ou négatif. Elle diffère de la vigueur hybride qui est très prononcée dans la première génération hybride et qui fait référence à une moyenne du caractère des hybrides qui excède positivement celle des parents (Rieseberg *et al.*, 2000).

Plusieurs études ont essayé de comprendre le mode d'action des gènes dans l'établissement de la ségrégation transgressive. Dans une population hybride F<sub>2</sub> obtenue à partir du croisement entre deux espèces sauvages de téosinte *Zea diploperennis* et *Z.mays ssp. parviglumis*, Westerbergh *et al.* (2002) expliquent la présence de transgressivité par des QTL à effets opposés dans chaque espèce parentale. Par contre, ils ne mettent pas en évidence d'effet de l'épistasie.

En conclusion de ce chapitre nous pouvons dire que l'étude de l'architecture génétique par la l'approche QTL ne permet pas de trancher entre le modèle polygénique de Fisher (1930), dans lequel les caractères quantitatifs sont contrôlés par un grand nombre de loci ayant chacun un faible effet sur l'expression du phénotype et le modèle oligogénique qui postule qu'une variation

phénotypique continue peut être la conséquence de quelques gènes à effets majeurs (Gottlieb, 1984 ; Tanksley, 1993). Il apparaît qu'en fonction du règne, de l'espèce, du caractère, et des mécanismes de spéciation les résultats diffèrent considérablement.

### **A.2.3 MECANISMES QUI CONDUISENT A LA DIFFERENCIATION INTERSPECIFIQUE**

#### **A.2.3.1 Facteurs agissant sur la diminution de la recombinaison au niveau des gènes de spéciation**

Dans sa revue sur la vision génique du processus de spéciation, Wu (2001) définit l'espèce comme une population qui est incapable de partager les gènes contrôlant les caractères adaptatifs en cas de contact avec une autre population. Néanmoins, les autres zones du génome non impliquées dans l'adaptation, peuvent être homogènes entre les deux populations. Une conséquence importante de la vision génique de la spéciation est que l'allèle ou le segment de chromosome devient l'unité de l'isolement plutôt que le génome entier (Rieseberg et Burke, 2001). Il est donc possible d'estimer l'efficacité d'un gène dans le processus d'isolement interspécifique en mesurant la longueur du segment chromosomique qu'il protège du flux de gènes (Rieseberg et Burke 2001). Cela revient donc à étudier les mécanismes par lesquels les facteurs agissant sur la diminution de la recombinaison au niveau des gènes de spéciation, peuvent étendre leurs effets sur des fragments de chromosomes plus longs et ainsi contribuer à la divergence de l'espèce. Deux types de facteurs peuvent agir sur la réduction du taux de recombinaison : des facteurs géniques et des facteurs chromosomiques (Rieseberg *et al.*, 1999a).

##### **A.2.3.1.1 Facteurs chromosomiques**

Afin de dissocier l'effet des réarrangements chromosomiques de l'effet des facteurs géniques dans la diminution du taux de recombinaison, deux approches peuvent être menées.

#### **A.2.3.1.1.1 Cartographie de QTL**

Quillet *et al.* (1995) ont par exemple mis en évidence que les QTL impliqués dans la stérilité des hybrides entre *Helianthus argophyllus* et *H. annuus* se situent dans 3 intervalles de groupes de liaisons différents et expliquent plus de 80 % de la variance phénotypique totale de la stérilité. Après observation des chromosomes pendant la méiose, ils révélèrent un taux significatif de mauvais appariements des chromosomes en étroite corrélation avec la viabilité du pollen, indiquant ainsi que les réarrangements chromosomiques (ou translocation) sont des facteurs fortement impliqués dans la stérilité hybride.

#### **A.2.3.1.1.2 Analyse des patterns d'introgression.**

Si le réarrangement chromosomique contribue à réduire la fitness hybride alors les blocs qui portent ces réarrangements seront contre-sélectionnés dans les descendances introgressées. C'est par ce biais que Rieseberg *et al.*, (1999b) ont mis en évidence dans 3 zones hybrides naturelles de *H. petiolaris* et *H. annuus*, une réduction significative de l'introgression de 26 segments chromosomiques dont 16 étaient liés à la stérilité. La comparaison du taux d'introgression à travers les segments chromosomiques colinéaires et réarrangés indique que près de 50% de la barrière à l'introgression est due à des réarrangements chromosomiques.

Ainsi comme l'avait prédit l'analyse QTL, les réarrangements chromosomiques doivent représenter une barrière au flux de gènes.

#### **A.2.3.1.2 Facteurs géniques**

Des facteurs géniques peuvent aussi affecter les taux de recombinaison (Sano, 1990), mais leur effet sur la réduction du flux de gènes ou de l'isolement reproductif n'a pas été étudié. On peut tout de même noter les résultats de Ting *et al.* (2000) qui mettent en évidence la monophilie de deux espèces proches *Drosophila mauritiana* et *D. simulans* en utilisant la généalogie d'un gène

(*Odysseus*) impliqué dans la stérilité hybride. Un point intéressant de leur étude est la mise en évidence d'un mélange de généalogies dans les régions éloignées de plus de 1.8 Kb du gène de spéciation. Cela suggère des processus d'entraînement (hitch-hiking) à très courte distance. Ainsi, au vu de leurs résultats, ils font l'hypothèse que les espèces divergentes, partiellement isolées reproductivement, doivent être perméables à l'introgession dans la plupart des régions du génome, excepté dans des petites régions proches du ou des gènes de spéciation.

### **A.2.3.2 Forces évolutives impliquées dans la différenciation interspécifique**

S'il est difficile comme nous venons de le voir de reconstruire les évènements génétiques qui mènent à la spéciation, il est encore plus difficile de mettre en évidence les mécanismes à l'origine de ces changements. Les questions posées sont : quelle est la part relative de la sélection et de la dérive génétique dans le processus de spéciation ? Comment expliquer le phénomène de spéciation malgré la présence de flux de gènes dans certaines populations ?

#### **A.2.3.2.1 Effet du flux de gènes**

La forme la plus rapide de spéciation par hybridation est la polyploïdisation dans laquelle l'isolement reproductif apparaît comme un sous-produit instantané de la duplication du génome de l'hybride. Une autre forme de spéciation par hybridation, qui n'est pas instantanée mais néanmoins rapide (Rieseberg *et al.*, 2000), est la spéciation d'hybrides homoploïdes. Elle consiste en l'apparition d'hybrides stables et fertiles sans changement dans le nombre de chromosomes (Stebbins 1957 ; Grant, 1958). Le processus par lequel de nouveaux hybrides homoploïdes apparaissent est relativement simple (Templeton, 1981). L'hybridation entre des espèces génétiquement ou chromosomiquement divergentes donne lieu à des hybrides moins viables (ou fertiles) que les parents. La sélection favorise les hybrides les plus viables (ou fertiles) qui s'hybrident entre eux et s'isolent des espèces parentales par des facteurs de stérilité génique ou

chromosomique (Stebbins, 1957 ; Rieseberg *et al.*, 1995), des réarrangements chromosomiques (Rieseberg *et al.*, 1995), ou des divergences écologiques (Templeton, 1981).

#### **A.2.3.2.2 Estimation de la part relative de la dérive génétique et de la sélection dans la différenciation entre les espèces**

Une récente méthode, basée sur l'étude de la direction des effets des QTL qui contribuent aux différences phénotypiques, permet de reconstruire l'histoire de caractères complexes. Si un caractère est le résultat d'une sélection directionnelle continue, les effets responsables de ces QTL doivent avoir la même direction dans l'espèce. Au contraire, si le caractère a divergé sous la neutralité, alors des QTL avec des directions opposées (ou antagonistes) doivent être observés. Cette approche a été formalisée par Orr (1998) qui a mis au point un test qui compare la proportion de QTL antagonistes dans une lignée donnée avec le nombre attendu sous l'hypothèse de neutralité. Dans une première version du test, le signe et l'effet des QTL sont pris en compte, alors que dans une deuxième version, les effets sont supposés identiques et seul le signe est pris en compte. Ce dernier test fut récemment utilisé par Rieseberg *et al.*, (2002) pour tester l'effet de la sélection directionnelle sur différents caractères. Leur analyse est basée sur les résultats de 86 études qui fournissent 2684 QTL impliqués dans 572 caractères.

Une première analyse impliquant des croisements entre une espèce domestiquée et une espèce sauvage met en évidence une sélection divergente plus forte pour les caractères domestiqués (taille du fruit, biomasse, solides solubles...) que ceux non domestiqués (morphologie des feuilles, architecture ...).

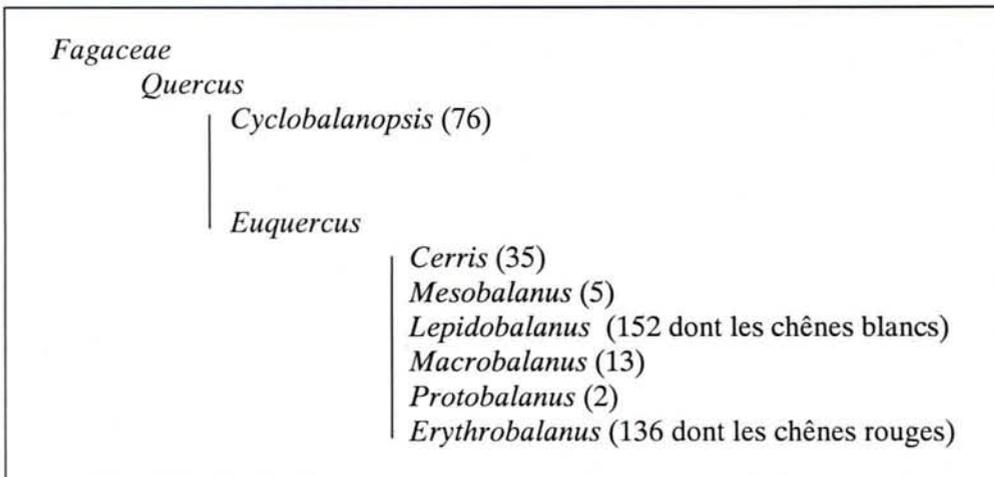
Dans une deuxième analyse qui implique des croisements entre des espèces sauvages, les caractères ont été classés par catégorie : règne animal, règne végétal, croisement intraspécifique, interspécifique, espèce allogame, autogame, caractère lié aux traits d'histoire de vie, caractère morphologique et caractère de durée de développement. Toutes les catégories de caractères dévient significativement de la neutralité, suggérant ainsi que la sélection directionnelle contribue fortement

à la différenciation des caractères dans tous les types d'organismes et tous les types de caractères. Afin de tester l'effet règne, type de croisement, type de système de croisement et type de caractère, une analyse de variance multivariée a ensuite été réalisée. Les résultats ne mettent pas en évidence de différence entre les différents règnes, ni entre les espèces autogames et les allogames. Cependant la différence est forte entre les croisements intra et interspécifiques indiquant que les différences entre les espèces résultent plus de la sélection directionnelle que les différences intra-espèce. Ces résultats corroborent les récents points de vue qui mettent l'accent sur le rôle de la sélection divergente comme une cause de la spéciation (Wu, 2001) et rejettent les modèles de spéciations basés sur des processus non sélectifs.

### **A.2.3.3 Conclusion du chapitre**

L'étude de l'architecture génétique par la méthode de détection de QTL ne permet pas de trancher entre un modèle polygénique et oligogénique du contrôle des caractères impliqués dans la différenciation interspécifique. Les résultats sont fortement hétérogènes entre les différentes études. Il ne faut pas perdre de l'esprit que la détection de QTL n'est pas une méthode de recherche exhaustive des gènes impliqués dans la différenciation entre les espèces, elle reste limitée par des contraintes statistiques et expérimentales qui pourraient parfois expliquer l'hétérogénéité des résultats. Cependant, cette méthode permet de dégrossir le terrain lorsqu'aucune connaissance n'a été acquise au préalable sur le contrôle génétique de la différenciation. Elle permet par ailleurs de faire des hypothèses sur les forces évolutives qui interviennent dans le processus de spéciation.

**Figure 1** : Taxonomie du genre *Quercus* selon Camus.



Le nombre d'espèces identifiées est indiqué entre parenthèses.

## A.3 PRESENTATION DU CHENE

### A.3.1 LE GENRE *QUERCUS*

#### A.3.1.1 Taxonomie

Les différentes espèces de chêne font partie du genre *Quercus* qui appartient à la famille des *Fagaceae*. Cette famille est importante puisqu'elle contient 8 genres et plus de mille espèces. Les hêtres et les châtaigniers en font par exemple partie.

La monographie des chênes de Camus (1934-54) est l'ouvrage de référence pour l'étude de la taxonomie du genre *Quercus*, même si cette classification est encore sujette à discussion. Au sein du genre *Quercus*, Camus distingue deux sous-genres : *Euquercus* et *Cyclobalanopsis* (Figure 1) qui couvrent deux zones géographiques distinctes. Le premier qui contient la section *lepidobalanus* des chênes blancs, se situe sur tout l'hémisphère Nord alors que le deuxième se situe en Asie.

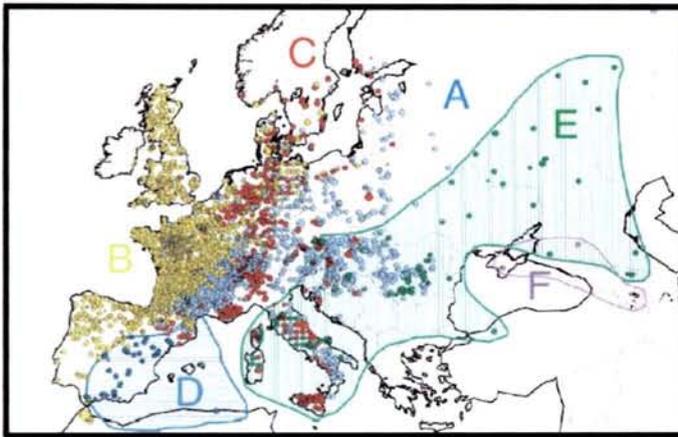
L'Europe compte environ 30 espèces de *Quercus*, dont les deux plus importantes en surface sont le chêne sessile, *Q. petraea* (Matt.) Liebl. et le chêne pédonculé, *Q. robur* L.

#### A.3.1.2 Histoire des chênes

Le genre *Quercus* a laissé beaucoup de traces fossiles sur les trois continents de l'hémisphère Nord (Manos et Stanford, 2001). Tout laisse à penser que la plupart des espèces actuelles étaient déjà génétiquement séparées dès le Miocène (Axelrod, 1983). Par ailleurs, l'existence d'espèces morphologiquement voisines au couple chêne sessile, chêne pédonculé en Amérique du Nord (*Quercus alba* et *Quercus montana*) suggère en fait que nos deux espèces étaient séparées depuis le Tertiaire.

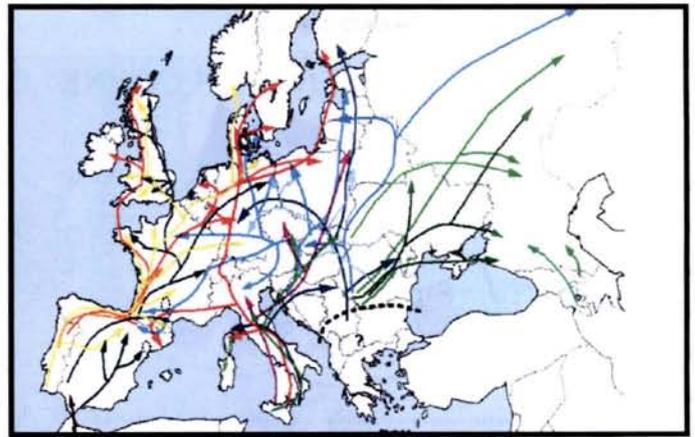
La distribution actuelle des chênes résulte de cycles de recolonisation-extinction engendrés par plus de 17 alternances de périodes glaciaires et interglaciaires au cours du quaternaire (il y a deux

**Figure 2 :** Distribution géographique des variants d'ADN chloroplastique chez les chênes blancs européens.



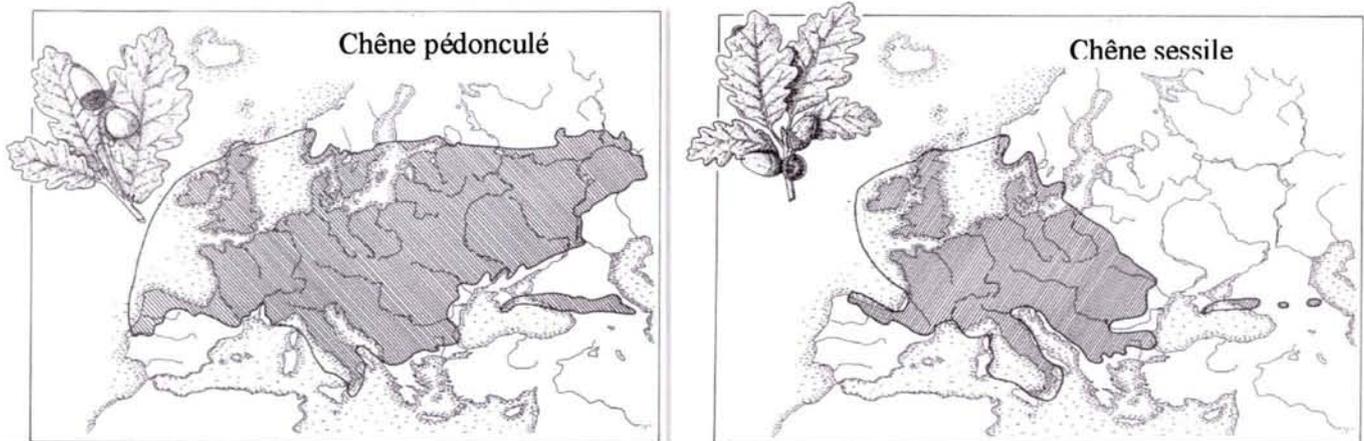
L'analyse systématique de la diversité de l'ADN du chloroplaste faite sur la base d'un échantillon de 2673 forêts différentes a révélé l'existence d'une quarantaine de variants chloroplastiques différents regroupés par leur similarité génétique en 6 grandes lignées représentées par des lettres et des couleurs différentes. (Petit *et al.*, 2002b)

**Figure 3:** Voies de recolonisation empruntées par les chênes blancs européens.

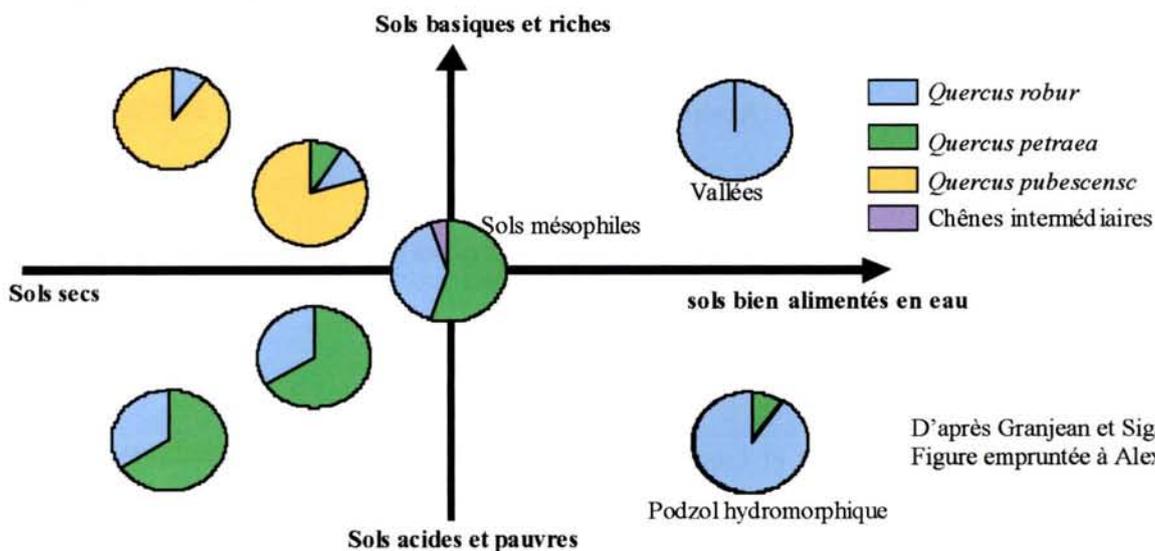


Les flèches dont la couleur correspond aux variants chloroplastiques indiquent les principales voies de migration suivies par les chênes (Petit *et al.*, 2002a)

**Figure 4 :** Répartition des chênes sessiles et pédonculés en Europe.



**Figure 5 :** Ecologie des chênes blancs européens



D'après Granjean et Sigaud (1987).  
Figure empruntée à Alexis Ducousso.

millions d'années) (Kremer *et al.*, 2002b). Ces changements climatiques ont entraîné l'extinction de certaines espèces en Europe, comme le Séquoia, le Liquidambar, le Tsuga, et le confinement d'autres dans des zones refuges. L'utilisation du pollen fossile par les paléobotanistes a permis de reconstruire la recolonisation des espèces forestières en Europe (Huntley et Birks, 1983 ; Brewer *et al.*, 2002). Ainsi, l'emplacement des zones refuges pour les chênes se situe au niveau des péninsules ibérique, italienne, et balkanique. Récemment, la recolonisation postglaciaire des chênes a pu être étudiée, plus en détail, en analysant la répartition géographique du polymorphisme de l'ADN chloroplastique qui n'est transmis que par les gamètes femelles et donc disséminé par les glands. Sur la base d'un inventaire systématique réalisé sur 2673 populations, 42 variants différents ont été dénombrés qui se regroupent en six lignées (Figure 2), (Petit *et al.*, 2002b). En se basant sur la distribution des haplotypes les plus fréquents, et en prenant en compte la distribution des pollens fossiles de chêne en Europe (Brewer *et al.*, 2002), une carte des voies de recolonisations des chênes blancs européens a pu être proposée (Figure 3) (Petit *et al.*, 2002a). Cependant, compte tenu de la similarité entre les grains de pollen et d'un large échange d'ADN chloroplastique entre les différentes espèces de chêne blanc (Petit *et al.*, 1997 ; Dumolin-Lapègue *et al.*, 1999 ; Petit *et al.*, 2002b), le statut espèce n'a pas été pris en compte dans ces études.

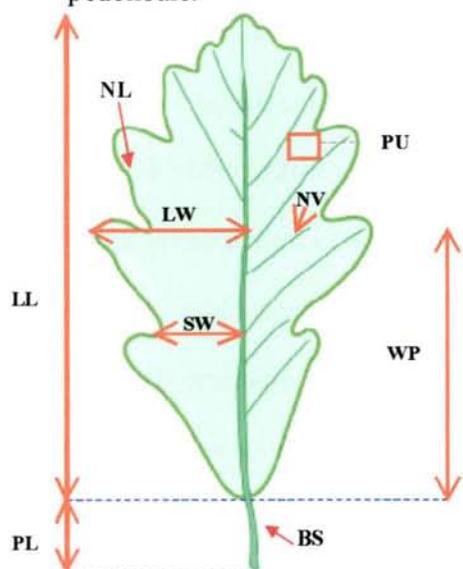
### A.3.2 CARACTERISTIQUES DES CHENES SESSILES ET PEDONCULES

#### A.3.2.1 Différenciation phénotypique

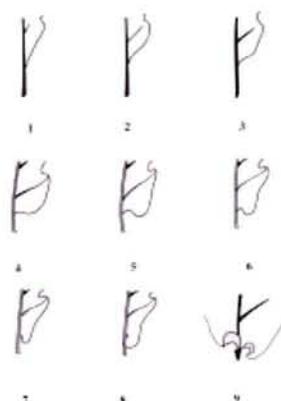
##### A.3.2.1.1 Ecologie

Le chêne sessile et le chêne pédonculé sont deux espèces sympatriques (Figure 4) qui néanmoins partagent des niches écologiques différentes (Figure 5). Le chêne pédonculé est une espèce pionnière qui colonise des milieux ouverts. Il préfère les sols humides, alluviaux, plus riches en minéraux, à pH plus basique. Le chêne sessile quant à lui est une espèce post-pionnière qui se met

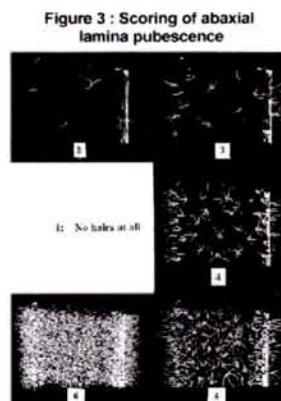
**Figure 6** : Caractères morphologiques foliaires qui différencient le chêne sessile et le chêne pédonculé.



Longueur du limbe (LL)  
 Longueur du pétiole (PL)  
 Largeur du plus gros lobe (LW)  
 Largeur du sinus (SW)  
 Longueur entre le lobe et la base (WP)  
 Nombre de lobes (NL)  
 Nombre de nervures intercalaires (NV)  
 Forme de l'oreillette (BS)  
 Pubescence (PU)



**Echelle de mesure de la forme de l'oreillette.**

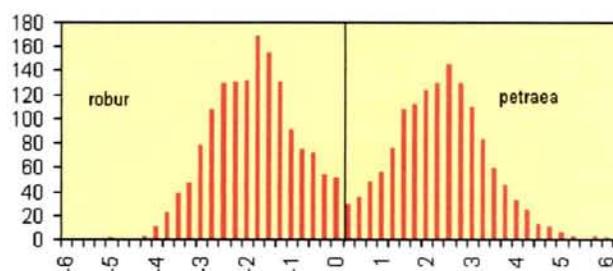


**Echelle de mesure de la pilosité**

**Tableau 2** : Amplitude de variation (95% des valeurs) des différents caractères morphologiques chez *Q. petraea* et *Q. robur* d'après Kremer *et al.* 2002a.

	<i>Q. robur</i>	<i>Q. petraea</i>
NV	1.6-8.0	0-3.1
PL (mm)	2.8-9.4	6.8-22.0
PU	1.0-4.4	2.2-5
NL	6.6-11.6	8.6-14.4
SW (mm)	8.3-18.4	12.4-25.4
BS	2.6-9.0	2.4-7.6

**Figure 7** : Distribution des valeurs de la première composante principale. Etude réalisée sur 2865 arbres (1346 sessiles et 1519 pédonculés) échantillonnés dans 8 sites européens.

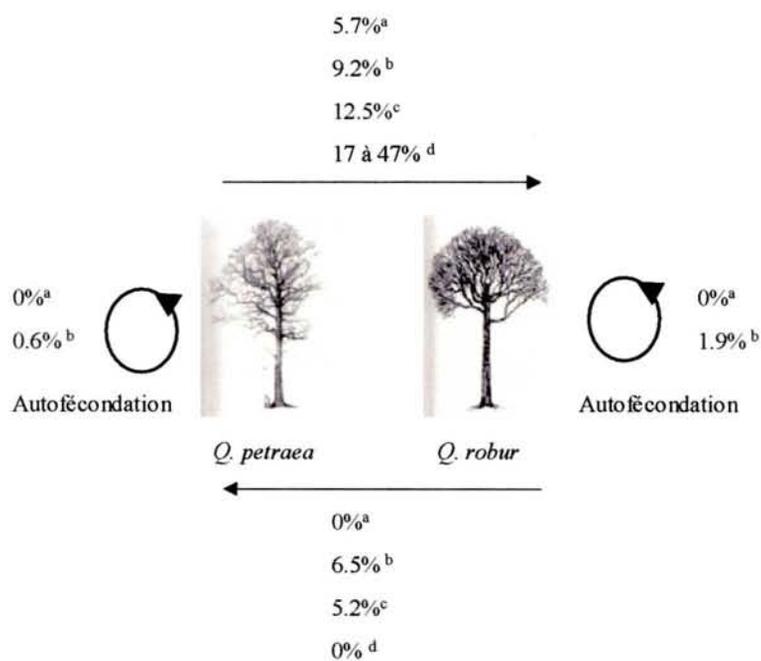


en place après le chêne pédonculé dans la dynamique de la forêt. Il pousse préférentiellement sur les sols secs, bien drainés, relativement pauvres en minéraux, à pH acides (Jones, 1959 ; Grandjean et Sigaud, 1987 ; Becker et Lévy, 1990). Ainsi le paysage forestier sera constitué préférentiellement de chênes pédonculés dans les vallées, et de chênes sessiles plus en altitude. Cependant, il n'est pas rare de trouver sur une même parcelle les deux espèces en mélange, notamment sur des sites constitués de mosaïques de sols humides et secs. Par exemple, sur 1400 sites du nord et de l'ouest de l'Allemagne analysés sur la base de caractères morphologiques foliaires, seuls 18 % étaient constitués de forêts monospécifiques (Kleinschmit et Kleinschmit, 2000).

#### **A.3.2.1.2 Morphologie**

Un premier critère de distinction est le gland. En effet, le gland du chêne pédonculé est allongé avec un long pédoncule (d'où son nom) alors que le chêne sessile possède un gland trapu avec un petit pédoncule. Plusieurs caractères foliaires indiqués sur la figure 6 différencient aussi les deux espèces (Kissling, 1980). La propriété discriminante de ces caractères a été mise en valeur par des analyses multivariées (Dupouey, 1983 ; Grandjean et Sigaud, 1987 ; Dupouey et Badeau, 1993 ; Kremer *et al.*, 2002a). Toutes mettent en évidence une distribution bimodale avec néanmoins un chevauchement dans les phénotypes intermédiaires (Figure 7). L'étude de Kremer *et al.* (2002a) réalisée sur 8 populations européennes différentes et qui utilise 3 types d'analyses multivariées indique que les caractères les plus discriminants sont : le nombre de nervures intercalaires, la longueur du pétiole, la forme de l'oreillette, la largeur du sinus et la pilosité. Le chevauchement dans les phénotypes intermédiaires pourrait s'expliquer par la présence de formes hybrides ou par une grande variabilité intraspécifique comme en témoigne le tableau 2 qui donne la variabilité phénotypique moyenne de différents caractères foliaires dans les 8 populations naturelles.

**Figure 8** : Taux d'hybridation entre le chêne sessile et le chêne pédonculé et taux d'autofécondation.



<sup>a</sup> Aas (1993); <sup>b</sup> Steinhoff (1993); <sup>c</sup> Kleinschmit *et al.*, (1995b); <sup>d</sup> Bacilieri *et al.*, (1995)

### A.3.2.2 Hybridation

Le chêne sessile et le chêne pédonculé sont des espèces allogames, monoïques : chaque arbre porte les deux sexes sur des fleurs séparées. Les floraisons des deux espèces sont synchrones à l'intérieur d'un peuplement mélangé (Bacilieri *et al.*, 1995) donc aucun obstacle d'ordre phénologique ne s'oppose à l'hybridation. L'hybridation entre le chêne sessile et le chêne pédonculé a été mise en évidence dans des croisements contrôlés d'une part (Aas, 1993 ; Steinhoff, 1993 ; Kleinschmit *et al.*, 1995b) et en population naturelle d'autre part (Bacilieri *et al.*, 1995, 1996 ; Streiff, 1998, sa thèse). L'ensemble des études indique une hybridation asymétrique entre les deux espèces avec un sens préférentiel du père sessile vers la mère pédonculé (Figure 8). L'autofécondation semble par contre très marginale. Une caractéristique importante des hybrides juvéniles est la ressemblance de leur morphologie foliaire avec le phénotype maternel (Bacilieri *et al.*, 1995 ; Kleinschmit *et al.*, 1995a). L'analyse morphologique de croisements contrôlés indique par ailleurs, que la période juvénile peut durer 5 ans (Kleinschmit *et al.*, 1995a). D'autre part, il a été montré en croisement contrôlé que les formes intermédiaires sont fertiles avec des performances proches de celle du chêne pédonculé et supérieures au chêne sessile dans les 7 premières années de croissance (Steinhoff, 1997). Cependant, la mise en évidence de formes hybrides dans les peuplements mélangés adultes à partir de marqueurs morphologiques est très inférieure au taux d'hybrides mis en évidence au stade graine à l'aide de marqueurs moléculaires. Deux explications à cela : soit les marqueurs morphologiques ne sont pas de bons diagnostics des formes hybrides, ce qui est vrai au stade juvénile, mais n'a pas encore pu être vérifié au stade adulte faute de disponibilité de matériel végétal âgé de pedigree connu. Soit il existe une barrière reproductive post-zygotique qui élimine progressivement les hybrides au cours de la vie du peuplement. Pour apporter des éléments de validité à cette deuxième hypothèse, il est nécessaire que des études soient menées sur la valeur sélective des hybrides. A l'heure actuelle, nous ne disposons d'aucune donnée sur la viabilité des hybrides ni en croisement contrôlé, ni en peuplement naturel, ni sur leur fertilité. Des observations

**Tableau 3** : Liste des différentes études menées sur la différenciation génétique entre le chêne sessile et le chêne pédonculé.

Marqueur	Nombre de marqueurs	Répartition des Populations	C	Marqueur diagnostique	P	Références
ADN chloroplastique	-	France	0.032 <sup>a</sup>	Non	-	Petit <i>et al.</i> ,1993
ADN ribosomal	2	-	-	Non	-	Muir <i>et al.</i> ,2001
ADN ribosomal	-	France	-	Non	0	Petit <i>et al.</i> ,1993
AFLP	170	Belgique	0.073 <sup>b</sup>	Non	-	Coart <i>et al.</i> ,2002
AFLP	155	Toute l'Europe	0.030 <sup>a</sup>	Non	-	Mariette <i>et al.</i> ,2002
AFLP	92	Hollande	-	Non	14	Bakker <i>et al.</i> , 2000
Séquences répétées	7	-	-	Non	-	Zoldos <i>et al.</i> ,2001
Isozymes	11	France	0.034 <sup>c</sup>	Non	54	Bacilieri <i>et al.</i> ,1995
Isozymes	8	Europe de l'est	0.15 <sup>a</sup>	Non	-	Gömöry <i>et al.</i> , 2001
			(0.004-0.467)			
Isozymes	1	Europe de l'est	0.543 <sup>a</sup>	Non	-	Gömöry, 2000
Isozymes	11	Toute l'Europe	0.03 <sup>a</sup>	Non	-	Zanetto <i>et al.</i> ,1994
			(0.004-0.11)			
Microsatellites	20	Toute l'Europe	-	Non	-	Muir <i>et al.</i> ,2000
Microsatellites	6	Toute l'Europe	0.013 <sup>a</sup>	Non	-	Mariette <i>et al.</i> ,2002
protéine	73	Toute l'Europe	-	Non	4.1	Barreneche <i>et al.</i> ,1996
RAPD	2800	Toute l'Europe	0.005 <sup>d</sup>	Non	2	Bodénès <i>et al.</i> ,1996a
RAPD	419	France	-	Non	3.3	Moreau <i>et al.</i> ,1994
SCAR	9	Toute l'Europe	(0.004-0.53) <sup>e</sup>	Non	-	Bodénès <i>et al.</i> ,1997

C = Coefficient de différenciation interspécifique moyen et gamme de variabilité entre parenthèses

P = % de locus qui présentent des différences de fréquences alléliques interspécifiques significatives

<sup>a</sup> = Coefficient de différenciation interspécifique de Nei.

<sup>b</sup> = Coefficient de différenciation interspécifique de Wright.

<sup>c</sup> = Distance génétique de Nei entre le chêne sessile et le chêne pédonculé

<sup>d</sup> = Diversité interspécifique nucléotidique entre *Q. robur* et *Q. petraea*

<sup>e</sup> = Coefficient de différenciation totale

non quantifiées suggèrent cependant que les individus issus de croisements contrôlés interspécifiques sont fertiles (Roussel, communication personnelle).

### A.3.2.3 Différenciation moléculaire

Le chêne sessile et le chêne pédonculé se différencient bien au stade adulte à l'aide de marqueurs morphologiques et écologiques. Néanmoins, la présence de formes intermédiaires et la difficulté de distinguer les espèces au stade juvénile, en raison de l'hérédité maternelle des caractères morphologiques distinctifs, illustrent les limites des marqueurs phénotypiques. Le développement des marqueurs moléculaires dans les années 80 a permis d'avoir une vision génétique de la différenciation interspécifique. La différenciation génétique entre le chêne sessile et le chêne pédonculé a été étudiée à l'aide de marqueurs de type isozyme, RAPD (Random Amplified DNA Polymorphism), SCAR (Sequence Characterized Amplified Region), protéine, AFLP (Amplified Fragment Length Polymorphism), microsatellite, ADNc (ADN complémentaire), ADNr (ADN ribosomique). Le tableau 3 fait la synthèse des travaux associés à ces études dont la plupart mettent en évidence des différences en terme de fréquence allélique, aucun marqueur diagnostique de l'espèce n'a pu être mis en évidence (marqueur dont un allèle serait présent chez tous les individus d'une espèce et absent chez les individus de l'autre espèce).

L'homogénéité génétique entre les deux espèces s'illustre bien par l'étude de Bodénès *et al.* (1996a) qui à partir de 2800 fragments RAPD ont seulement obtenu des différences significatives de fréquence allélique pour 2% des fragments. L'homogénéité entre les deux génomes a aussi été mise en évidence, récemment au niveau de 7 séquences hautement répétées (100 à 700 répétitions) présentes dans les deux espèces et qui montrent le même pattern d'hybridation in situ (Zoldos *et al.*, 2001).

Si le génome semble en grande partie homogène, certaines études ont cependant mis en évidence des marqueurs fortement discriminants. Par exemple, l'étude de la variabilité génétique d'une déshydrogénase NAD dépendante dans 5 populations de l'Europe de l'est indique que la

composante interspécifique de la variabilité compte pour 54 % de la diversité génétique totale (Gömöry, 2000). Des fortes valeurs de différenciation interspécifique ont aussi été obtenues pour des SCAR (Bodénès *et al.*, 1997) et pour des marqueurs AFLP. Coart *et al.* (2002) ont par exemple obtenu un marqueur AFLP (parmi 170 étudiés) qui présente 71% de différence de fréquence allélique entre les deux espèces. Similairement, parmi les 155 marqueurs AFLP étudiés par Stéphanie Mariette au cours de sa thèse (2001), deux d'entre-eux possèdent un coefficient de différenciation interspécifique supérieur à 40%. Afin de tester le pouvoir discriminant de ces 2 marqueurs dans des populations naturelles, S. Mariette a réalisé un test d'identification basé sur des fréquences alléliques. Elle a ainsi pu montrer que le marqueur le plus discriminant ( $G_{ST}=0.55$ ) permet de séparer les deux espèces dans 87% des cas avec cependant des fluctuations en fonction de la provenance de l'arbre (entre 68 et 98 %). A l'heure actuelle ce sont les meilleurs marqueurs diagnostiques que nous possédons. Ils donnent de meilleurs résultats que ceux obtenus à partir de marqueurs microsatellites qui fournissent une bonne assignation des espèces dans 78% des cas lorsqu'on cumule 20 microsatellites (Muir *et al.*, 2001).

Il semble donc que le génome du chêne sessile et du chêne pédonculé soient très homogènes excepté pour quelques régions fortement différenciées. Ce constat s'illustre bien par la distribution en L des valeurs de  $G_{ST}$  de 155 marqueurs AFLP (Mariette *et al.*, 2002).

#### **A.3.2.4 Où se situent les chênes sessile et pédonculé dans le processus de spéciation ?**

##### **A.3.2.4.1 Espèce ou sous-espèce ?**

Les débats scientifiques autour du concept d'espèce mentionnés dans le premier chapitre, s'illustrent très bien avec le cas du chêne. En effet de nombreux auteurs ont remis en question leur statut d'espèce compte tenu de leur possible hybridation (Burger, 1975 ; Kleinschmit et Kleinschmit, 2000). Ainsi, si l'on reprend les principaux concepts :

*Le chêne sessile et le chêne pédonculé sont ils deux espèces typologiques ?*

Ce concept taxonomiste mis en place par Linné permet, comme nous l'avons vu, de bien distinguer les deux espèces sur la base d'analyse multivariées. Néanmoins, la distribution bimodale des caractères avec la présence d'un chevauchement dans les phénotypes intermédiaires ne permet pas la classification de certains individus.

*Le chêne sessile et le chêne pédonculé sont-ils deux espèces biologiques ?*

Si on applique la définition stricte de Mayr, le chêne sessile et le chêne pédonculé ne sont pas deux espèces car ils s'hybrident. On parlerait plutôt de sous-espèces. Cependant, en dépit d'une interfertilité mise en évidence par l'étude de la reproduction dans les peuplements naturels, les deux formes parentales prédominent dans les forêts.

*Le chêne sessile et le chêne pédonculé sont-ils deux espèces écologiques ?*

Cette notion semble mieux appropriée au cas des chênes car les deux espèces possèdent des exigences écologiques bien distinctes. Cependant comme nous l'avons noté, la plupart des peuplements comprennent les deux espèces.

*Le chêne sessile et le chêne pédonculé sont-ils un complexe d'espèce ?*

Ce concept fut proposé par Pernès en 1984. Il définit le complexe d'espèce comme étant un ensemble de taxons qui échangent des gènes directement ou par espèces relais. Cette définition correspond très bien à la situation du chêne : différences morphologiques et écologiques associées à une hybridation importante et une faible différenciation génétique.

L'espèce, comme nous l'avons déjà dit, est un aboutissement. Il n'est donc pas surprenant qu'il soit difficile de classer le chêne sessile et le chêne pédonculé comme deux espèces distinctes si le processus de spéciation est en cours. Si les observations ne donnent pas de réponse claires pour classer les chênes en espèces ou sous-espèces, les réponses en terme de processus de spéciation le sont encore moins.

#### A.3.2.4.2 Mécanisme de spéciation

Trois hypothèses de mécanisme de spéciation pourraient expliquer la présence de différences phénotypiques entre le chêne sessile et le chêne pédonculé associées à un flux de gènes important et une faible différenciation génétique (Kleinschmit et Kleinschmit, 2000).

1- Le chêne sessile et le chêne pédonculé dérivent d'un même ancêtre commun et évoluent vers l'établissement de deux espèces distinctes.

Dans cette première hypothèse, la spéciation serait le résultat de la sélection divergente. Des indicateurs en faveur de cette hypothèse sont les exigences écologiques distinctes qui engendrent des adaptations différentielles entre les deux espèces et la présence d'hybrides contre-sélectionnés.

2- Le chêne sessile et le chêne pédonculé sont à l'équilibre entre la convergence causée par le flux de gènes et la divergence causée par l'adaptation aux différentes conditions écologiques.

Cette deuxième hypothèse est soutenue par la présence de flux de gènes entre les deux espèces et l'observation d'hybrides localement bien adaptés associés à une contre-sélection des hybrides dans les sites extrêmes.

3- Le chêne sessile et le chêne pédonculé étaient deux espèces bien séparées qui sont en train de converger en une seule espèce.

Dans cette troisième hypothèse, la spéciation est le résultat de la sélection stabilisatrice. Un élément en sa faveur est la présence de flux de gènes entre les espèces, par contre un point défavorable est l'observation de barrière post-reproductive qui élimine les hybrides.

Il semble néanmoins que le scénario ne soit pas aussi manichéen que les trois scénarios évoqués précédemment. Tout d'abord, les éléments fossiles semblent indiquer que les espèces existent depuis le tertiaire. Comment expliquer dès lors que les espèces se soient maintenues en dépit des flux de gènes interspécifiques? Cette question amène à s'interroger sur le rôle des flux de gènes interspécifiques chez les chênes.

#### A.3.2.4.3 Rôle de l'hybridation interspécifique dans l'évolution des chênes

L'analyse comparative de la distribution géographique des variants de l'ADN chloroplastique révèle des stratégies particulières de colonisation propre au chêne sessile et au chêne pédonculé. En effet, chaque fois que le chêne sessile et le chêne pédonculé coexistent dans la même forêt, ils partagent le même variant chloroplastique (Petit *et al.*, 1997). L'explication la plus probable serait que l'homogénéisation des génomes chloroplastiques entre les deux espèces qui partagent une forêt résulte d'hybridations récurrentes et unidirectionnelles (Kremer *et al.*, 2002b). Cette hypothèse est renforcée par la mise en évidence d'une hybridation asymétrique entre les deux espèces. L'hybridation préférentielle du père sessile vers la mère pédonculé permettrait au chêne sessile de s'introduire progressivement dans le chêne pédonculé par back-cross successifs. Ainsi, le flux interspécifique par pollen peut être vu comme une opportunité pour le chêne sessile (espèce post-pionnière) d'être entraîné par le chêne pédonculé (espèce pionnière) pour occuper des nouveaux milieux.

Ces processus de colonisation par hybridation répétée au cours de l'histoire évolutive des deux chênes au cours du quaternaire a conduit à l'homogénéisation progressive des génomes, excepté au niveau des régions génomiques soumises à la sélection disruptive qui maintient des phénotypes parentaux adaptés à des conditions de milieux particuliers. Pour résumer, il faut donc suffisamment de divergence génétique pour pouvoir coloniser ces espaces particuliers et pas trop pour toujours autoriser l'hybridation. C'est cet équilibre qui serait aujourd'hui atteint.

Chapitre B  
PROBLEMATIQUE

## CHAPITRE B PROBLEMATIQUE

### B.1 INTRODUCTION ET RAPPELS

Le chêne sessile et le chêne pédonculé sont les deux espèces de feuillus majoritaires en Europe. Ils partagent les mêmes zones géographiques (espèces sympatriques) en ayant néanmoins des exigences écologiques différentes. On a cependant montré que la plupart des peuplements possèdent des zones de mélange à la limite des exigences écologiques de chacune des deux espèces facilitant l'hybridation. L'hybridation entre le chêne sessile et le chêne pédonculé est asymétrique (Bacilieri *et al.*, 1996). Le sens préférentiel du croisement va du père sessile vers la mère pédonculé.

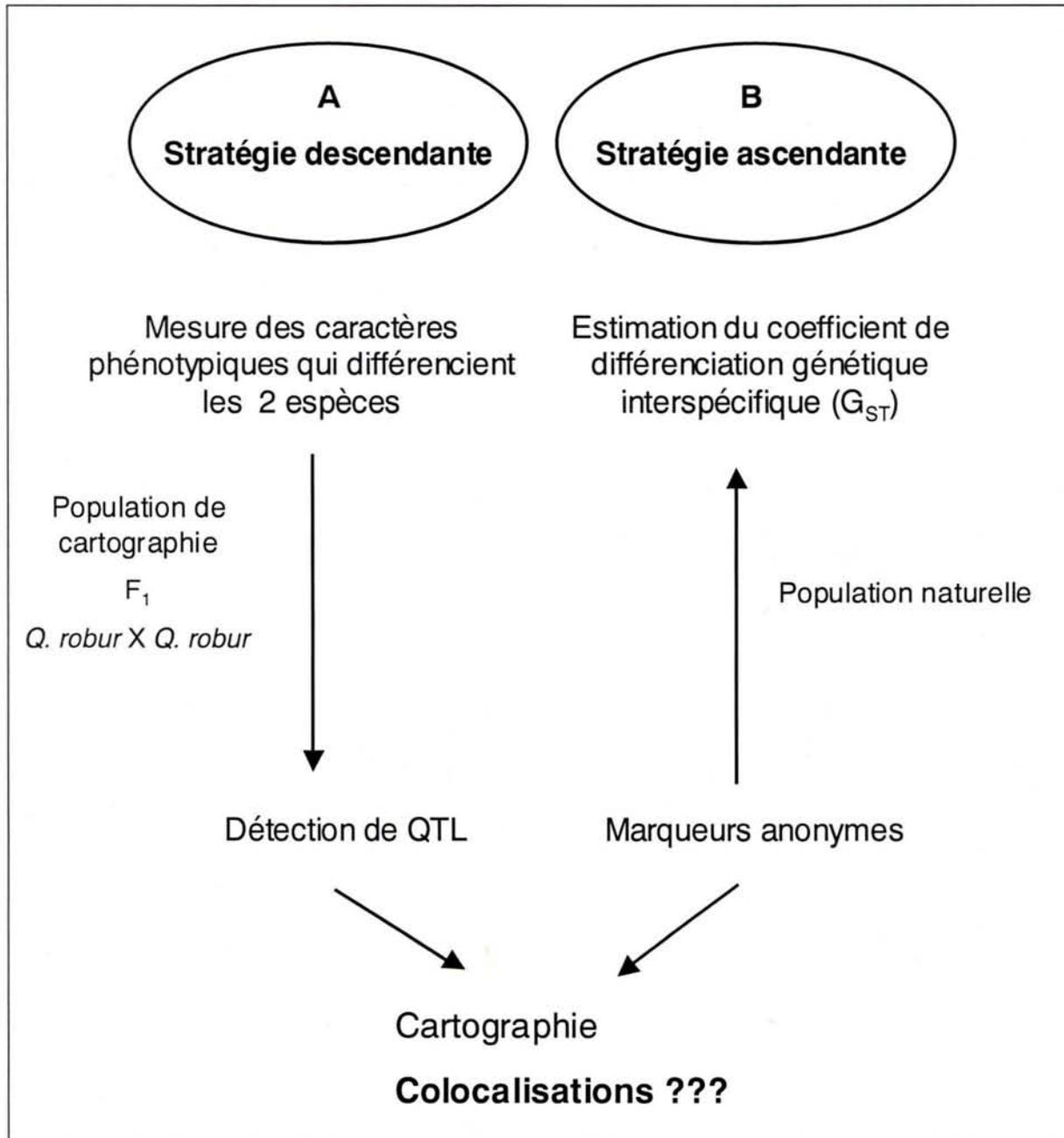
La différenciation écologique des deux espèces s'accompagne d'une différenciation morphologique. Plusieurs critères morphologiques de la feuille et du gland manifestent des différences quantitatives entre les deux espèces cependant des phénotypiques intermédiaires ne peuvent permettre le classement de l'ensemble des individus.

Les marqueurs génétiques pourraient permettre la classification de tous les individus. Malheureusement, les études menées à ce jour n'ont pu mettre en évidence de marqueurs spécifiques de l'espèce. Seules des variations en terme de fréquence allélique ont été obtenues. L'hybridation facilitée par la sympatrie entre les deux espèces, est certainement la cause de cette faible différenciation génétique.

Au vu de ces résultats, les chênes sessile et pédonculé ne peuvent être classés en tant qu'espèces biologiques. La meilleure façon de les classer pour le moment est de les définir comme un complexe d'espèces.

En adoptant deux types d'approche, cette thèse a pour but d'étudier la distribution des régions génomiques qui différencient les deux espèces.

**Figure 9** : Stratégies adoptées pour étudier la différenciation génétique entre le chêne sessile et le chêne pédonculé.



Concrètement les questions posées sont :

- Où se situent dans le génome les zones impliquées dans la différenciation ?
- Sont-elles réunies ou dispersées ?
- Sont-elles impliquées dans l'expression d'un ou de plusieurs caractères ?
- Quelles sont leurs effets sur ces caractères ?

## **B.2 STRATEGIES DEVELOPPEES**

Nous allons utiliser la carte génétique du chêne pédonculé (Barreneche *et al.*, 1998) pour localiser des régions génomiques et des marqueurs anonymes impliqués dans la différenciation génétique entre le chêne sessile et le chêne pédonculé.

Deux stratégies ont été suivies (Figure 9)

### **B.2.1 LA STRATEGIE DESCENDANTE : DU PHENOTYPE AU GENOME**

Nous savons qu'il existe des différences phénotypiques entre les espèces, qui sont couramment utilisées pour leur distinction. En recherchant les QTL contrôlant ces caractères phénotypiques, nous devrions localiser les régions génomiques impliquées dans la différenciation. Cette approche descendante a cependant certaines limites inhérentes à la détection de QTL. D'une part, la sous-estimation du nombre de QTL liée au fait que la détection est réalisée dans une seule famille, et d'autre part, une imprécision de la localisation des QTL liée à la détection statistique des QTL (grands intervalles de confiance).

### **B.2.2 LA STRATEGIE ASCENDANTE : DU GENOME AUX REGIONS DISCRIMINANTES**

Elle consiste à faire un screening du génome pour la différenciation interspécifique. L'idée est de prospecter le génome de manière systématique à l'aide des différentes techniques de marquage

disponible et d'estimer pour chaque marqueur un coefficient de différenciation interspécifique classiquement utilisé en génétique des populations ( $G_{ST}$ ). La méthode suppose bien évidemment que chaque marqueur soit cartographié sur une carte génétique et ait fait l'objet d'analyse des fréquences alléliques à l'intérieur de chaque espèce. Elle souffre également d'un certain nombre de contraintes. D'une part, elle fait l'hypothèse qu'il y ait un effet d'entraînement d'un gène sélectionné sur les fréquences alléliques d'un marqueur voisin. Si cet effet d'entraînement est faible, il faudra faire un screening avec une très haute densité de marqueurs sous peine de ne pas piéger les régions génomiques différenciées. Comme la détection de QTL, la méthode risque également de sous-estimer le nombre de régions à cause d'une densité incomplète de marqueurs. Si l'effet d'entraînement est uniquement dû à la liaison génétique, il risque d'être de courte distance compte tenu de la très forte allogamie et des tailles de populations importantes chez les chênes. Par contre, si l'effet d'entraînement est dû à un effet sélectif portant sur plusieurs gènes voisins (gènes coadaptés), il risque de se porter sur une plus longue distance.

La synthèse de ces deux approches sera faite sous forme d'analyse de colocalisation des QTL informatifs (stratégie descendante) et de marqueurs impliqués dans la différenciation (stratégie ascendante) sur la même carte génétique.

Chapitre C  
DETECTION DE QTL

## CHAPITRE C DETECTION DE QTL

La stratégie descendante a pour objectif de localiser les QTL qui contrôlent des caractères phénotypiques fortement impliqués dans la différenciation phénotypique des deux espèces. La sélection de ces caractères s'est faite sur la base de données antérieures relatives à la morphologie et à l'écologie des espèces. Pour résumer, notre sélection a porté sur trois catégories de caractères :

- **Morphologie foliaire** : Il existe une très grande bibliographie sur le sujet (Kremer *et al.*, 2002a) montrant, entre autres, que la pilosité des feuilles, leur nervation, la longueur du pétiole sont parmi les caractères les plus distinctifs des deux espèces.
- **Ecophysiologie** : la réponse des plantes à une contrainte hydrique (qu'il s'agisse d'excès (anoxie) ou de manque d'eau (sécheresse) a fait l'objet de très nombreux travaux. Pour résumer les plants de chêne sessile sont plus résistants à la sécheresse et plus sensibles à l'hydromorphie que ceux de chêne pédonculé (Granjean et Sigaud, 1987 ; Ponton *et al.*, 2001).
- **Croissance en hauteur** : autant les pépiniéristes que les scientifiques ont remarqué que la croissance juvénile du chêne pédonculé est plus importante (et plus polycyclique) que celle du chêne sessile dans le jeune âge (Ponton *et al.*, 2002).

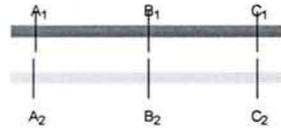
En plus de la détection proprement dite des QTL, notre analyse portera sur la distribution au sein du génome de ces QTL, en s'interrogeant sur leur regroupement préférentiel ou pas. A ce titre, leur distribution sera utilement comparée à celle de QTL de caractères non impliqués dans la différenciation. C'est la raison pour laquelle l'analyse a été complétée par une détection de QTL impliqués dans des caractères qui ne manifestent aucune différenciation interspécifique. Cette deuxième catégorie de QTL devant constituer en quelque sorte les QTL témoins de l'hypothèse nulle (distribution aléatoire des QTL dans le génome). Nous avons autant que possible essayé de garder les mêmes catégories de caractères pour les caractères non différenciés.

- **Morphologie foliaire** : les caractères de taille de feuilles (longueur et largeur du limbe) (Kremer *et al.*, 2002a)
- **Ecophysiologie** : Les dates de débourrement des bourgeons, bien que très variables à l'intérieur de chaque espèce ne manifestent aucune différence interspécifique (Bacilieri *et al.*, 1995)
- **Aptitude au bouturage** : les deux espèces peuvent se reproduire par multiplication végétative. Alors que l'on observe à l'intérieur de chaque espèce une forte variation de l'aptitude au bouturage entre arbres, en moyenne les deux espèces ne se différencient pas pour ce caractère (Spethmann, 1986).

Pour chacune de ces catégories de caractères, les QTL seront recherchés dans une famille en ségrégation. La détection de QTL étant souvent affectée par une forte imprécision (grands intervalles de confiance) et une forte variation (selon les milieux et selon les saisons), nos observations ont toujours porté sur des données pluriannuelles et sur deux plantations expérimentales. En plus, les données ont été recueillies sur des copies végétatives des descendants de la famille de cartographie, multipliant ainsi le nombre d'observations.

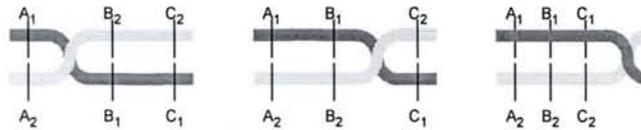
La première partie de ce chapitre est consacrée au principe de construction d'une carte génétique, ainsi qu'aux différentes méthodes statistiques de détection de QTL. Puis dans un second temps, les résultats de la détection de QTL seront présentés.

## Encadré 1 : Calcul du taux de recombinaison



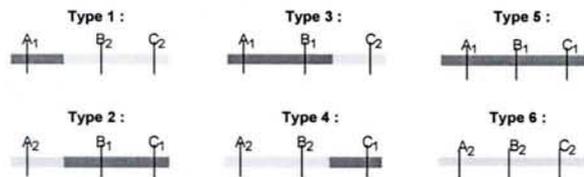
Paire de chromosomes homologues d'un individu diploïde et hétérozygote au locus A, B, C.

Méiose : appariement des chromosomes homologues et échanges de fragments d'ADN (crossing-over)



Remarque : Les doubles crossing-over ne sont pas considérés dans cet exemple.

Les gamètes haploïdes formés aux trois loci seront soit de type parental (absence de crossing-over) soit de type recombiné (crossing-over)



Le taux de recombinaison ( $r$ ) entre deux loci est la proportion de gamètes recombinés par rapport au nombre total de gamètes formés ( $N$ ) :

$r$	A	B
B	$\frac{\text{type1} + \text{type2}}{N}$	
C	$\frac{\text{type1} + \text{type2} + \text{type3} + \text{type4}}{N}$	$\frac{\text{type3} + \text{type4}}{N}$

Le taux de recombinaison  $r$  est relié à la distance séparant deux loci. Plus cette distance est grande plus la probabilité d'observer des crossing over est grande. Il y a indépendance entre les loci lorsque le nombre de gamètes de type parental est égal au nombre de gamètes de type recombiné ( $r=0.5$ ).

## C.1 MATERIEL ET METHODES

### C.1.1 CONSTRUCTION D'UNE CARTE GENETIQUE

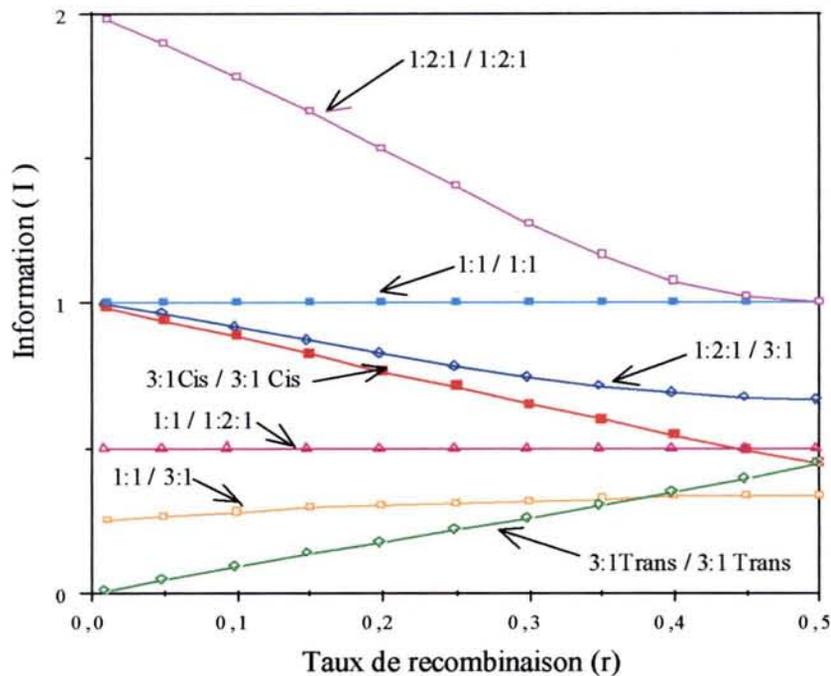
#### C.1.1.1 Principe

L'établissement d'une carte génétique nécessite :

- (1) la création d'une descendance en ségrégation,
- (2) la caractérisation moléculaire de la descendance et des parents,
- (3) l'utilisation d'outils d'analyse statistique et de méthodes de calculs mathématiques pour estimer les distances génétiques entre les marqueurs moléculaires d'un même groupe de liaison, puis de les ordonner à l'intérieur de chaque groupe.

La technique de cartographie est abondamment détaillée par de Vienne (1997). Une fois l'ensemble des descendants typés à l'aide de marqueurs moléculaires, un premier test de conformité de ségrégation mendélienne (test du  $\chi^2$ ) est réalisé pour chaque marqueur. Un  $\chi^2$  significatif indique une distorsion de ségrégation qui peut avoir une origine soit biologique, comme la liaison du marqueur avec un locus d'incompatibilité, soit statistique, et dans ce cas résultant d'un échantillonnage trop faible. Le taux de recombinaison entre chaque marqueur peut ensuite être calculé comme le rapport du nombre de gamètes recombinés au nombre de gamètes total (Encadré 1). Le taux de recombinaison  $r$  peut aussi être estimé par la méthode du maximum de vraisemblance. Cette estimation dépend de l'effectif de la descendance, du type de descendance, et du type de marqueur. Avant de déclarer deux marqueurs liés, il est nécessaire d'effectuer un test de liaison. Il s'agit alors de tester l'indépendance de ségrégation entre les marqueurs sous l'hypothèse nulle  $H_0$ : " $r=0.5$ " (il y a autant de gamètes parentaux que de gamètes recombinés). Le test de liaison peut être réalisé en utilisant le  $\chi^2$  ou le rapport de vraisemblance appelé "LOD score" (Log of the odds ratio, décrit par Morton en 1955). Connaissant les taux de recombinaison, il est ensuite possible de calculer la distance génétique entre

**Figure 10 :** Evolution de l'information du taux de recombinaison (inverse de la variance du taux de recombinaison) en fonction du type de ségrégation pour des couples de marqueurs. Les informations sont calculées par rapport à la situation observée dans un croisement backcross pour laquelle les allèles au marqueur ségrègent dans des proportions  $\frac{1}{2} : \frac{1}{2}$ .



Equations des fonctions d'information  $I(r)$  correspondant à différentes configurations de ségrégation de couples de marqueurs dans différents types de croisements :

Croisement de type backcross :

$I(r)_{1:1/1:1} = 1/(r(1-r))$  (Allard, 1956)

Croisement de type  $F_2$  :

-Cas de 2 marqueurs codominants (meilleure situation)

$I(r)_{1:2:1/1:2:1} = 2(1-3r+3r^2)/r(1-r)(1-2r+2r^2)$  (Allard, 1956)

-Cas de 2 marqueurs dominants en position Cis

$I(r)_{3:1\text{ cis} / 3:1\text{ cis}} = 2(3-4r+2r^2)/(r(2-r)(3-2r+r^2))$  (Ritter *et al.*, 1990)

-Cas de 2 marqueurs dominants en position Trans

$I(r)_{3:1\text{ trans} / 3:1\text{ trans}} = 2(1+2r^2)/(2+r^2)(1-r^2)$  (Allard, 1956)

-Cas d'un marqueur codominant et d'un marqueur dominant :

$I(r)_{1:2:1/3:1} = 2+(1-r)^2/r(2-r)+(1-2r)^2/2(1-r+r^2)+r^2/(1-r^2) + (1-2r)^2/2r(1-r)$  (Allard, 1956)

Croisement entre deux parents hétérozygotes (stratégie double pseudo-testcross, cf encadré 2))

En plus des 5 configurations présentées au dessus il est possible de trouver les cas suivants :

-Cas d'un marqueur en configuration backcross et d'un marqueur codominant

$I(r)_{1:1/1:2:1} = 1/2r(1-r)$  (Plomion *et al.*, 1997)

-Cas d'un marqueur en configuration backcross et d'un marqueur dominant

$I(r)_{1:1/3:1} = (1+2r-2r^2)/2r(1-r^2)(2-r)$  (Ritter *et al.*, 1990)

les marqueurs. Deux types de distance sont généralement utilisées : la distance de Haldane (1919) qui considère que les crossing-overs sont indépendants, et la distance de Kosambi (1944) qui tient compte des interférences entre les crossing-overs. Ayant estimé les distances entre tous les marqueurs 2 à 2, il est alors possible de les ordonner à l'intérieur des groupes de liaison. Des tests 3 points et multipoints sont alors utilisés pour préciser leur ordre.

La précision de l'estimation du taux de recombinaison  $r$  diffère selon la descendance utilisée et les couples de marqueurs pris en compte. Elle peut être mesurée par une fonction  $I(\hat{r})$ , appelée fonction d'information (Fisher, 1921) qui correspond à l'inverse de la variance de l'estimateur  $\hat{r}$  du taux de recombinaison  $r$  :  $I(\hat{r}) = 1/V(\hat{r})$ . On peut ainsi construire dans une descendance donnée, une fonction d'information pour chaque couple de marqueurs et les comparer entre elles (Allard, 1956 ; Ritter *et al.*, 1990, Ritter et Salamini, 1996). La figure 10 représente l'évolution de l'information du taux de recombinaison en fonction du type de ségrégation pour différents couples de marqueurs. A titre d'exemple, les couples de marqueurs les plus informatifs sont ceux qui ségrègent en proportion 1:2:1 et 1:2:1 alors que des couples de marqueurs qui ségrègent en proportion 3:1 et 1:1 seront moins informatifs que des marqueurs qui ségrègent en proportion 1:1 et 1:1.

### C.1.1.2 Cas particulier des arbres forestiers

Chez les arbres forestiers contrairement aux plantes annuelles, les populations sont fortement hétérozygotes car peu sélectionnées. C'est pourquoi il est possible de construire une carte génétique à partir d'une famille  $F_1$ . L'analyse de la ségrégation dans une descendance issue de parents hétérozygotes est possible grâce à la stratégie du double pseudo-testcross (Grattapaglia *et al.*, 1994) qui consiste à décomposer le croisement en 2 testcross. Un marqueur est exploité s'il est hétérozygote chez un parent et homozygote chez l'autre. Ainsi une carte génétique est construite pour chacun des 2 parents (Encadré 2). Les marqueurs ponts qui ségrègent chez les deux parents permettent ensuite d'identifier et d'orienter les groupes homologues. Enfin il est aussi possible de construire une carte

## Encadré 2 : La Stratégie double pseudo-testcross

La stratégie double pseudo-testcross permet de construire deux cartes parentales à partir de marqueurs dominants et/ou codominants qui ségrègent dans une population F<sub>1</sub>. La stratégie consiste à décomposer le croisement en deux testcross. Ainsi les marqueurs informatifs pour la construction de la carte femelle seront hétérozygotes chez la femelle et homozygotes chez le mâle (et vice versa).

### 1) Cas des marqueurs dominants

Cas	Profils parentaux		génotypes parentaux		Profils dans la descendance F1	ségrégation	Carte génétique
	♀	♂	♀	♂			
1	—		A <sub>1</sub> /A <sub>0</sub>	A <sub>0</sub> /A <sub>0</sub>	—	1:1	♀
2		—	A <sub>0</sub> /A <sub>0</sub>	A <sub>1</sub> /A <sub>0</sub>	—	1:1	♂
3	—	—	A <sub>1</sub> /A <sub>0</sub>	A <sub>1</sub> /A <sub>0</sub>	—	3:1	♀ et ♂ mais peu informatifs

### 2) Cas des marqueurs codominants

Cas	Profils parentaux		génotypes parentaux		Profils dans la descendance F1	Ségrégation	Carte génétique
	♀	♂	♀	♂			
4	==	—	A <sub>1</sub> /A <sub>2</sub>	A <sub>1</sub> /A <sub>1</sub>	== —	1:1	♀
5	—	==	A <sub>1</sub> /A <sub>1</sub>	A <sub>1</sub> /A <sub>2</sub>	== —	1:1	♂
6	==	==	A <sub>1</sub> /A <sub>2</sub>	A <sub>1</sub> /A <sub>2</sub>	— == —	1:2:1	♀ et ♂
7	==	==	A <sub>1</sub> /A <sub>2</sub>	A <sub>1</sub> /A <sub>3</sub>	== == — ==	1:1:1:1	♀ et ♂
8	==	==	A <sub>1</sub> /A <sub>2</sub>	A <sub>3</sub> /A <sub>4</sub>	== — == ==	1:1:1:1	♀ et ♂

Les marqueurs dominants de configuration 3 ainsi que les marqueurs codominants de configuration 6 ne peuvent être décomposés sous forme de backcross. Ils peuvent être utilisés pour trouver les groupes homologues entre le mâle et la femelle ou servir par la suite de marqueurs pont pour construire la carte consensus.

consensus qui regroupe l'ensemble des marqueurs, tout en sachant que le taux de recombinaison est beaucoup moins bien estimé lorsque l'on confronte des marqueurs qui ségrègent en proportion 1:1 avec des marqueurs qui ségrègent en proportion 3:1 (Figure 10).

### **C.1.1.3 Caractéristiques de la carte génétique du chêne pédonculé**

En 1998, Barreneche *et al.* ont publié la première carte génétique du chêne pédonculé à partir de 94 descendants issus du croisement intraspécifique entre deux parents pédonculés. L'utilisation de la stratégie du double pseudo-testcross a permis d'obtenir deux cartes parentales comportant un total de 307 marqueurs (271 RAPD, 18 microsatellites, 1 minisatellite, 10 SCAR, 6 isozymes, 1 5SrDNA) repartis sur douze chromosomes, ce qui correspond au nombre haploïde de chromosomes chez le chêne. La longueur totale de la carte femelle estimée, à 1192 cM, et celle du mâle, estimée à 1235 cM, indiquent que les deux cartes parentales couvrent entre 85 et 90 % du génome. La carte génétique du chêne pédonculé a depuis été complétée et sera bientôt publiée. Elle comporte aujourd'hui environ 500 marqueurs. Les marqueurs rajoutés sont essentiellement de type AFLP et microsatellites.

### **C.1.1.4 Construction de la carte jalon du chêne pédonculé, support de la détection de QTL.**

La puissance de détection des QTL étant très dépendante de l'effectif de la descendance, 181 descendants supplémentaires ont été typés pour la construction de la carte jalon (soit 275 individus au total). Compte tenu de la faible répétabilité des marqueurs RAPD et de la mise au point de la technique AFLP dans notre laboratoire il nous a semblé plus judicieux de construire la "carte QTL" en utilisant des marqueurs dominants de type AFLP qui fournissent un grand nombre de marqueurs, associés à des marqueurs codominants de type microsatellites et SCAR qui permettent de faire le pont entre les deux cartes parentales. Les différents marqueurs ont été choisis de sorte qu'ils soient bien répartis sur le génome. Les deux cartes parentales ont été construites avec le logiciel Mapmaker

(Lander *et al.*, 1987) en utilisant la stratégie backcross comme indiquée dans l'encadré 2. Elles contiennent des marqueurs AFLP, microsatellites, SCAR et RAPD. Dans une première étape les marqueurs ont été réunis au sein des différents groupes de liaison en se fixant un LOD supérieur à 7, puis ils ont été ordonnés à l'intérieur de chaque groupe en se fixant un seuil de LOD égal à 3. Des détails sur la construction de la carte et sur sa composition sont donnés dans l'article en annexe 1. La carte femelle est composée de 75 marqueurs et couvre 902cM, avec en moyenne un marqueur tous les 14 cM. La carte mâle contient 72 marqueurs et couvre 933 cM, avec en moyenne un marqueur tous les 15.5 cM. La taille de la carte indiquée dans l'article est légèrement différente (femelle=890cM et mâle=929cM) car depuis la soumission de l'article une quarantaine d'individus qui n'avaient pas été génotypés ont été complétés.

## **C.1.2 DETECTION DE QTL**

### **C.1.2.1 Principales méthodes de détection de QTL.**

La détection de QTL est fondée sur la recherche systématique de déséquilibre de liaison physique entre des locus marqueurs et des locus contrôlant des caractères quantitatifs. Pour cela il faut disposer d'une descendance en ségrégation, déterminer pour chaque descendant le génotype d'un ensemble de locus marqueurs, mesurer la valeur du caractère quantitatif étudié sur chaque descendant et mettre en oeuvre des méthodes biométriques pour rechercher les locus marqueurs dont le génotype est corrélé au caractère et estimer les paramètres génétiques des QTL détectés.

Différentes approches statistiques permettent de rechercher les marqueurs corrélés aux caractères. La plus simple est l'analyse de variance qui consiste à rechercher une relation entre le génotype de chaque locus marqueur et la valeur du caractère quantitatif considéré. La régression multiple est une autre approche qui permet de relier la variabilité du caractère au polymorphisme de plusieurs marqueurs à la fois. L'avantage de la détection de QTL en prenant les marqueurs individuellement est sa simplicité. Ces analyses sont robustes aux écarts à la normalité et une carte génétique saturée n'est

pas nécessaire. Cependant la limitation principale réside dans l'erreur portant sur la position et l'effet des QTL lorsque la densité de marqueurs est faible (de Vienne et Causse, 1997). Beaucoup plus puissantes, les méthodes de cartographie d'intervalle (interval mapping, Lander et Botstein, 1989) permettent de positionner les QTL dans l'intervalle de marqueurs adjacents. Le modèle de l'interval mapping fait l'hypothèse que sur un chromosome, un seul QTL est situé entre deux marqueurs liés par un taux de recombinaison  $r$ . Dans une population backcross composée de  $n$  individus, Lander et Botstein (1989) proposent d'utiliser le modèle linéaire suivant pour tester la position d'un QTL dans un intervalle de marqueurs  $i$  et  $i+1$  :

$$y_j = \mu + bg_j + e_j \quad j=1,2, \dots,n$$

Où  $y_j$  est la valeur du caractère de l'individu  $j$ ,  $\mu$  est la moyenne du modèle,  $b$  est l'effet du QTL (différence entre la moyenne du caractère des individus homozygotes et hétérozygotes,  $\mu_{QQ} - \mu_{Qq}$ ),  $g_j$  est une variable indicatrice de valeur 0 ou 1 avec une probabilité dépendante du génotype des marqueurs flanquants et de la distance de la position à tester de ces marqueurs et  $e_j$  est l'erreur résiduelle du modèle. Ce modèle statistique est ensuite utilisé pour construire un test de rapport de vraisemblance basé sur les hypothèses :  $H_0 : b = 0$  et  $H_1 : b \neq 0$ . La présence d'un QTL est ensuite

estimée en calculant en chaque position du chromosome le LOD score qui est le logarithme décimal

du rapport de vraisemblance :  $LOD = \log_{10} \frac{L(\hat{\mu}, \hat{b}, \hat{\sigma}^2)}{L(\hat{\mu}, 0, \hat{\sigma}^2)}$  où  $L(\hat{\mu}, \hat{b}, \hat{\sigma}^2)$  sont les estimateurs du maximum

de vraisemblance de  $\mu$ ,  $b$  et  $\sigma^2$  sous l'hypothèse alternative  $H_1$  ( $b$  étant l'effet du QTL),  $L(\hat{\mu}, 0, \hat{\sigma}^2)$

sont les estimateurs du maximum de vraisemblance sous  $H_0$ . Lorsque le LOD score dépasse un seuil  $C$

déterminé, on déclare la présence d'un QTL. Les deux principales limites du modèle "interval

mapping" sont d'une part, qu'il est sensible à des distributions non normales des caractères quantitatifs

et d'autre part, qu'il reste un modèle à un "QTL". L'effet et la position du QTL détecté seront biaisés

par le bruit de fond engendré par les autres QTL. En effet, lorsque deux "QTL" ségrégent l'un au

voisinage de l'autre, et lorsqu'ils agissent dans la même direction, un seul "QTL" fantôme peut être

détecté au milieu des deux "QTL" (Martinez et Curnow, 1992), au contraire si les deux QTL ont le

même effet mais de signe opposé, aucun QTL ne sera détecté (Jansen et Stam, 1994). Afin d'avoir une estimation plus précise de QTL multiples, des méthodes qui combinent la cartographie d'intervalle et la régression linéaire ont été proposées (Jansen, 1993 ; Zeng , 1994). Cela revient à prendre les autres QTL présents dans le génome, représentés par les marqueurs qui leur sont proches, comme co-facteur dans le modèle, pour réduire la part de variance résiduelle induite par leur ségrégation.

### C.1.2.2 Cas particulier des arbres forestiers

Le nombre de QTL détectés et l'importance de leur effet sont propres à la population analysée. Compte tenu de la maturité sexuelle tardive des arbres forestiers, la détection de QTL est surtout menée dans des descendance  $F_1$ . Une ségrégation au QTL n'est alors observée que si l'un des parents est hétérozygote pour ce locus. Malgré le taux d'hétérozygotie relativement élevé des arbres (Hamrick *et al.*, 1992), un croisement unique ne révèle qu'une partie des QTL intervenant pour un caractère donné. La détection de QTL dans plusieurs familles augmente alors la probabilité de voir co-ségréger un marqueur et un caractère quantitatif . Ainsi, des plans de croisements factoriels ou diallèles ont été réalisés pour plusieurs espèces (mélèze, peuplier, pin maritime, eucalyptus). Afin de maximiser l'efficacité de détection de QTL il est aussi possible chez certains arbres d'avoir recours au clonage des descendants (peuplier, eucalyptus, chêne). La répétition des individus permet de diminuer la variance environnementale associée aux caractères, et donc d'améliorer l'estimation de la variance génétique additive expliquée par l'association marqueur/caractère (Bradshaw et Foster, 1992).

**Tableau 4** : Description de la population de cartographie et des dispositifs expérimentaux utilisés pour la détection de QTL.

	Dispositif 1	Dispositif 2	Dispositif 3
Type de test	Parc à pied mères	Test clonal	Test clonal
Localisation	Pierroton 44°44' N 00°41' W	Bourran 44°20' N 00°24' E	Bourran 44°20' N 00°24' E
Année de bouturage		Eté 1997	Eté 1998
Date d'installation	Printemps 1995	Hiver 1998	Printemps 2000
Espacement	1.5 × 1.5	1.5 × 4	1.5 × 4
Nombre de descendants (clones)	278	174	207
Nombre de boutures		1080	2196
Nombre moyen de boutures par clone		6.2	10.6
Nombre de blocs		36	183
Nombre de clones par bloc		30	12
Caractères mesurés	-Débourrement en 1999 -Aptitude au bouturage en 1997,1998, 1999, et 2000.	-Morphologie foliaire en 1999 -Débourrement en 2000 -Efficience d'utilisation de l'eau en 2000 et 2001	-Morphologie foliaire* en 2000 -Débourrement en 2002 -Croissance en 2002 -Efficience d'utilisation de l'eau en 2002

\* les feuilles ont été prélevées en pépinière avant la plantation sur le terrain.

### C.1.2.3 Détection de QTL chez le chêne pédonculé

#### C.1.2.3.1 Présentation de la famille en ségrégation et des dispositifs de mesure

##### C.1.2.3.1.1 La famille

Une famille intraspécifique de *Q. robur* a été obtenue en croisant deux arbres adultes. La femelle (3P) est localisée sur le site de l'INRA de Pierroton et le mâle (A4) à Arcachon.

Le croisement a permis d'obtenir plus de 400 glands dont 278 ont fourni des plantules. Un sous ensemble de 94 arbres a été utilisé pour la construction des cartes génétiques parentales (Barreneche *et al.* 1998) et les 184 pleins frères supplémentaires ont été génotypés pour la cartographie de QTL.

Après 3 ans passés en pépinière, les 278 descendants ont été transplantés dans un parc à clone.

Après deux ans de transplantation, à la fin février, chaque plant a été recépé dans le but de les maintenir juvénile. Le recépage est depuis réalisé chaque hiver. Les rejets de souches sont ensuite utilisés au mois de juin pour réaliser des boutures. En moyenne 20 boutures par descendant sont réalisées chaque année dont plus de 60 % sont viables. Les boutures âgées de 2 ans peuvent ensuite être installées sur le terrain pour l'évaluation phénotypique en vue de la détection de QTL.

##### C.1.2.3.1.2 Les dispositifs

Les caractères phénotypiques étudiés pour la détection de QTL ont été mesurés dans trois types de dispositifs dont la description est donnée au tableau 4.

-Le dispositif 1, correspond au parc à pied-mères localisé à Pierroton. Il rassemble 278 descendants issus du croisement contrôlé entre 3P et A4. Le bouturage de chaque descendant en plusieurs copies a permis l'installation de 2 tests clonaux (dispositif 2 et 3) dans une unité expérimentale de l'INRA située à Bourran dans le Lot et Garonne.

-Le dispositif 2, mis en place au printemps 1999 est composé de 1080 boutures issues de la saison de bouturage 1997. Il comprend 174 clones répétés en moyenne 6.2 fois. Les copies végétatives sont réparties dans 36 blocs incomplets comprenant chacun 30 clones en simple copie.

-Le dispositif 3, mis en place au printemps 2000 est composé de 2196 boutures réalisées en 1998 à partir de 207 clones (en moyenne 10.6 copies par clone). Les copies végétatives sont réparties dans 183 blocs incomplets comprenant chacun 12 clones en simple copie.

Les deux dispositifs clonaux sont distants d'une centaine de mètres et sont installés sur des sols de composition identique.

### **C.1.2.3.2      Analyses statistiques préliminaires**

Le modèle de base de la génétique quantitative décompose la valeur phénotypique d'un individu de la façon suivante :

$$P = \mu + G + E$$

Avec  $P$  la valeur phénotypique qui est le résultat de la mesure effectuée sur un individu,  $\mu$  la moyenne générale des valeurs phénotypiques dans la population considérée,  $G$  la valeur génétique qui représente l'effet moyen du génotype sur un caractère donné et  $E$  la valeur environnementale qui est une somme de facteurs de milieux contrôlés (blocs, observateur, date de mesure) et non contrôlés (erreur de mesure, micro-milieu).

#### **C.1.2.3.2.1      Ajustement des données individuelles à l'effet bloc**

Les dispositifs expérimentaux composés de blocs incomplets randomisés permettent de prendre en compte l'hétérogénéité du milieu, et donc de diminuer la part de la variance environnementale dans la variance phénotypique. Pour chaque caractère mesuré, l'effet bloc a été testé par le modèle d'ANOVA suivant :

$$Y_{ij} = \mu + B_j + E_{ij}$$

avec  $Y_{ij}$  la valeur de la variable pour l'individu  $i$  localisé dans le bloc  $j$ ,  $\mu$  sa moyenne générale et  $E_{ij}$  l'erreur résiduelle. Un ajustement des données individuelles à l'effet bloc a ensuite été réalisé si son effet était significatif :

$$\hat{Y}_i = \mu + \varepsilon_i$$

avec  $\hat{Y}_i$  la valeur de la variable ajustée pour l'individu  $i$ ,  $\mu$  sa moyenne générale et  $\varepsilon_i$  l'erreur résiduelle.

#### C.1.2.3.2.2 Estimation des composantes de la variance

##### C.1.2.3.2.2.1 Estimation des composantes de la variance à l'intérieur de la descendance

La multiplication végétative de chaque descendant nous a permis d'observer de nombreux individus de même génotype dans diverses conditions de micro-milieu. La valeur phénotypique d'un clone est alors la moyenne des valeurs phénotypiques des copies de ce clone. Cela nous permet à l'intérieur de la famille de pleins frères de décomposer la variance phénotypique selon le modèle suivant :

$$Z = \mu + clone + bouture \quad (I)$$

Où  $Z$  est la variance phénotypique dans la descendance,  $\mu$  la moyenne générale du caractère dans la descendance,  $clone$  la variance entre-clone (équivalent à la variance génétique dans le croisement) et  $bouture$  la variance intra-clone qui correspond à la variance environnementale (ou résiduelle). Si la variation entre les différentes copies d'un même clone (variance intra-clone) est nulle alors la valeur phénotypique d'un individu sera, si on l'exprime en écart à la moyenne de la population, la valeur génétique (ou valeur clonale). Si par contre la variance intra-clone n'est pas nulle, cela signifie qu'il existe des facteurs environnementaux qui agissent sur l'expression du phénotype. Certains de ces facteurs peuvent être contrôlés et pourront être pris en compte (effet bloc par exemple), alors que d'autres résultant de micro-phénomènes locaux s'appliquant de manière différente à chaque individu ne pourront être pris en compte.

Afin d'estimer, à l'échelle de la descendance, la part de variance phénotypique expliquée par la variance entre clones nous avons calculé la répétabilité ( $R$ ) qui est égale au rapport entre la variance clonale et la variance des moyennes de clone :

$$R = \frac{\sigma_{clone}^2}{\sigma_{clone}^2 + \frac{\sigma_{bouture}^2}{n}}$$

où  $n$  est le nombre de copie végétative par clone. Si  $n$  égale 1 la répétabilité est égale à l'héritabilité au sens large.

#### C.1.2.3.2.2 Estimation des composantes de la variance à l'intérieur de la population d'où sont issus les parents

Dans un deuxième temps, nous avons estimé les différentes composantes de la variance à l'échelle non pas de la descendance mais de la population d'où les parents ont été échantillonnés.

Si l'on considère plusieurs familles de plein frères, la variance génétique totale est la somme des variances intra famille ( $\sigma_w^2$ ) et inter famille ( $\sigma_b^2$ ) :

$$V_G = \sigma_w^2 + \sigma_b^2 \quad (a)$$

Il est ensuite possible d'estimer chacune de ces composantes :

- Estimation de la variance génétique totale  $V_G$

Si les effets d'épistasie sont négligés (interaction entre allèles de locus différents), la variance génétique totale  $V_G$  est égale à la somme de la variance génétique additive  $V_A$  (somme des effets moyens des allèles d'un même locus) et de la variance de dominance  $V_D$  (interaction entre les allèles d'un même locus) :

$$V_G = V_A + V_D \quad (b)$$

- Estimation de la variance inter famille  $\sigma_b^2$

Le modèle de la génétique quantitative prédit que la variance entre famille de pleins frères est égale à la covariance entre pleins frères

$$\sigma_b^2 = Cov(\text{plein frères}) = \frac{1}{2}V_A + \frac{1}{4}V_D \quad (c)$$

- Estimation de la variance intra famille  $\sigma_w^2$

Etant donné que les individus multipliés végétativement sont des pleins frères, la variance clonale  $\sigma_{clone}^2$  est égale à la variance intra famille

$$\sigma_w^2 : \sigma_{clone}^2 = \sigma_w^2 \quad (d)$$

- Estimation de la variance intra famille  $\sigma_{clone}^2$

Il est possible de déduire des relations (a), (b), (c) et (d) que la variance clonale est égale à :

$$\sigma_{clone}^2 = \frac{1}{2}V_A + \frac{3}{4}V_D$$

Connaissant la variance clonale, il est ensuite possible d'estimer la variance génétique additive en faisant des suppositions sur la variance de dominance. Si on considère que la variance de dominance est généralement inférieure à la variance d'additivité chez les arbres forestiers (Cornelius, 1994), on peut estimer la variance d'additivité en prenant des valeurs extrêmes de variance de dominance. Nous avons choisi d'estimer la variance d'additivité en considérant deux situations contrastées ( $V_D=0$ ) et ( $V_D=1/2 V_A$ ).

Ainsi, si  $V_D=0$ , alors

$$V_A = 2 \sigma_{clone}^2$$

et si  $V_D=1/2 V_A$ , alors

$$V_A = \frac{8}{7} \sigma_{clone}^2$$

Connaissant  $V_A$  il est aussi possible d'estimer, dans la population d'où sont issus les parents, l'héritabilité au sens strict qui est la proportion de variance génétique additive dans la variance phénotypique :  $h^2 = \frac{V_A}{V_P}$ . Cette relation peut encore s'écrire :

$$h^2 = \frac{V_A}{V_D + V_A + V_E}$$

avec  $V_E$  égale à la variance intra clone ( $\sigma_{bouture}^2$ ).

Dans le cas où l'on considère la dominance nulle ( $V_D=0$ ), l'héritabilité au sens strict est égale à :

$$h^2 = \frac{2\sigma_{clone}^2}{2\sigma_{clone}^2 + \sigma_{bouture}^2}$$

Dans le cas où l'on considère  $V_D=1/2 V_A$ , l'héritabilité au sens strict est égale à :

$$h^2 = \frac{\frac{8}{7}\sigma_{clone}^2}{\frac{12}{7}\sigma_{clone}^2 + \sigma_{bouture}^2}$$

### C.1.2.3.3 Détection de QTL

#### C.1.2.3.3.1 Le modèle statistique

La recherche de QTL a été réalisée en utilisant le logiciel MultiQTL (Britvin *et al*, 2001), (<http://esti.haifa.ac.il/~poptheor>). Le modèle choisi dans notre étude est un modèle à un QTL par chromosome, sans utiliser les options proposées de restauration des données manquantes, et de normalisation de la valeur des LOD score. Afin de réduire le bruit de fond et l'influence des QTL situés sur les autres chromosomes, le modèle statistique utilisé est le composite interval mapping qui cumule la méthode de cartographie par intervalle (interval mapping) et une méthode d'ajustement des co-facteurs par régression linéaire. La procédure d'ajustement des co-facteurs consiste à analyser sur

### Encadré 3 : Seuil de détection des QTL

Etant donné que dans les deux cartes parentales du chêne pédonculé le nombre de marqueurs diffère suivant le chromosome considéré, nous avons calculé la correspondance entre l'erreur de type I au niveau du génome et l'erreur de type I au niveau du chromosome pour les 24 chromosomes de la façon suivante :

- ❖ Calcul de l'erreur de type I au niveau du marqueur :

$$\alpha_{mj} = 1 - (1 - \alpha_{gj})^{(1/Mj)} \quad j = 1 \text{ ou } 2$$

avec  $\alpha_{mj}$  l'erreur au niveau du marqueur dans la carte parentale  $j$ ,  $\alpha_{gj}$  l'erreur au niveau du génome dans la carte  $j$ , et  $Mj$  le nombre total de marqueurs répartis sur la carte  $j$ .

- ❖ Calcul de l'erreur de type I au niveau de chaque chromosome  $\alpha_{cij}$  :

$$\alpha_{cij} = 1 - (1 - \alpha_{mj})^{Mij} \quad i = 1, 2 \dots 12$$

Avec  $\alpha_{cij}$  l'erreur au niveau du chromosome  $i$  de la carte  $j$  et  $Mij$  le nombre de marqueurs sur le chromosome  $i$  de la carte  $j$ .

La relation entre l'erreur de type I au niveau de chaque chromosome et l'erreur de type I au niveau du génome peut alors s'écrire :

$$\alpha_{cij} = 1 - \left\{ 1 - \left[ 1 - \left( 1 - \alpha_{gj} \right)^{\frac{1}{Mj}} \right]^{Mij} \right\}$$

**Tableau 6** : Correspondance dans les cartes parentales entre l'erreur de type I au niveau du génome, du marqueur et du chromosome

Carte femelle													
$\alpha_g$	$\alpha_m$	$\alpha_{c1}$ (6)	$\alpha_{c2}$ (8)	$\alpha_{c3}$ (3)	$\alpha_{c4}$ (9)	$\alpha_{c5}$ (7)	$\alpha_{c6}$ (6)	$\alpha_{c7}$ (6)	$\alpha_{c8}$ (5)	$\alpha_{c9}$ (6)	$\alpha_{c10}$ (8)	$\alpha_{c11}$ (6)	$\alpha_{c12}$ (4)
0.05	0.0007	0.0041	0.0055	0.0021	0.0062	0.0048	0.0041	0.0041	0.0034	0.0041	0.0055	0.0041	0.0028
0.10	0.0014	0.0085	0.0113	0.0043	0.0127	0.0099	0.0085	0.0085	0.0071	0.0085	0.0113	0.0085	0.0057
0.15	0.0030	0.0179	0.0238	0.0090	0.0268	0.0209	0.0179	0.0179	0.0150	0.0179	0.0238	0.0179	0.0120
0.20	0.0048	0.0285	0.0378	0.0144	0.0425	0.0332	0.0285	0.0285	0.0238	0.0285	0.0378	0.0285	0.0191
0.30	0.0069	0.0406	0.0537	0.0205	0.0602	0.0472	0.0406	0.0406	0.0339	0.0406	0.0537	0.0406	0.0272
0.50	0.0093	0.0546	0.0722	0.0277	0.0808	0.0634	0.0546	0.0546	0.0457	0.0546	0.0722	0.0546	0.0378
Carte mâle													
$\alpha_g$	$\alpha_m$	$\alpha_{c1}$ (4)	$\alpha_{c2}$ (8)	$\alpha_{c3}$ (5)	$\alpha_{c4}$ (6)	$\alpha_{c5}$ (7)	$\alpha_{c6}$ (5)	$\alpha_{c7}$ (4)	$\alpha_{c8}$ (7)	$\alpha_{c9}$ (6)	$\alpha_{c10}$ (6)	$\alpha_{c11}$ (9)	$\alpha_{c12}$ (5)
0.05	0.0007	0.0028	0.0057	0.0035	0.0043	0.0050	0.0035	0.0028	0.0050	0.0043	0.0043	0.0064	0.0035
0.10	0.0015	0.0058	0.0116	0.0073	0.0087	0.0102	0.0073	0.0058	0.0102	0.0087	0.0087	0.0131	0.0073
0.15	0.0023	0.0090	0.0179	0.0112	0.0134	0.0156	0.0112	0.0090	0.0156	0.0134	0.0134	0.0201	0.0112
0.20	0.0031	0.0123	0.0245	0.0154	0.0184	0.0214	0.0154	0.0123	0.0214	0.0184	0.0184	0.0275	0.0154
0.30	0.0497	0.0197	0.0391	0.0246	0.0294	0.0343	0.0246	0.0197	0.0343	0.0294	0.0294	0.0438	0.0246
0.50	0.0096	0.0378	0.0741	0.0470	0.0561	0.0652	0.0470	0.0378	0.0652	0.0561	0.0561	0.0830	0.0470

Entre parenthèses est indiqué le nombre de marqueur dans le chromosome considéré

chaque chromosome la présence du QTL par interval mapping tout en soustrayant l'effet des QTL situés sur les autres chromosomes. La position et l'effet des QTL n'étant pas connus à l'avance l'algorithme est itératif. Concrètement, dans le premier cycle d'analyse, le QTL le plus significatif est recherché parmi les 12 chromosomes par interval mapping, et son effet est annulé. Dans le deuxième cycle, le deuxième QTL le plus significatif est recherché parmi les 11 chromosomes restants et son effet est annulé à nouveau. La procédure est répétée jusqu'à ce qu'il n'y ait plus de QTL détectés sur les chromosomes restants. En faisant l'hypothèse d'absence d'interaction, l'effet de chaque QTL détecté est ensuite ré-évalué en l'ajustant à l'effet des QTL provenant des différents chromosomes dans l'ordre décroissant de leur significativité. Cette opération est répétée jusqu'à ce que l'effet des QTL diffère de moins de 2% entre deux itérations successives. Une dernière étape consiste à annuler les effets ré-évalués de l'ensemble des QTL détectés et de recommencer la procédure sur les chromosomes restants jusqu'à ce qu'il n'y ait plus d'effet détecté.

#### *C.1.2.3.3.2 Seuil de détection des QTL*

L'erreur de type I de la détection de QTL qui correspond à la probabilité d'obtenir un QTL par le simple fait du hasard est testée par le test de permutation proposé par Churchill et Doerge (1994).

L'association initiale entre les marqueurs et les phénotypes est cassée en permutant 1000 fois les valeurs des caractères tout en respectant l'ordre des marqueurs. Les données permutées sont analysées par la même procédure que les données initiales et la fréquence d'obtention d'une valeur de LOD supérieure à la valeur de LOD initiale est considérée comme une mesure empirique de l'erreur de type I. Les tests de permutations réalisés par MultiQTL donnent une erreur de type I des QTL au niveau du chromosome. Dans un premier temps, afin d'être le plus exhaustif possible, nous avons fixé le seuil de détection des QTL à 5 % au niveau du chromosome ce qui correspond à une probabilité d'environ 50 % au niveau du génome (Encadré 3)!

Dans un deuxième temps, afin de considérer la signification des QTL au niveau du génome et non pas au niveau du chromosome nous avons calculé l'erreur de type I au niveau de chaque chromosome pour différents seuils d'erreur de type I au niveau du génome (5, 10, 15, 20, 30, et 50 %) (Encadré 3).

#### C.1.2.3.4 Intervalle de confiance des QTL

L'intervalle de confiance de la position des QTL est obtenu par Bootstrap (Visscher *et al.*, 1996) qui est une méthode de ré-échantillonnage aléatoire avec remise. 1000 échantillons aléatoires de 275 individus (taille de la descendance) sont générés en tirant au hasard 275 fois avec remise un descendant dans la population de cartographie sans casser la liaison génotype-phénotype. A l'extrême, il est possible d'obtenir un échantillon de 275 individus identiques. Chaque échantillon est ensuite analysé par composite interval mapping afin de construire un histogramme de distribution des fréquences de valeurs de LOD maximum dans chaque intervalle du chromosome considéré. L'intervalle de confiance à 95 % de la position des QTL peut alors être calculé :

$$IC_{(95\%)} = \text{moyenne} \pm 1.96 \text{ écart-type}$$

#### C.1.2.3.5 Estimation des effets

Dans la stratégie "double pseudo testcross" une carte est réalisée pour chaque parent. Le modèle utilisé pour estimer la variance de moyenne de clone  $\sigma_{\bar{z}}^2$  [ $\sigma_{\bar{z}}^2 = \sigma_{clone}^2 + \frac{\sigma_{bouture}^2}{n}$ ] expliquée par un QTL dans chacune des cartes est un modèle à un facteur fixé, le locus marqueur, dont les niveaux sont les deux génotypes possibles Qq et qq

$$\bar{Z}_{ij} = \mu + \alpha_i + E_{ij}$$

Avec  $\bar{Z}_{ij}$  la moyenne des copies végétatives du clone  $j$  de la classe génotypique  $i$  au locus marqueur ( $i = Qq$  ou  $qq$ ),  $\alpha_i$  est l'effet du locus marqueur de génotype  $i$  et  $E_{ij}$  est la variable résiduelle.

Le logiciel MultiQTL calcule la proportion  $P$  de variance de moyenne de clone expliquée par le QTL comme une héritabilité au sens large basée sur des moyenne de clones.

$$P = \frac{p(1-p)d^2}{\sigma_{\bar{z}}^2} = \frac{1/4 d^2}{\sigma_{\bar{z}}^2}$$

où l'effet du QTL  $d$  peut être estimé de la façon suivante :

$$d = X_{Qq} - X_{qq}$$

avec  $X_{Qq}$  et  $X_{qq}$  les valeurs moyennes du caractère dans les deux classes génotypiques alternatives.

$\sigma_{\bar{z}}^2$  est la variance des moyennes de clones et  $p$  la fréquence de l'allèle Q dans la population (à l'équilibre de Hardy-Weinberg la fréquence attendue est 0.5). Connaissant  $P$  et la répétabilité  $R$  il est alors possible de calculer la proportion  $Q$  de variance clonale expliquée par le QTL :

$$Q = \frac{1/4 d^2}{\sigma_{clone}^2} = \frac{P}{R}$$

Il est aussi possible d'estimer la proportion  $S$  de variance génétique additive expliquée par le QTL dans la population d'où sont issus les parents du croisement :

$$S = \frac{1/4 d^2}{\sigma_A^2}$$

quand  $V_D=0$ , alors  $S= Q/2$  et quand  $V_D=1/2 V_A$ , alors  $S= 7Q/8$ .

Finalement nous pouvons calculer la proportion de variance phénotypique expliquée par le QTL dans la population d'où sont issus les parents du croisement :

$$PV = \frac{1/4 d^2}{\sigma_P^2}$$

avec  $PV= S* h^2$ .

### C.1.3 DISTRIBUTION DES QTL LE LONG DES GROUPES DE LIAISON

L'un des objectifs de notre étude est de comparer la distribution des QTL dans le génome pour les deux types de caractères : ceux manifestant une différenciation phénotypique entre les deux espèces et ceux n'en manifestant pas.

#### C.1.3.1 Hypothèse de distribution totalement aléatoire (Processus poissonien)

L'analyse de la distribution observée sera faite par rapport à la situation d'une distribution totalement aléatoire (CSR Complete Spatial Randomness). Comment construire une distribution totalement aléatoire ? Tout d'abord nous allons subdiviser de manière arbitraire le génome de longueur L (en cM) en N intervalles de taille t (en cM). Dans ce qui suit, nous considérerons comme intervalle la distance moyenne séparant les deux estimations de la localisation du QTL : celle qui correspond au LOD score maximum et celle qui correspond à la moyenne des 1000 ré-échantillonnages bootstraps. Soit n le nombre de QTL étudiés et  $\lambda$  la densité de QTL par cM, la densité par intervalle considéré est donc  $\lambda t$ .

Sous l'hypothèse de distribution totalement aléatoire, la probabilité qu'un intervalle  $X_i$  comprenne k QTL suit alors une distribution de Poisson de moyenne  $\lambda t$ .

$$\Pr ( X_i = k ) = \frac{e^{-\lambda t} (\lambda t)^k}{k!}$$

#### C.1.3.2 Comparaison de la distribution observée à l'Hypothèse nulle de distribution totalement aléatoire

A l'aide de ces probabilités nous pouvons donc construire la distribution attendue sous l'hypothèse  $H_0$  de distribution aléatoire des QTL. La distribution observée pourra donc simplement être comparée à la distribution  $H_0$  par un test G à l-2 degrés de liberté (où l est le nombre de classes considérées). La perte de 2 ddl (au lieu de 1 dans un test G classique) est liée à la construction de l'hypothèse nulle de la distribution aléatoire des QTL qui est dépendante de la densité de QTL par intervalle ( $\lambda t$ ).

Dans un deuxième temps, nous nous sommes s'interrogés sur l'écart de la distribution observée à la distribution aléatoire. Pour cela on peut construire la statistique suivante, encore appelée indice de dispersion I (rapport entre la variance et la moyenne du nombre de QTL par intervalle). Sous l'hypothèse de distribution poissonnienne, la moyenne devrait être égale à la variance

$$I = \frac{s^2}{\bar{x}}$$

La statistique I suit une loi de chi deux à N-1 degrés de liberté.

Dans le cas où I est significativement supérieur à 1 (les intervalles contenant plusieurs QTL sont excédentaires par rapport à la distribution aléatoire) les QTL observent une distribution agrégée.

Dans le cas où I est significativement inférieure à 1 (les intervalles contenant plusieurs QTL sont moins nombreux par rapport à la distribution aléatoire) les QTL observent une distribution régulière.

### C.1.3.3 Colocalisations entre les QTL détectés

Pour tester si le chevauchement des QTL détectés dans différents dispositifs est dû au hasard ou pas, il est possible de calculer la probabilité de co-occurrence de QTL dans les mêmes régions génomiques observée dans deux dispositifs sous l'hypothèse nulle de distribution aléatoire.

Soit  $G$  le nombre de QTL détectés dans le dispositif qui présente le plus grand nombre de QTL détectés, soit  $P$  le nombre de QTL dans le second dispositif, et soit  $C$  le nombre de QTL partagés par les deux dispositifs. Nous nous intéressons au calcul de la probabilité d'occurrence de  $C$  intervalles communs, sachant  $P$  et  $G$  et sous l'hypothèse de distribution au hasard des QTL. Cette probabilité peut être calculée de la manière suivante. Elle est égale au rapport du nombre de combinaisons de  $C$  parmi  $G$  au sein d'un ensemble de  $N$  intervalles, rapporté au nombre total de combinaisons de  $P$  parmi  $N$  (Feltus *et al.*, 2003) :

$$Pvalue = \frac{C_G^C \times C_{N-G}^{P-C}}{C_N^P}$$

**Figure 11** : Caractères morphologiques foliaires mesurés pour la détection de QTL.

*5 Variables mesurées*

- LL longueur du limbe
- PL longueur de pétiole
- LW largeur maximale du lobe
- SW largeur du sinus
- WP longueur de la base du limbe au niveau de sa largeur maximale

*2 Variables comptées*

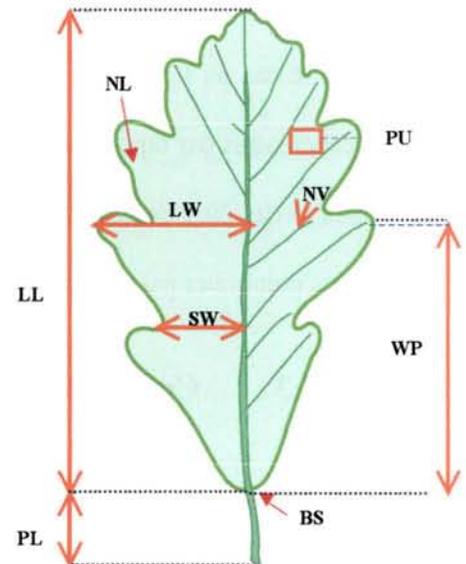
- NL nombre de lobes
- NV nombre de veines intercalaires

*3 Variables observées*

- BS forme de l'oreillette
- PUI pilosité de la face inférieur entre les nervures
- PU<sub>v</sub> pilosité de la face inférieur de la nervure médiane

*5 Variables calculées*

- OB forme du limbe  $100 \cdot WP / LL$
- PR longueur relative du pétiole  $100 \cdot PL / (LL + PL)$
- LDR profondeur relative du lobe  $100 \cdot (LW - SW) / LW$
- PV pourcentage de veination  $100 \cdot NV / NL$
- LWR largeur relative du lobe  $100 \cdot LW / LL$



## C.2 RESULTATS

### C.2.1 DETECTION DE QTL IMPLIQUES DANS LES CARACTERES QUI DIFFERENCIENT LE CHENE SESSILE ET LE CHENE PEDONCULE

#### C.2.1.1 QTL impliqués dans la morphologie foliaire

Ce chapitre a fait l'objet d'une publication acceptée moyennant des modifications mineures dans le journal *Heredity* (voir annexe 1)

##### C.2.1.1.1 Mesures des caractères foliaires

Quinze caractères morphologiques foliaires qui différencient le chêne sessile et le chêne pédonculé ont été mesurés en population naturelle et dans la population de cartographie du chêne pédonculé. Ces différents caractères sont présentés sur la figure 11.

##### C.2.1.1.1.1 *Population naturelle*

Les caractères morphologiques foliaires ont été mesurés au cours d'études antérieures dans un peuplement mixte composé de *Q. petraea* et *Q. robur* localisé à la Petite Charnie dans le Mayenne (Bacilieri *et al.*, 1995; Kremer *et al.*, 2002a). La surface d'étude s'étend sur 4.6 ha (250 m×250 m) et comprend un total de 426 arbres adultes âgés de plus de 100 ans (199 *Q. petraea*, 215 *Q. robur* et 12 de formes intermédiaires). Sur chaque arbre, 5 feuilles ont été prélevées. Les données morphologiques analysées antérieurement (Bacilieri *et al.*, 1995 ; Kremer *et al.*, 2002a) ne sont pas présentées dans la thèse, seules seront reportées les composantes de la variance de chacun des caractères. Le modèle de variance utilisé pour estimer les composantes de la variances est indiqué dans l'article en annexe 1.

**Tableau 5** : Décomposition de la variation de la morphologie foliaire en population naturelle.

Caractère	<i>Q. petraea</i>		<i>Q. robur</i>		
	<i>moyenne</i>	<i>Variabilité</i> Intraspécifique <sup>1</sup>	<i>Moyenne</i>	<i>Variabilité</i> Intraspécifique <sup>1</sup>	<i>Variabilité</i> Interspécifique <sup>2</sup>
LL (mm)	91.4	0.59	77.5	0.64	0.23
PL (mm)	16	0.70	4.4	0.52	0.81
LW (mm)	28.2	0.61	24.5	0.63	0.14
SW (mm)	20.5	0.62	13.4	0.50	0.59
WP (mm)	49.7	0.42	44.4	0.50	0.08
NL	11.2	0.52	8.7	0.50	0.51
NV	0.35	0.32	3.1	0.49	0.67
BS	3.9	0.47	6.8	0.48	0.72
PUI	3.4	0.58	1.2	0.74	0.77
OB %	54.1	0.16	57.3	0.17	0.07
PR %	14.8	0.68	5.4	0.50	0.84
LDR %	26.7	0.47	44.5	0.44	0.67
PV %	3.32	0.34	36.0	0.48	0.70
LWR %	30.8	0.51	31.6	0.40	0.02

<sup>1</sup> La variabilité intraspécifique représente la proportion de variation à l'intérieur d'une espèce qui est due à la variation entre les arbres de cette espèce (dérive du modèle 2 décrit dans l'annexe 1).

<sup>2</sup> La variabilité interspécifique est une estimation de la variation expliquée par la différenciation entre les deux espèces (dérive du modèle 1 décrit dans l'annexe 1)

#### C.2.1.1.1.2 Population de cartographie

Les quinze caractères morphologiques foliaires ont été mesurés dans deux dispositifs de la population de cartographie du chêne pédonculé. La première campagne de mesure (expérimentation 1) a été effectuée en 1999 sur des feuilles prélevées sur des boutures réalisées en 1997 âgées de 2 ans et installées depuis 1 an dans le dispositif 2 (Tableau 4). 2 feuilles ont été prélevées sur chacune des 1080 boutures qui composent le dispositif soit 174 clones mesurés en moyenne 6 fois. La deuxième campagne de mesure (expérimentation 2) a été réalisée sur des feuilles prélevées sur des boutures réalisées en 1998 âgées de 2 ans encore installées en pépinière (avant la mise en place dans le dispositif 3 décrit dans le tableau 4). 2 feuilles ont été prélevées sur 1523 boutures provenant de 210 descendants (7 copies végétatives en moyenne par clone).

#### C.2.1.1.2 Résultats

##### C.2.1.1.2.1 Décomposition de la variance phénotypique en population naturelle

Le tableau 5 donne la variabilité intra et interspécifique des différents caractères morphologiques. Les caractères relatifs à la dimension de la feuille (LL, LW, WP) ainsi que leurs variables dérivées (OB, LWR) présentent une faible variabilité interspécifique (entre 2 et 23%). Les caractères les plus discriminants (variance interspécifique supérieure à 50%) sont dans l'ordre décroissant : la longueur du pétiole (PL et PR) , la pubescence sur le limbe (PUI), la forme de l'oreillette (BS), le nombre de veines intercalaires (NV et PV), la largeur du sinus (SW), et le nombre de lobes (NL). Cependant, en dépit de fortes variations interspécifiques la gamme de variation à l'intérieur d'une espèce chevauche la gamme de variation de l'autre espèce. Par exemple chez *Q. petraea* la longueur du pétiole qui est le caractère le plus discriminant, varie entre 3 et 33 mm, alors que chez *Q. robur* PL varie entre 0 et 12 mm.

**Tableau 6 :** Valeur de F correspondant à l'effet bloc et à l'effet clone.

	LL	PL	LW	SW	WP	NL	NV	BS	PUv	PUI	OB	PR	LDR	PV	LWR
Effet bloc (1)	6.3	3.4	3.6	2.1	3	2	4.2	4.7	3.6	14.1	1.4 <sup>ns</sup>	1.1 <sup>ns</sup>	1.5 <sup>ns</sup>	1.7 <sup>ns</sup>	1.7 <sup>ns</sup>
Effet clone (1)*	6	4.7	4.8	4.6	3	5	6.1	7.2	13.4	20.2	1.1 <sup>ns</sup>	2.5	2.8	1.8	1.1 <sup>ns</sup>
Effet clone (2)	2.9	3.9	2.9	3.8	2.8	4.1	6.3	3.1	10.1	6.3	1.8	2.7	3.4	3.7	1.7

Le chiffre entre parenthèses correspond au numéro de l'expérimentation

*ns* indique que le test n'est pas significatif

\* effet clone ajusté à l'effet bloc

**Tableau 7 :** Décomposition de la variation de la morphologie foliaire dans la population de cartographie.

Caractère	Expérimentation 1			Expérimentation 2		
	Moyenne	CV <sub>G</sub>	CV <sub>R</sub>	Moyenne	CV <sub>G</sub>	CV <sub>R</sub>
LL(mm)	58.5	0.10	0.26	69.8	0.11	0.27
PL(mm)	2.6	0.04	0.18	3.2	0.15	0.29
LW(mm)	17.8	0.10	0.29	20.7	0.13	0.32
SW(mm)	6.4	0.16	0.37	8.4	0.20	0.41
WP(mm)	34.6	0.11	0.34	41.9	0.13	0.33
NL	7.7	0.09	0.22	8.5	0.10	0.20
NV	5.8	0.13	0.40	4.2	0.26	0.37
BS	5.1	0.09	0.27	4.6	0.14	0.33
PUv	1.5	0.28	0.43	1.7	0.36	0.40
PUI	2.8	0.15	0.26	2.2	0.24	0.35
OB%	59.6	0.00	0.33	61.1	0.12	0.45
PR%	4.4	0.10	0.38	4.5	0.13	0.33
LDR%	62.9	0.07	0.21	58.9	0.09	0.20
PV%	76.7	0.07	0.51	52.9	0.31	0.63
LWR%	31.0	0.03	0.34	30.0	0.09	0.37

CV<sub>G</sub> coefficient de variation génétique  $\frac{\sigma_{clone}}{\mu}$  calculé à partir du modèle I

CV<sub>R</sub> coefficient de variation résiduelle  $\frac{\sigma_{bouture}}{\mu}$  calculé à partir du modèle I

### C.2.1.1.2.2 Décomposition de la variance phénotypique dans la population de cartographie

Dans une analyse préliminaire, l'effet bloc a été testé sur les données de l'expérimentation 1 (Tableau 6). Afin de diminuer la variance environnementale, un ajustement des valeurs individuelles à l'effet bloc a été effectué pour 10 des 15 variables. L'effet clone a ensuite été testé dans les deux dispositifs afin de voir s'il existe des différences significatives entre les moyennes de clones. L'effet clone était significatif pour toutes les variables indiquant ainsi un contrôle génétique des caractères, excepté pour OB et LWR mesurés dans l'expérimentation 1 (Tableau 6).

La normalité de la distribution testée avec la méthode de Box et Cox (1964) indique que seul PUV s'écarte de la normalité. Une transformation  $1/x^2$  est recommandée dans l'expérimentation 1 et  $1/x$  dans l'expérimentation 2.

Afin de comparer la variation des différents caractères entre eux, nous avons calculé les coefficients de variation génétique ( $CV_G = \frac{\sigma_{clone}}{\mu}$ ) et résiduelle ( $CV_R = \frac{\sigma_{bouture}}{\mu}$ ) qui sont indépendants de l'unité de mesure (Tableau 7). Le coefficient de variation génétique pour l'ensemble des caractères est plus faible dans l'expérimentation 1 (en moyenne  $CV_G = 0.10$ ) que dans l'expérimentation 2 (en moyenne  $CV_G = 0.17$ ). Les coefficients de variation résiduelle sont par contre similaires (0.32 et 0.35).

Une deuxième façon de comparer les composantes de la variance des différents caractères a été (1) de calculer la répétabilité R qui donne la proportion de variance des moyennes de clones dans la variance phénotypique à l'intérieur de la population de cartographie et (2) d'estimer l'héritabilité au sens strict qui donne la proportion de variance génétique additive dans la variance phénotypique à l'intérieur de la population d'où ont été échantillonnés les parents (Tableau 8).

Les héritabilités et les répétabilités sont supérieures dans l'expérimentation 2, ce qui s'explique au vu des coefficients de variation, par une variabilité génétique plus forte dans l'expérimentation 2 plutôt qu'une variabilité environnementale plus faible (Tableau 7). La pilosité est le caractère le plus héritable (aussi bien PUI que PUV) et OB ou LWR les caractères les moins hértables. La plupart des caractères qui sont fortement impliqués dans la différenciation interspécifique (PL, SW, NV, BS, PU,

**Tableau 8 :** Répétabilité et héritabilité des caractères de morphologie foliaire.

Caractère	Expérimentation 1			Expérimentation 2		
	Héritabilité		R	Héritabilité		R
	$V_D=0$	$V_D=1/2 V_A$		$V_D=0$	$V_D=1/2 V_A$	
LL	0.22	0.13	0.56	0.25	0.15	0.66
PL	0.08	0.05	0.30	0.34	0.20	0.74
LW	0.21	0.12	0.54	0.25	0.15	0.65
SW	0.26	0.15	0.62	0.33	0.20	0.73
WP	0.17	0.10	0.49	0.24	0.14	0.65
NL	0.23	0.14	0.58	0.35	0.21	0.75
NV	0.17	0.1	0.49	0.48	0.30	0.84
BS	0.2	0.11	0.53	0.27	0.16	0.68
PUv	0.46	0.28	0.8	0.61	0.39	0.90
PUI	0.38	0.23	0.74	0.48	0.30	0.84
OB%	0	0	0	0.12	0.07	0.44
PR%	0.12	0.07	0.38	0.23	0.14	0.63
LDR%	0.18	0.11	0.51	0.29	0.18	0.70
PV%	0.04	0.023	0.17	0.32	0.19	0.73
LWR%	0.02	0.009	0.07	0.12	0.07	0.43

R : répétabilité de la moyenne de clones

**Tableau 9:** Coefficient de corrélation de Pearson entre les moyennes de clones.

	LL	<u>PL</u>	LW	<u>SW</u>	WP	<u>NL</u>	<u>NV</u>	<u>BS</u>	<u>PUv</u>	<u>PUI</u>
LL	0.05	<b>0.47</b>	<b>0.82</b>	<b>0.42</b>	<b>0.83</b>	<b>0.26</b>	<b>0.22</b>	<b>0.17</b>	-0.06	<b>0.24</b>
<u>PL</u>	<b>0.50</b>	<b>0.17</b>	<b>0.47</b>	<b>0.27</b>	<b>0.35</b>	0.10	<b>0.15</b>	<b>-0.24</b>	-0.08	0.01
LW	<b>0.85</b>	<b>0.49</b>	0.09	<b>0.56</b>	<b>0.65</b>	0.10	0.09	<b>0.23</b>	-0.10	<b>0.16</b>
<u>SW</u>	<b>0.69</b>	<b>0.40</b>	<b>0.76</b>	<b>0.24</b>	<b>0.48</b>	0.06	<b>-0.28</b>	-0.07	-0.06	0.06
WP	<b>0.88</b>	<b>0.44</b>	<b>0.73</b>	<b>0.68</b>	0.04	<b>0.27</b>	0.10	0.09	-0.05	<b>0.27</b>
<u>NL</u>	<b>0.25</b>	<b>-0.16</b>	<b>0.13</b>	0.07	<b>0.38</b>	0.14	<b>0.45</b>	<b>0.30</b>	<b>0.27</b>	<b>0.27</b>
<u>NV</u>	<b>0.23</b>	<b>0.17</b>	0.09	0.00	<b>0.19</b>	-0.02	0.09	<b>0.26</b>	0.14	0.08
<u>BS</u>	<b>0.25</b>	<b>-0.22</b>	<b>0.28</b>	<b>0.19</b>	<b>0.29</b>	<b>0.31</b>	0.13	0.11	0.05	0.03
<u>PUv</u>	<b>-0.25</b>	<b>-0.26</b>	<b>-0.26</b>	<b>-0.22</b>	<b>-0.16</b>	<b>0.15</b>	-0.11	0.04	<b>0.55</b>	<b>0.22</b>
<u>PUI</u>	-0.06	-0.08	-0.07	<b>-0.14</b>	0.00	<b>0.20</b>	<b>-0.17</b>	-0.10	<b>0.28</b>	<b>0.20</b>

Les valeurs en gras correspondent à des corrélations significatives

Les valeurs au dessus de la diagonale correspondent à des corrélations dans l'expérimentation 1

Les valeurs au dessous de la diagonale correspondent à des corrélations dans l'expérimentation 2

Les valeurs sur la diagonale correspondent à des corrélations entre les 2 expérimentations

Les caractères soulignés présentent la plus grande variabilité interspécifique (voir tableau 5)

NL) et leur variables dérivées (PR, LDR, PV) sont aussi ceux qui présentent les plus fortes valeurs d'héritabilité.

#### *C.2.1.1.2.3 Corrélations entre caractères*

Les corrélations entre les moyennes de clones sont données dans le tableau 9. Les variables dérivées n'ont pas été incluses. Dans chaque expérimentation les caractères de dimension de la feuille sont fortement corrélés entre eux (LL, SW, LW, et WP) ce qui laisse suggérer la présence de pléiotropie (un même gène qui contrôle plusieurs caractères). Si l'on se focalise sur les caractères qui différencient le plus les deux espèces, parmi les 21 paires possibles de caractères, 8 sont significatives dans l'expérimentation 1 et 13 dans l'expérimentation 2. La corrélation la plus forte est obtenue entre SW et PL (0.27 et 0.40 respectivement dans l'expérimentation 1 et 2). Concernant les corrélations entre les deux dispositifs, seules 4 corrélations sur 10 sont significatives. La plus forte corrélation étant obtenue pour PUv (0.55). Les faibles corrélations entre les deux années de mesure peuvent être liées à des facteurs environnementaux (dispositif en pépinière et dispositif sur le terrain) ou à des facteurs génétiques.

#### *C.2.1.1.2.4 Détection de QTL*

Dans un premier temps, je propose de focaliser les résultats sur les QTL significatifs au seuil de 5 % au niveau du génome (Tableau 10).

Aucun QTL n'a été détecté pour deux caractères (LW et OB) parmi les 15 étudiés. Pour OB cela s'explique par une faible variance génétique (Tableau 10).

Pour les 13 caractères restants entre 1 et 2 QTL ont été détectés. Ces QTL expliquent en moyenne entre 10.7 et 13.8 % de la variance des moyennes de clones ( $P$ ) à l'intérieur de la population de cartographie, entre 7.8 et 40% de la variance additive ( $S$ ) et entre 2.35 et 2.8 % de

**Tableau 10** : Liste des QTL détectés au seuil de 5% au niveau du génome pour les différents caractères morphologiques foliaires étudiés.

Expérimentation 1

Caractère	GL	N	Lod	$\alpha_{ch}$	P1	P2	IC	D	P	V <sub>D</sub> =0		V <sub>D</sub> =1/2V <sub>A</sub>	
										S	PV	S	PV
PL	1f	124	3.9	0.004	55.4	46.7	20.2	0.32	13.6	23	1	40	1.9
PR	10f	134	3.2	0.005	63.1	60.9	18.5	0.48	9.7	13	1.5	22	1.5
LWR	10f	131	2.8	0.005	68.9	65.5	19	2.26	9.9	69	1.1	120	1.1
SW	3f	126	3.7	0.002	22.3	22.1	6.9	1.00	15.5	13	3.3	22	3.4
SW	3m	137	3.6	0.003	19.3	17.2	9.2	0.81	10.9	9	2.3	15	2.4
LDR	3f	126	3.7	0.002	20.5	18.2	16.3	-4.99	14.5	14	2.6	25	2.7
LDR	3m	137	3.0	0.003	19.1	18.5	7.9	-3.84	9.9	10	1.8	17	1.8
PUI	12f	110	4.8	0.003	14.9	17.6	11.9	0.37	17.2	12	4.4	20	4.7
PUI	1m	108	3.0	0.003	29.3	30.8	12.9	-0.33	12.9	9	3.3	15	3.5
NL	2f	89	3.9	0.005	67	50.7	39.7	0.94	22.9	20	4.6	34	4.7
NV	3f	126	4.7	0.002	20.2	15.8	15.7	-0.91	16.4	17	2.9	29	2.9
PV	3f	126	5.7	0.002	20.4	18.3	12.8	-12.19	21.7	66	2.7	115	2.7
Valeur moyenne									13.8	23	2.7	40	2.8

Expérimentation 2

Caractère	GL	N	Lod	$\alpha_{ch}$	P1	P2	IC	D	P	V <sub>D</sub> =0		V <sub>D</sub> =1/2V <sub>A</sub>	
										S	PV	S	PV
LL	11f	171	3.2	0.004	2.7	7.3	27.8	6.64	8.8	7	1.7	12	1.7
LL	6m	184	2.5	0.003	26	25.3	10.3	-5.70	6	5	1.1	8	1.2
PL	1f	156	4.2	0.004	39	44.2	17.1	0.41	11.2	8	2.5	13	2.7
PL	11f	171	3.4	0.004	17.7	19.4	23.3	0.36	7.9	5	1.8	9	1.9
PR	1F	165	3.6	0.004	9	12.2	25.5	0.66	11.8	9	2.2	16.4	2.2
SW	3f	162	3.1	0.002	17.9	18.3	5.4	1.51	8.4	6	1.9	10	2
SW	3m	160	2.5	0.003	43.2	27.6	33.1	1.61	9.7	7	2.1	12	2.2
LDR	3f	162	7.1	0.002	17.9	17.7	3.0	-6.14	19.1	14	4	24	4.2
LDR	3m	160	4.5	0.003	41.1	34.2	21	-6.41	8.2	6	1.8	10	1.8
BS	4m	131	2.6	0.004	86.3	74.4	33.7	-0.54	8.4	6	1.7	11	1.8
WP	11f	171	3.0	0.004	5.2	9.6	30.4	4.69	8.4	7	1.6	11	1.6
NL	8m	166	3.1	0.005	11.6	16	34.1	-0.71	8.4	6	2	10	2.1
NV	5f	168	7.4	0.005	31.4	32.2	14.1	1.00	17.6	11	5.1	18	5.4
PV	5f	168	6.2	0.005	31.4	30.7	6.1	15.53	15.7	12	3.4	19	3.6
Valeur moyenne									10.7	7.8	2.35	13.1	2.4

Avec GL le groupe de liaison (f=femelle ; m=mâle), N le nombre d'individus pris en compte dans l'analyse, Lod la valeur du Lod score au niveau de la position la plus probable du QTL,  $\alpha_{ch}$  l'erreur de type I au niveau du chromosome qui correspond à une erreur de type I de 5% au niveau du génome, P1 la position la plus probable du QTL obtenue d'après la valeur maximale de LOD, P2 la position la plus probable du QTL obtenue après 1000 ré-échantillonnages réalisés par bootstrap, IC l'intervalle de confiance du QTL obtenu après 1000 ré-échantillonnages réalisés par bootstrap, D l'effet de substitution allélique (correspond à la différence entre la moyenne des individus hétérozygotes et la moyenne des individus homozygotes), P le pourcentage de variance des moyennes de clones expliquée par le QTL, S le pourcentage de variance additive expliquée par le QTL en supposant la variance de dominance nulle (V<sub>D</sub>=0) ou égale à la moitié de la variance d'additivité (V<sub>D</sub>=1/2V<sub>A</sub>), PV le pourcentage de variance phénotypique expliquée par le QTL dans la population d'où sont issus les parents du croisement en supposant V<sub>D</sub>=0 ou V<sub>D</sub>=1/2V<sub>A</sub>.

**Tableau 11** : Liste des QTL détectés au seuil de 5 % au niveau du chromosome pour les différents caractères morphologiques foliaires étudiés.

<i>Expérimentation 1</i>											<i>Expérimentation 2</i>												
caractère	<i>n</i>	<i>GL</i>	<i>Lod</i>	$\alpha_c$	$\alpha_g$	<i>P1</i>	<i>P2</i>	<i>IC</i>	<i>D</i>	<i>P</i>	caractère	<i>n</i>	<i>GL</i>	<i>Lod</i>	$\alpha_{ch}$	$\alpha_g$	<i>P1</i>	<i>P2</i>	<i>IC</i>	<i>D</i>	<i>P</i>		
LL	2	7F	2.4	0.009	10	31.1	33.5	28.6	4.76	8.7	LL	2	11F	3.21	0.001	5	2.7	7.26	27.8	6.64	8.8		
		11F	1.9	0.025	20	62.5	54.7	35.9	-4.14	6.3			6M	2.47	0	5	26	25.3	10.3	-5.70	6		
LW	1	11F	2.4	0.005	10	63.0	54.6	40.2	-1.54	8.5	LW	1	11F	2.16	0.014	15	2.6	12.8	40.2	2.04	6		
WP	2	5M	1.6	0.045	50	36.9	50.9	51.2	3.37	7.1	WP	2	11F	3.03	0.001	5	5.2	9.63	30.4	4.69	8.4		
		11F	2.2	0.014	15	63.5	61.5	23.1	-3.36	7.9			6M	2.19	0.007	10	26	24.5	13.6	-3.94	5.3		
NL	2	2F	3.9	0	5	67.0	50.7	39.8	0.94	22.9	NL	1	8M	3.08	0.001	5	11.6	16	34.1	-0.71	8.4		
		12M	3.4	0.004	10	38.1	38	14	-0.69	12.8			PL	3	11F	3.44	0	5	17.7	19.4	23.3	0.36	7.9
1F	3.9	0	5	55.4	46.7	20.2	0.32	13.6	1F	4.25	0	5			39	44.2	17.1	0.41	11.2				
4M	2.0	0.017	20	8.7	20.6	47.7	-0.23	6.4	3F	1.80	0.017	15			21.3	21.6	10.3	0.26	4.8				
NV	2	2F	2.0	0.026	20	66.7	54.7	35.9	0.68	10.5	NV	3	12F	2.27	0.011	15	47.6	44.4	21.4	-0.56	5.7		
		3F	4.7	0	5	20.2	15.8	15.7	-0.91	16.4			3F	1.49	0.028	50	14.4	15.5	11.9	-0.55	5.3		
SW	5	3F	3.7	0	5	22.3	22.1	6.9	1.00	15.5	SW	2	5F	7.38	0	5	31.4	32.2	14.1	1.00	17.6		
		3M	3.6	0.002	5	19.3	17.2	9.2	0.82	10.9			3F	3.08	0	5	17.9	18.3	5.4	1.51	8.4		
		5M	2.6	0.007	10	83.2	73.5	42.7	0.70	8			3M	2.50	0.003	5	43.2	27.6	33.1	1.61	9.7		
		6M	1.7	0.034	50	14.7	16.2	22.0	0.54	4.4			PUI	2	6M	2.19	0.012	20	34.2	30.4	22.7	0.31	6.3
		9F	1.7	0.031	30	51.1	50.2	27.6	-0.63	6.1					7M	1.55	0.038	50	40.6	34.6	21.6	0.27	4.4
PUI	4	1F	2.8	0.005	10	34.0	37.2	15.6	-0.24	7.8	PUv	4	4F	1.99	0.027	20	80.5	73.7	58.4	0.35	7.1		
		1M	3.0	0	5	29.3	30.8	12.9	-0.33	12.9			4M	1.59	0.047	50	8.5	13.5	43.1	-0.28	4.1		
		6F	2.5	0.007	10	7.6	31.8	44.9	-0.26	9.2			7F	1.86	0.032	30	73.6	75.9	62.5	-0.32	6.4		
		12F	4.8	0	5	14.9	17.6	11.9	0.37	17.2			9F	2.80	0.005	10	62.8	59.1	25.1	0.41	9.6		
PUv	3	7F	1.8	0.023	20	59.7	50.5	57.4	-0.21	6.2	BS	5	11F	1.39	0.043	50	76.3	66.6	37.4	-0.39	3.7		
		8F	2.1	0.009	15	42.2	40	19.5	0.27	9.8			1F	1.51	0.043	50	7.87	13.6	29.6	-0.44	5.3		
		9F	2.7	0.002	20	59.7	54.6	30.2	0.28	10.3			3M	2.50	0.006	10	0.7	3.4	12.0	0.50	6.9		
												6M	2.19	0.007	10	26	24.5	13.6	-3.94	5.3			
												4M	2.6	0.003	5	86.3	74.4	33.7	-0.54	8.4			

Suite **Tableau 11** :

<i>Expérimentation 1</i>										<i>Expérimentation 2</i>												
caractère	<i>n</i>	<i>GL</i>	<i>Lod</i>	$\alpha_c$	$\alpha_g$	<i>PI</i>	<i>P2</i>	<i>IC</i>	<i>D</i>	caractère	<i>n</i>	<i>GL</i>	<i>Lod</i>	$\alpha_{ch}$	<i>ag</i>	<i>PI</i>	<i>P2</i>	<i>IC</i>	<i>D</i>	<i>P</i>		
LDR <sub>f(LW,SW)</sub>	5	3F	3.7	0	5	20.5	18.2	16.3	-4.99	LDR <sub>f(LW,SW)</sub>	3	2F	2.31	0.01	10	73.3	68.3	26.3	-4.45	8.2		
		3M	3	0.001	5	19.1	18.5	7.9	-3.8			3F	7.11	0	5	17.9	17.7	3.0	-6.14	18.1		
		6F	1.7	0.029	30	53.4	55.1	30	-3.94			3M	4.45	0	5	41.1	34.2	21.0	-6.41	19.1		
		6M	1.6	0.04	50	11.5	15.1	22.5	-3			LWR <sub>f(LW,LL)</sub>	2	11F	1.40	0.017	15	62.5	60.6	18.1	-2.61	3.8
		9F	1.5	0.048	50	33.0	39.3	37.4	3.3					7M	1.10	0.041	50	22.9	25.7	16.8	-2.37	3.1
LWR <sub>f(LW,LL)</sub>	2	3M	2.1	0.012	20	47.1	39.8	30.2	2.09	PR <sub>f(PL,LL)</sub>	1	1F	3.57	0	5	9	12.2	25.5	0.66	11.8		
		10F	2.8	0.001	5	68.9	65.5	12.9	2.26	PV <sub>f(NV,NL)</sub>	2	5F	6.22	0	5	31.4	30.7	6.1	15.53	15.7		
PR <sub>f(PL,LL)</sub>	5	1F	2.3	0.007	10	55.4	46.9	22	0.41			8M	2.47	0.005	10	11.6	15	24.7	10.33	6.8		
		4M	2.7	0.005	10	8.7	10.1	19.0	-0.4													
		8M	2.2	0.016	20	95.4	90.4	38.8	-0.48													
		10F	3.2	0.002	5	63.1	60.9	18.5	0.48													
PV <sub>f(NV,NL)</sub>	2	11F	1.5	0.048	50	50.6	40.9	33.9	0.29													
		3F	5.7	0	5	20.4	18.3	12.8	-12.19													
		3M	2.4	0.005	10	17.8	17	11.9	-8.56													

Avec *GL* le groupe de liaison (F=femelle ; M=mâle), *n* le nombre de QTL détectés pour le caractère considéré, *Lod* la valeur du Lod score au niveau de la position la plus probable du QTL,  $\alpha_c$  l'erreur de type I au niveau du chromosome obtenue après 1000 permutations des données,  $\alpha_g$  l'erreur de type I (en pourcentage) au niveau du génome qui correspond à  $\alpha_c$ , *PI* la position la plus probable du QTL d'après la valeur maximale de LOD, *P2* la position la plus probable du QTL obtenue après 1000 ré-échantillonnages réalisés par bootstrap, *IC* l'intervalle de confiance du QTL obtenu après 1000 ré-échantillonnages réalisés par bootstrap, *D* l'effet de substitution allélique (correspond à la différence entre la moyenne des individus hétérozygotes et la moyenne des individus homozygotes), *P* le pourcentage de la variance des moyennes de clones expliquée par le QTL.

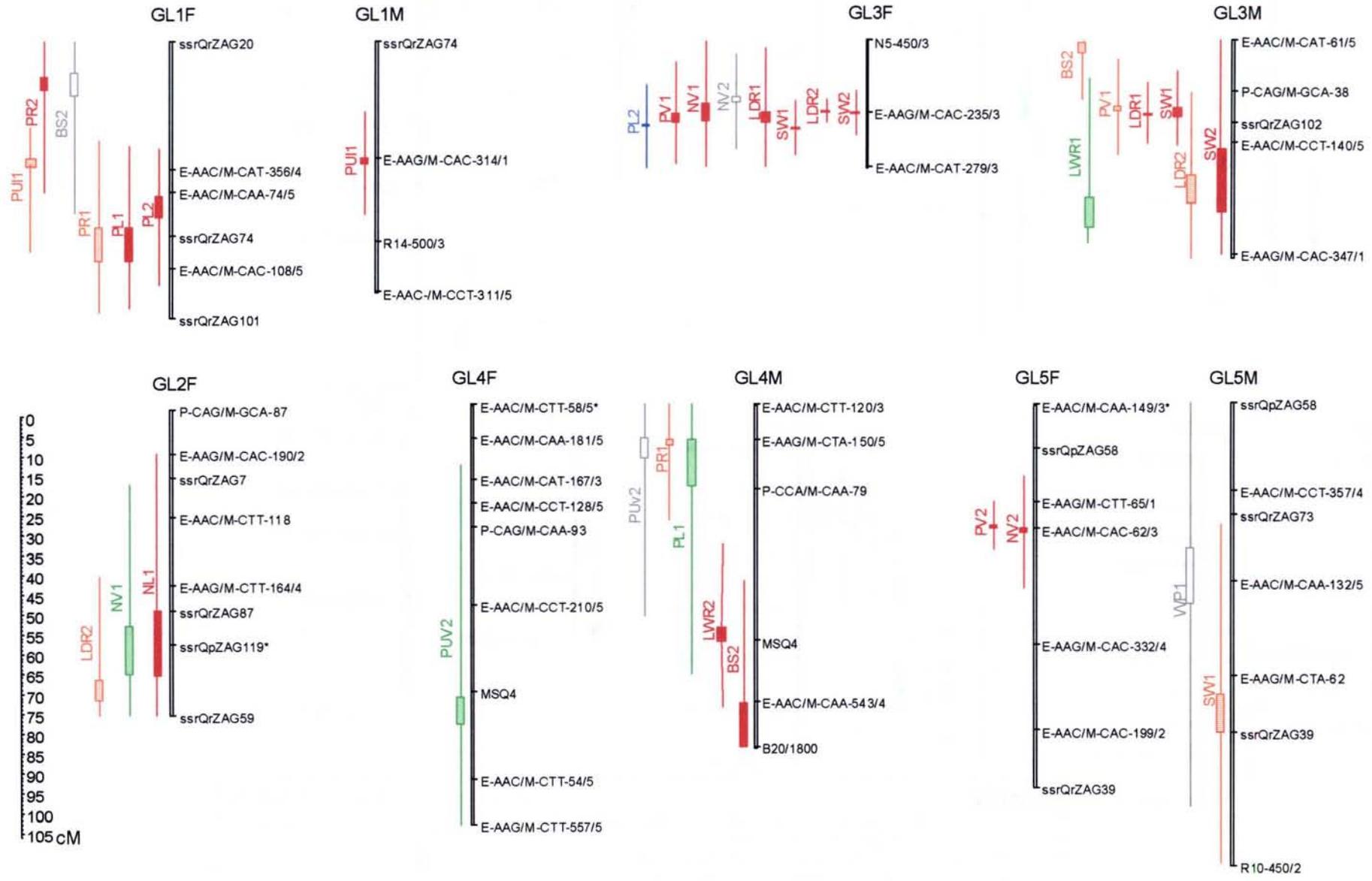
la variance phénotypique ( $PV$ ) à l'intérieur de la population de *Q. robur*. En dépit de fortes répétabilité la part de variance des moyennes de clone ( $P$ ) expliquée par les QTL reste modérée, aucun QTL à effet majeur n'est observé. Pour 7 caractères fortement impliqués dans la différenciation entre les deux espèces (NL, PL, SW, NV, LDR, PR et PV) des QTL ont été détectés dans les deux dispositifs sur 8 groupes de liaisons (1, 2, 3, 4, 5, 8, 11 et 12) (Figure 12). Pour PL (GL1) et SW (GL3) les QTL sont détectés sur le même groupe de liaison dans les deux expérimentations.

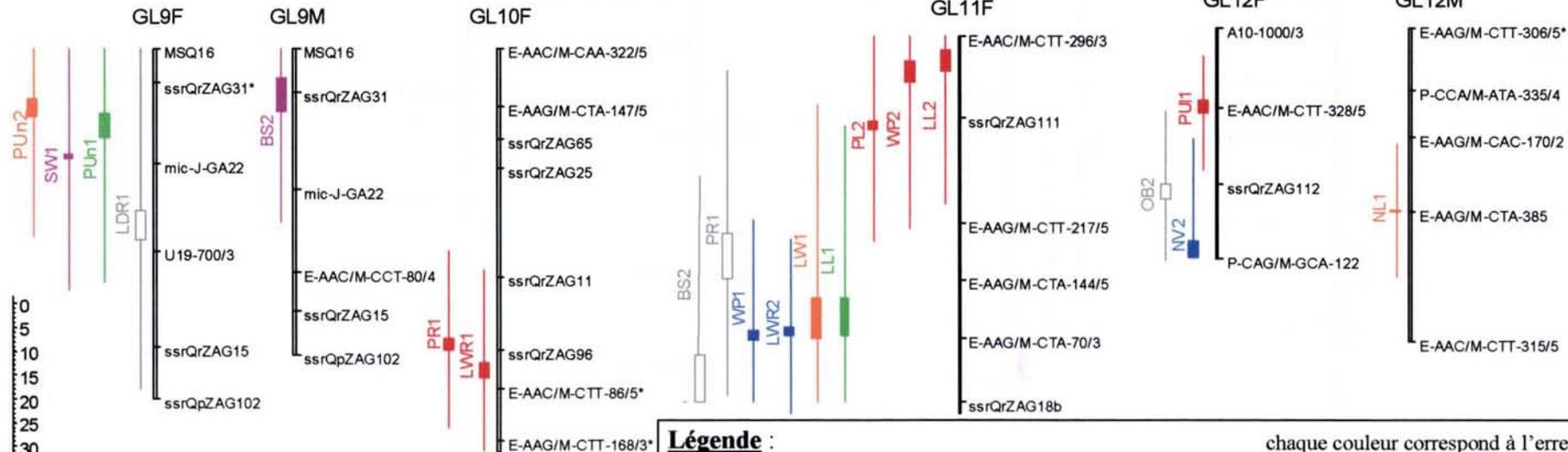
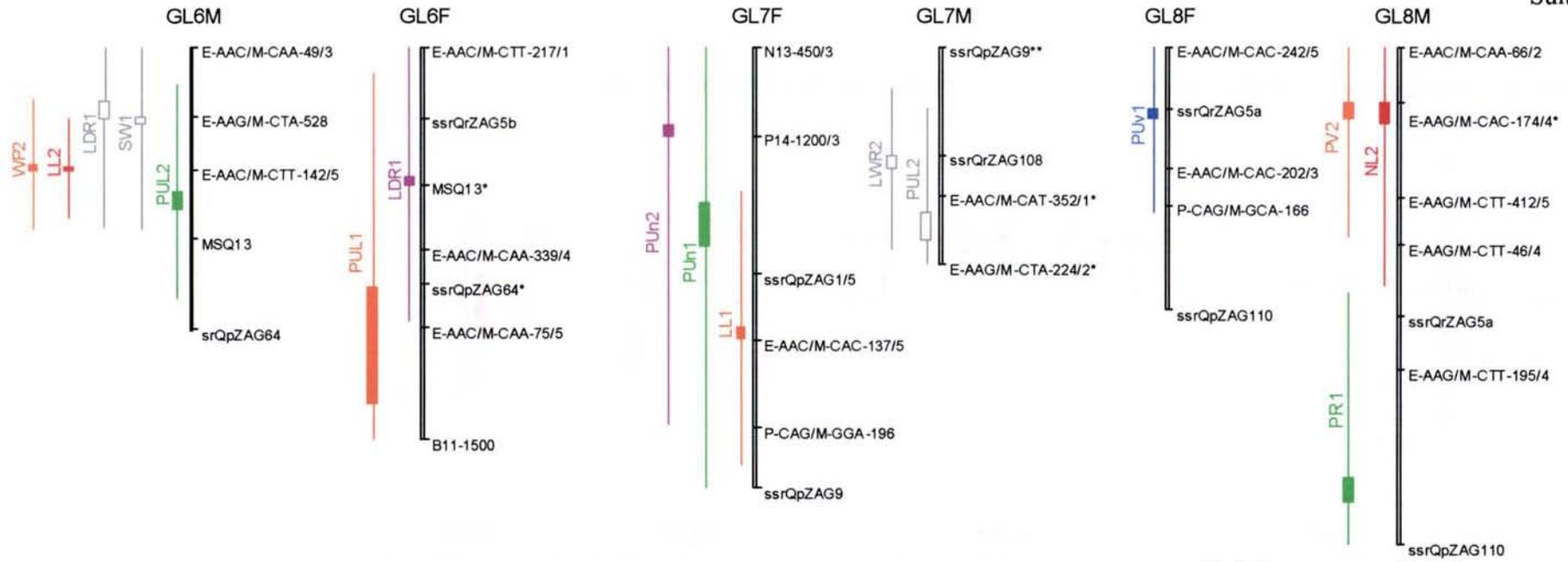
Des colocalisations entre deux QTL impliqués dans la différenciation entre les deux espèces sont observées sur le GL1 (PL, PR, et PUI), et le GL3 (PV, NV, SW, LDR).

Si l'on considère maintenant l'ensemble des QTL détectés au seuil de 5 % au niveau du chromosome, 37 QTL ont été détectés dans l'expérimentation 1 et 33 QTL dans l'expérimentation 2. Entre 0 (OB) et 5 QTL (SW, BS) sont détectés par caractère avec une variance des moyennes de clones expliquée par le QTL qui varie entre 3.1 et 22.9% (Tableau 11). Il ne semble pas qu'il y ait de relation entre répétabilité et nombre de QTL détectés ou proportion de variance expliquée par les QTL. Par exemple pour la pubescence qui est le caractère le plus héritable, 2 et 4 QTL ont été détectés respectivement dans les expérimentations 1 et 2 expliquant chacun entre 4.1 et 17.2 % de la variance des moyennes de clone (Tableau 11). Un contrôle génétique similaire est observé pour la variable LDR qui présente une répétabilité plus faible, où 3 et 5 QTL ont été détectés dans les expérimentations 1 et 2 expliquant chacun entre 5.5 et 19.1 % de la variance des moyennes des clones.

Six QTL impliqués dans 5 caractères parmi les 15 étudiés, sont détectés sur les mêmes groupes de liaisons dans les expérimentations 1 et 2 avec un effet  $D$  similaire et de même signe (PL sur GL1F, NV sur GL3F, SW sur GL3F, LDR sur GL3F et  $PU_V$  sur GL7F et GL9F). Les autres QTL spécifiques d'une expérimentation peuvent s'expliquer par les faibles corrélations observées entre les deux expérimentations.

**Figure 12 :** Distribution sur la carte génétique du chêne pédonculé des QTL impliqués dans la morphologie foliaire.





0  
5  
10  
15  
20  
25  
30  
35  
40  
45  
50 cM

**Légende :**

P1 : position la plus probable du QTL d'après la valeur maximale du LOD  
 P2 : position la plus probable du QTL après 1000 ré-échantillonnages réalisés par bootstrap  
 IC : intervalle de confiance du QTL obtenu après 1000 ré-échantillonnages.

chaque couleur correspond à l'erreur de type 1 au niveau du génome :

5%	20%
10%	30%
15%	50%

Si l'on se focalise sur les 7 caractères les plus discriminants sans prendre en compte leur variable dérivée (NL, PL, SW, NV, BS, PUI et PU<sub>v</sub>) entre 1 et 5 QTL ont été détectés dans chaque expérimentation ayant chacun des effets faibles à modérés. Les pourcentages de variance expliquée plus élevés dans l'expérimentation 1 sont certainement surestimés compte tenu du plus faible nombre d'individus pris en compte dans l'analyse. Les clusters entre caractères discriminants décrits en se fixant un seuil de 5% au niveau du génome sont renforcés lorsque l'on prend en compte tous les QTL détectés au seuil de 5% au niveau du chromosome (Figure 12) :

- sur GL1F dans une fenêtre de 30 cM: PUI<sub>1</sub>, PL<sub>1</sub>, PL<sub>2</sub>, BS<sub>2</sub> avec PUI<sub>1</sub> présent aussi sur GL1M
- sur GL2F dans une fenêtre de 60 cM NV<sub>1</sub> et NL<sub>1</sub>
- sur GL3F dans une fenêtre de 20 cM: NV<sub>1</sub>, NV<sub>2</sub>, PL<sub>2</sub>, SW<sub>1</sub>, SW<sub>2</sub>
- sur GL3M dans une fenêtre de 25 cM: BS<sub>2</sub>, SW<sub>1</sub>, SW<sub>2</sub>
- sur GL4M dans une fenêtre de 30 cM: PL<sub>1</sub> et PU<sub>v2</sub>
- sur GL6M dans une fenêtre de 40 cM: PUI<sub>2</sub>, SW<sub>1</sub> avec PUI<sub>1</sub> aussi présent sur GL6F
- sur GL7F dans une fenêtre de 30 cM: PU<sub>v1</sub>, PU<sub>v2</sub> et PUI<sub>2</sub> sur GL7M
- sur GL9F dans une fenêtre de 10 cM: PU<sub>v1</sub>, PU<sub>v2</sub>, SW<sub>1</sub> et BS<sub>2</sub> sur GL9M
- sur GL7F dans une fenêtre de 30 cM: PU<sub>v1</sub>, PU<sub>v2</sub> et PUI<sub>2</sub> sur GL7M

#### C.2.1.1.3 Conclusion

Les caractères impliqués dans la différenciation phénotypique entre le chêne sessile et le chêne pédonculé sont contrôlés par plusieurs gènes ayant des effets faibles à modérés malgré de forte héritabilité pour certains d'entre eux. Les QTL sont répartis sur l'ensemble du génome avec néanmoins la présence de quelques clusters notamment sur le chromosome 3.

### C.2.1.2 QTL impliqués dans l'efficacité d'utilisation de l'eau

Le chêne sessile et le chêne pédonculé présentent des réponses très différentes vis à vis de la sécheresse. On a pu observer chez le chêne pédonculé un dépérissement après des périodes de sécheresses prolongées alors que le chêne sessile est moins affecté. Ce fut par exemple le cas en France après les sécheresses de 1976 (Becker et Levy, 1982). La distribution écologique des deux espèces de chênes est en accord avec ces observations, le chêne pédonculé étant présent sur des sols fortement hydromorphes alors que le chêne sessile sera préférentiellement sur des sols plus secs. Afin de comprendre les bases écophysiologiques de ces différentes niches écologiques, la tolérance à la sécheresse a pu être étudiée en mesurant l'efficacité d'utilisation de l'eau en adoptant l'approche de discrimination isotopique du carbone. La théorie isotopique prévoit, pour une concentration atmosphérique en CO<sub>2</sub> donnée, une relation linéaire entre, (1) la discrimination isotopique du carbone ( $\Delta$ ), définie en première approximation comme la différence entre la composition isotopique de l'air et celle de la plante ( $\Delta = \delta^{13}C_{\text{air}} - \delta^{13}C_{\text{plante}}$ ) et (2), la concentration intercellulaire en CO<sub>2</sub>, elle-même linéairement liée à l'efficacité d'utilisation de l'eau intrinsèque foliaire ( $WUE_{\text{int}}$ ) (Farquhar *et al.*, 1989).  $WUE_{\text{int}}$ , défini comme le rapport entre le taux d'assimilation de CO<sub>2</sub> et la conductance stomatique pour la vapeur d'eau ( $A/g$ ) est relié à l'efficacité d'utilisation de l'eau par l'équation :  $WUE = \frac{WUE_{\text{int}}}{v}$  où  $v$ , une variable imposée par le climat, est la différence de pression partielle de H<sub>2</sub>O entre la feuille et l'atmosphère. Pour nos expérimentations les différents génotypes sont comparés dans des conditions environnementales équivalentes ( $v$  est une constante).

C'est par cette approche que Ponton *et al.*, 2001 ont mis en évidence dans des peuplements mixtes de chênes sessiles et chênes pédonculés adultes, une forte variabilité interspécifique (jusqu'à 23% de la variabilité totale) pour l'efficacité d'utilisation de l'eau dans la cellulose bois. Le chêne pédonculé présente une plus forte discrimination isotopique que le chêne sessile et par conséquent une plus faible efficacité d'utilisation de l'eau intrinsèque (moins 13% en moyenne).

**Tableau 12** : Décomposition de la variance pour l'efficiencia d'utilisation de l'eau.

caractère	Moyenne	Min	Max	CV <sub>G</sub>	CV <sub>R</sub>	R	Héritabilité	
							V <sub>D</sub> =0	V <sub>D</sub> =1/2V <sub>A</sub>
δ <sup>13</sup> C_1	-25.77	-27.98	-23.42	0.018	0.024	0.74	0.51	0.32
δ <sup>13</sup> C_2	-26.00	-27.74	-24.07	0.026	0.019	0.79	0.80	0.51
δ <sup>13</sup> C_3	-26.26	-27.60	-24.76	0.018	0.019	0.81	0.63	0.40

CV<sub>G</sub> coefficient de variation génétique  $\frac{\sigma_{clone}}{\mu}$  calculé à partir du modèle I

CV<sub>R</sub> coefficient de variation résiduelle  $\frac{\sigma_{bouture}}{\mu}$  calculé à partir du modèle I

R répétabilité des moyennes de clones

**Tableau 13** : Coefficient de corrélation de Pearson.

	δ <sup>13</sup> C_1	δ <sup>13</sup> C_2
δ <sup>13</sup> C_2	<b>0.65</b>	
δ <sup>13</sup> C_3	<b>0.63</b>	<b>0.71</b>

En gras les corrélations sont significatives

Les différences interspécifiques pour la réponse à la sécheresse mesurée sur des plantules indiquent de faibles interactions génotype milieu pour la discrimination isotopique du carbone (Ponton *et al.*, 2002). Ce qui est en accord avec un contrôle génétique fort illustré par de fortes héritabilités. ( $h^2 = 0.6$  à  $0.9$ , Farquhar *et al.*, 1989).

#### C.2.1.2.1 Matériel et Méthodes

L'efficacité d'utilisation de l'eau a été mesurée durant 3 années consécutives par nos collègues de l'INRA de Nancy dans le cadre du projet OAKFLOW. C'est à eux que nous empruntons ces données, fruit d'une coopération entre les deux laboratoires.

-en 2000, 173 clones en moyenne répétés 5.5 fois ont été mesurés dans le dispositif 2 (Tableau 4).

-en 2001, 121 clones en moyenne répétés 2 fois ont été mesurés dans le dispositif 2 (Tableau 4)

-en 2002, 207 clones en moyenne répétés 5 fois ont été mesurés dans le dispositif 3 (Tableau 4)

Le  $\delta^{13}C_{\text{plante}}$  a été mesuré sur une feuille située sur la première unité de croissance de chaque arbre en utilisant un analyseur élémentaire (dosage de la concentration en carbone de l'échantillon) couplé à un spectromètre de masse.

#### C.2.1.2.2 Résultats

##### C.2.1.2.2.1 Décomposition de la variance phénotypique

L'effet clone testé par ANOVA était hautement significatif dans les 3 expérimentations. Les répétabilités des moyennes de clones et les héritabilités sont fortes pour les 3 années de mesures, variant entre 74 et 81% pour les répétabilités et entre 32 et 80% pour l'héritabilité (Tableau 12). Ces fortes héritabilités sont en accord avec celles estimées par Farquhar *et al.* (1989) et indiquent que l'efficacité d'utilisation de l'eau est un caractère fortement contrôlé génétiquement. La stabilité entre les années et les dispositifs de la performance des moyennes des clones est confirmée par de fortes corrélations entre les années de mesures qui varient entre 0.63 et 0.71 (Tableau 13).

**Tableau 14** : Liste des QTL impliqués dans l'efficacité d'utilisation de l'eau détectés au seuil de 5% au niveau du chromosome.

Caractère	GL	N	Lod	$\alpha_{chr}$	$\alpha_G$	P1	P2	IC	D	P	$V_D=0$		$V_D=1/2V_A$	
											S	PV	S	PV
$\delta^{13}C_{-1}$	1f	144	2.1	0.015	20%	68	47.4	57.8	0.27	6.3	4.2	2.2	7.4	2.4
$\delta^{13}C_{-2}$	2m	88	3.6	0.001	5%	97.6	90.8	32.5	0.52	12.7	8.0	6.4	14.0	7.2
$\delta^{13}C_{-2}$	3f	74	1.4	0.031	50%	0	3.8	16.5	0.42	8.5	5.3	4.3	9.4	4.8
$\delta^{13}C_{-3}$	4f	171	1.8	0.03	50%	8.4	50.8	85.3	-0.21	4.7	2.9	1.8	5.1	2.0
$\delta^{13}C_{-3}$	6m	178	2.5	0.004	5%	14.6	15	11.5	0.20	4.4	2.7	1.7	4.7	1.9
$\delta^{13}C_{-1}$	8m	130	5.6	0.001	5%	47.9	44.8	17.0	0.42	15.8	9.7	6.1	17.0	6.8
$\delta^{13}C_{-3}$	8m	160	1.9	0.024	30%	31.4	34.3	40.4	0.18	3.6	2.2	1.4	3.9	1.5
$\delta^{13}C_{-2}$	8m	92	1.9	0.026	30%	31.6	41.4	35.3	0.36	5.4	3.4	2.7	5.9	3.1
$\delta^{13}C_{-1}$	9m	153	2.2	0.016	20%	15.7	15.6	25.1	0.21	3.8	2.6	1.3	4.5	1.4
$\delta^{13}C_{-3}$	9m	156	2.8	0.003	5%	0	6.67	31.0	0.21	4.9	3.0	1.9	5.3	2.1
$\delta^{13}C_{-3}$	10m	100	1.8	0.038	50%	0	9.25	32.1	0.23	5.4	3.3	2.1	5.8	2.3
$\delta^{13}C_{-3}$	11f	175	1.7	0.04	50%	17.7	26.6	41.0	0.19	4.1	2.5	1.6	4.4	1.8
$\delta^{13}C_{-1}$	11m	160	10.53	0.001	5%	57.1	57.7	9.0	-0.51	22.0	14.8	7.6	25.9	8.2
$\delta^{13}C_{-3}$	11m	190	11.46	0.001	5%	64.3	64.6	7.9	-0.44	21.7	13.4	8.4	23.4	9.3
$\delta^{13}C_{-2}$	11m	106	11.13	0.001	5%	64.2	63.8	4.2	-0.89	34.4	21.6	17.2	37.9	19.5
Moyenne										10.5	6.7	4.5	11.6	4.9

GL : le groupe de liaison (m = carte mâle et f = carte femelle)

N : le nombre d'individus pris en compte dans l'analyse

Lod : la valeur de Lod maximale obtenue sur le chromosome après 1000 permutations

$\alpha_{chr}$  : L'erreur de type I au niveau du chromosome obtenue après 1000 permutations

$\alpha_G$  : L'erreur de type I en pourcentage au niveau du génome

P1 : La position la plus probable du QTL sur le chromosome obtenue par permutation

P2 : La position du QTL la plus probable après 1000 ré-échantillonnage par bootstrap

IC : intervalle de confiance du QTL obtenu après 1000 ré-échantillonnage par bootstrap

D : effet de substitution allélique au QTL

P : le pourcentage de variance des moyennes de clones expliquée par le QTL dans la descendance

S : le pourcentage de variance additive expliquée par le QTL dans la population d'où sont issus les parents pour deux valeurs contrastées de variance de dominance ( $V_D=0$  et  $V_D=1/2V_A$ )

PV : le pourcentage de variance phénotypique expliquée par le QTL dans la population d'où sont issus les parents pour deux valeurs contrastées de variance de dominance ( $V_D=0$  et  $V_D=1/2V_A$ )

$\delta^{13}C_{-1}$  : mesure réalisée en 2000 dans le dispositif 2

$\delta^{13}C_{-2}$  : mesure réalisée en 2001 dans le dispositif 2

$\delta^{13}C_{-3}$  : mesure réalisée en 2002 dans le dispositif 3.

**Tableau 15** : Erreur de type I des colocalisations entre QTL impliqués dans la discrimination isotopique du carbone détectés sur plusieurs années.

Carte	Années	GL	N	P	G	C	Pvalue
Mâle	1 - 2	8 ; 11	121	3	3	2	<b>0.001</b>
Mâle	1 - 3	8 ; 9 ; 11	121	3	5	3	<b>3.5E-05</b>
Mâle	2 - 3	8 ; 11	121	3	5	2	<b>0.004</b>

GL : les groupes de liaison où une colocalisation est observée

N : le nombre d'intervalle comparés. Correspond au rapport entre la taille de la carte en cM et l'intervalle moyen de la position la plus probable du QTL (moyenne de |P1-P2| dans le tableau 14).

P : nombre de QTL détecté l'année qui donne le plus petit nombre de QTL

G : nombre de QTL détecté l'année qui donne le plus grand nombre de QTL

C : nombre de colocalisations observées entre deux années

Pvalue : Probabilité d'observer C colocalisations par chance entre deux années

$$Pvalue = \frac{C_G^C \times C_{N-G}^{P-C}}{C_N^P}$$

#### **C.2.1.2.2.2 Détection de QTL**

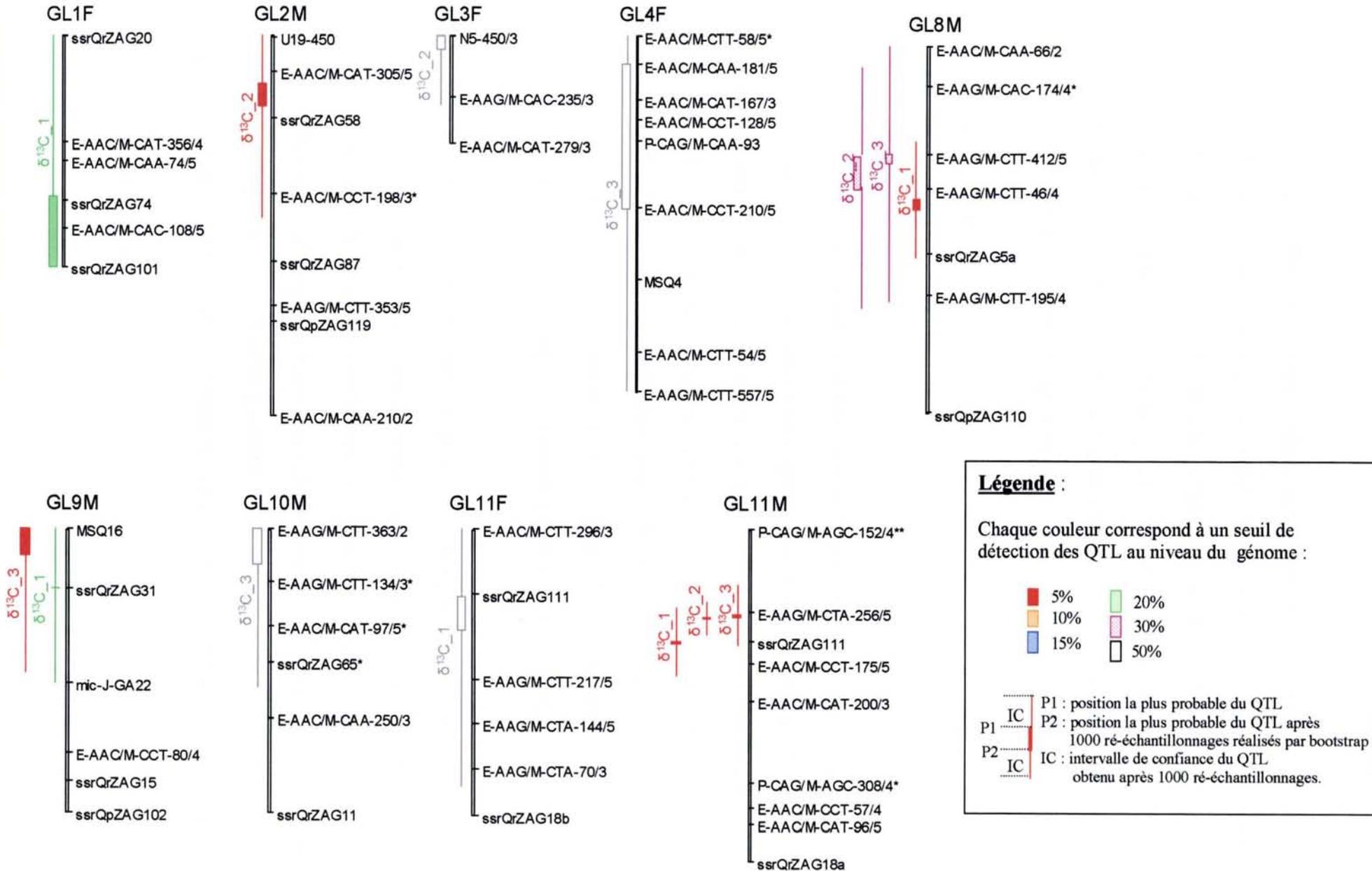
Quinze QTL significatifs à 5% au niveau du chromosome ont été détectés qui expliquent entre 3.6 et 34.4% de la variance des moyennes de clone (Tableau 14) et entre 2.2 et 37.9 % de la variance additive. Sur les groupes de liaisons 11m et 8m deux QTL présentent des effets majeurs expliquant plus de 15.8 % de la variance des moyennes de clones. Sur le groupe de liaison 11 l'importance du QTL dans l'expression de l'efficacité de l'utilisation de l'eau est renforcée par la détection de QTL durant 3 années consécutives avec des variances de moyenne de clones qui varient entre 21.7 et 34.4%.

Seules des colocalisations sur la carte mâle sont observées (Figure 13). Le groupe de liaison 9m présente une colocalisation entre l'expérimentation 1 et 3 et les groupes de liaison 8m et 11m présentent des colocalisations entre les trois expérimentations. Si l'on considère qu'en moyenne l'intervalle de la position du QTL est de 7.73cM (moyenne de la différence entre P1 et P2 voir tableau 14) alors 121 intervalles (933cM/7.73) peuvent être comparés sur la carte mâle pour réaliser le test de colocalisation (voir chapitre C.1.3.3). Le test de colocalisation est significatif entre toutes les paires d'années (Tableau 15) ce qui traduit une bonne stabilité dans le temps mais aussi dans l'espace de ces 3 QTL.

#### **C.2.1.2.2.3 Conclusion**

L'efficacité d'utilisation de l'eau est un caractère fortement héritable, stable entre les années et entre les dispositifs comme en témoignent les fortes corrélations entre les moyennes de clones ainsi que les colocalisations entre les QTL détectés sur trois années sur les groupes 8m, 9m et 11m. L'efficacité d'utilisation de l'eau semble être contrôlée par un gène à effet majeur situé sur le groupe de liaison 11 associé à plusieurs gènes ayant des effets faibles à modérés. L'importance de ce QTL est confirmée par sa présence chez le châtaignier sur le même groupe de liaison (Manuela Casasoli, communication personnelle).

**Figure 13 :** Distribution sur la carte génétique du chêne pédonculé des QTL impliqués dans la discrimination isotopique du carbone.



### C.2.1.3 QTL impliqués dans la croissance

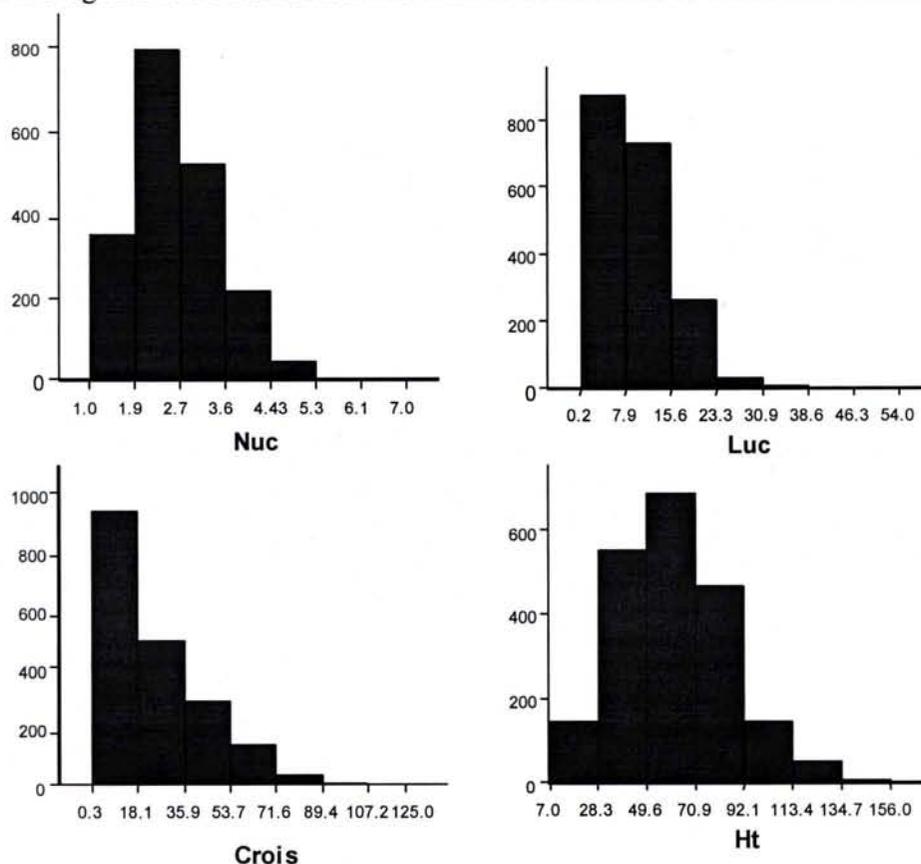
La croissance en hauteur du chêne n'est pas continue mais polycyclique. On parle souvent de la pousse de mai, de la pousse de la St Jean et de la pousse d'août, chaque pousse étant appelée une Unité de Croissance (UC). La croissance n'est pas clairement impliquée dans la différenciation entre le chêne sessile et le chêne pédonculé au stade adulte. Cependant, dans les deux premières années de croissance le chêne pédonculé présente une croissance plus rapide que le chêne sessile (Ponton *et al.*, 2002), illustrant ainsi la capacité du chêne pédonculé à dominer le chêne sessile dans les premiers stades de régénération des forêts.

#### C.2.1.3.1 Mesure des caractères

Différentes composantes de la croissance ont été mesurées durant l'hiver 2001 sur des boutures installées dans le dispositif 3 (Tableau 4) qui terminaient leur troisième saison de croissance depuis le bouturage. Les caractères mesurés étaient :

- le nombre d'unité de croissance ( $N_{UC}$ ),
- la longueur moyenne d'une UC ( $L_{UC}$ ),
- la croissance totale de l'arbre durant la troisième saison (Crois), égale à la somme de toutes les UC,
- la hauteur totale de chaque arbre à la fin de la troisième saison de végétation (Ht).

**Figure 14 :** Histogramme de distribution des différents caractères de croissance en hauteur.



**Tableau 16 :** Décomposition de la variance pour la croissance en hauteur.

Caractère	Moyenne	Min	Max	CV <sub>G</sub>	CV <sub>R</sub>	R	Héritabilité	
							VD=0	VD=1/2VA
Ht	62.2 cm	7 cm	156 cm	0.13	0.34	0.58	0.23	0.14
Crois	24.4 cm	0.3 cm	125 cm	0.25	0.70	0.54	0.21	0.12
L <sub>UC</sub>	9.5 cm	0,2 cm	54 cm	0.18	0.59	0.45	0.16	0.09
N <sub>UC</sub>	2.4	1	7	0.08	0.39	0.26	0.07	0.04

CV<sub>G</sub> coefficient de variation génétique  $\frac{\sigma_{clone}}{\mu}$  calculé à partir du modèle I

CV<sub>R</sub> coefficient de variation résiduelle  $\frac{\sigma_{bouture}}{\mu}$  calculé à partir du modèle I

R répétabilité des moyennes de clones

**Tableau 17 :** Coefficient de corrélation de Pearson entre moyennes de clones.

	Ht	Crois	L <sub>UC</sub>
Ht			
Crois	<b>0.51</b>		
L <sub>UC</sub>	<b>0.44</b>	<b>0.91</b>	
N <sub>UC</sub>	<b>0.40</b>	<b>0.72</b>	<b>0.39</b>

En gras les valeurs sont significatives

### C.2.1.3.2 Résultats

#### C.2.1.3.2.1 Décomposition de la variance phénotypique

La distribution des valeurs phénotypiques a été testée par la méthode de Box et Cox (1964). Les histogrammes de distribution sont présentés sur la figure 14. Les valeurs individuelles pour le nombre d'unité de croissance ( $N_{UC}$ ) et pour la hauteur totale (Ht) sont normalement distribuées alors que pour la croissance annuelle (Crois) et la longueur moyenne d'une unité de croissance ( $L_{UC}$ ) une transformation  $\sqrt{x}$  a été réalisée pour la détection de QTL.

La variation intra clone pour l'ensemble des caractères de croissance est très forte comme en témoignent les forts coefficients de variation résiduelle qui peuvent atteindre jusqu'à 70% pour la croissance annuelle (Crois) (Tableau 16). Ces variations individuelles peuvent notamment s'expliquer par l'effet de la transplantation des boutures. En dépit d'une forte variance résiduelle, les répétabilités sont relativement fortes (supérieures à 45%) excepté pour le nombre d'unité de croissance ( $N_{UC}$ ) où la répétabilité est modérée ( $R=26\%$ ). De fortes répétabilités pour les caractères de croissance en hauteur ont aussi été observées chez d'autres espèces d'arbres forestiers. Chez le saule et le peuplier par exemple, la répétabilité de la croissance annuelle en hauteur varie en fonction des années entre 57 et 70% pour le saule (Tsarouhas *et al.*, 2002) et entre 69 et 89% pour le peuplier (Wu et Stettler, 1996).

#### C.2.1.3.2.2 Corrélations entre caractères

Les 4 caractères de croissance présentent tous des corrélations significatives. Les corrélations les plus fortes sont obtenues pour les couples Crois /  $L_{UC}$  et Crois /  $N_{UC}$  avec un r respectif de 0.91 et 0.72 indiquant ainsi que les clones qui poussent le plus tendent à avoir un plus grand nombre d'unités de croissance et de plus longues unités de croissance. (Tableau 17).

**Tableau 18** : Détection de QTL impliqués dans la croissance en hauteur.

Caractère	GL	N	LoD	$\alpha_{chr}$	$\alpha_G$	P1	P2	IC	D	P	S		PV	
											$V_D=0$	$V_D=1/2 V_A$	$V_D=0$	$V_D=1/2 V_A$
Ht	5m	164	2.5	0.004	5%	53.4	48.9	28.4	-5.8	6.7	5.8	1.3	10.1	1.4
Ht	6m	188	3.2	0.001	5%	28.3	29	13.4	-5.7	6.4	5.5	1.3	9.7	1.3
Ht	10m	170	7.5	0.001	5%	18.4	19.1	12.8	-9.6	17.3	14.9	3.5	26.1	3.6
Ht	11f	172	3.5	0.001	5%	40.0	36.3	30.8	6.9	9.1	7.8	1.8	13.7	1.9
Crois	3m	191	5.1	0.001	5%	22.1	27	19.3	-0.5	9.8	9.5	1.8	16.7	1.8
Crois	10f	178	3.9	0.001	5%	12.2	16.1	26.5	-0.5	9.5	9.2	1.7	16.2	1.8
Crois	10m	170	8.2	0.001	5%	15.5	16.5	9.6	-0.7	18.7	18.2	3.4	31.8	3.5
L <sub>UC</sub>	3m	191	5.5	0.001	5%	21.3	24.6	18.2	-0.3	10.7	11.6	1.8	20.3	1.9
L <sub>UC</sub>	10m	170	7.0	0.001	5%	15.8	17.3	11.3	-0.3	16.4	17.8	2.8	31.2	2.9
L <sub>UC</sub>	10f	170	2.0	0.014	15%	5.1	18.3	42.1	-0.2	5.9	6.4	1.0	11.2	1.0
N <sub>UC</sub>	7f	166	1.8	0.032	30%	30.7	35.9	32.5	0.1	4.1	7.8	0.6	13.6	0.6
N <sub>UC</sub>	10f	178	6.1	0.001	5%	12.2	16.1	17.0	-0.3	13.7	25.9	1.9	45.3	1.9
N <sub>UC</sub>	10m	170	3.1	0.001	5%	17.9	23.7	22.9	-0.2	8.9	16.8	1.2	29.5	1.2
N <sub>UC</sub>	11f	193	2.5	0.008	5%	51.1	44.4	25.1	0.2	5.1	9.6	0.7	16.9	0.7
Moyenne										10.2	11.9	1.8	20.9	1.8

Avec

GL : le groupe de liaison (m = carte mâle et f = carte femelle)

N : le nombre d'individus pris en compte dans l'analyse

Lod : la valeur de Lod maximale obtenue sur le chromosome après 1000 permutations

$\alpha_{chr}$  : L'erreur de type I au niveau du chromosome obtenue après 1000 permutations

$\alpha_G$  : L'erreur de type I en pourcentage au niveau du génome

P1 : La position la plus probable du QTL sur le chromosome d'après la valeur de LOD maximale

P2 : La position du QTL la plus probable après 1000 ré-échantillonnages réalisés par bootstrap

IC : intervalle de confiance du QTL obtenu après 1000 ré-échantillonnages réalisés par bootstrap

D : effet de substitution allélique au QTL

P : le pourcentage de variance des moyennes de clones expliquée par le QTL dans la descendance

S : le pourcentage de variance additive expliquée par le QTL dans la population d'où sont issus les parents pour deux valeurs contrastées de variance de dominance ( $V_D=0$  et  $V_D=1/2V_A$ )

PV : le pourcentage de variance phénotypique expliquée par le QTL dans la population d'où sont issus les parents pour deux valeurs contrastées de variance de dominance ( $V_D=0$  et  $V_D=1/2V_A$ )

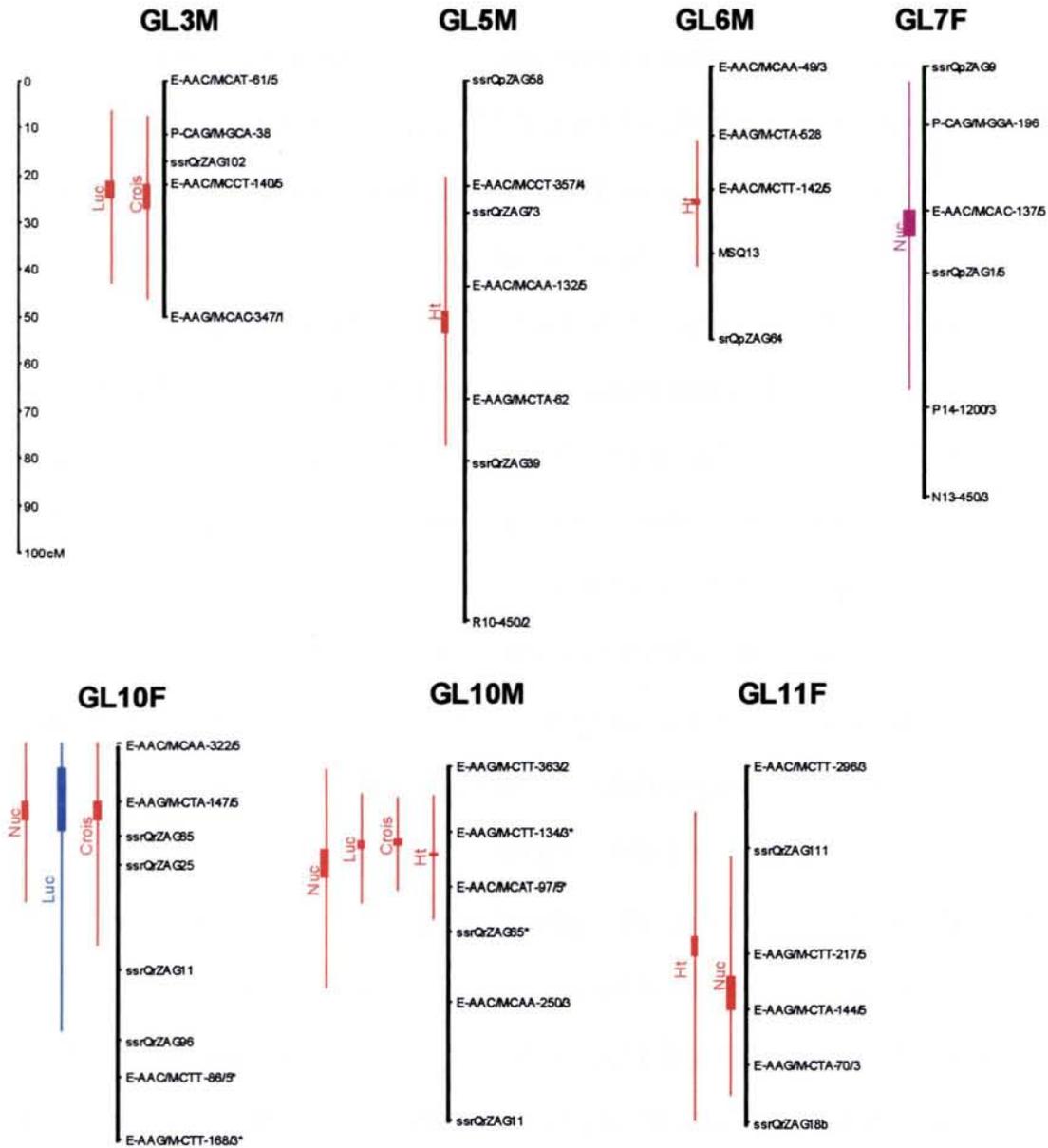
### C.2.1.3.2.3 Détection de QTL impliqués dans la croissance

Entre 3 et 4 QTL ont été détectés pour chaque composante de la croissance qui expliquent chacun entre 4.1 et 18, % de la variance clonale et entre 5.5 et 45.3% de la variance additive (Tableau 18).

Des résultats similaires ont été obtenus chez *Pinus Taeda*, où suivant les années et les pedigrees, entre 2 et 4 QTL impliqués dans la croissance annuelle ont été détectés qui expliquent chacun entre 7.3 et 30% de la variance phénotypique (Kaya *et al.*, 1999). Chez le pin sylvestre (Lerceteau *et al.*, 2000) la contribution individuelle des 3 QTL impliqués dans la hauteur totale de l'arbre est plus modérée variant entre 11.5 et 12.9% cependant, leur contribution globale à la variance phénotypique atteint 25.8%. Enfin chez le saule, 4 QTL ont été détectés pour la hauteur totale qui expliquent entre 13 et 18% de la variance phénotypique et 3 QTL ont été détectés pour la croissance annuelle qui expliquent entre 12 et 17% (Tsarouhas *et al.*, 2002). L'interaction entre locus a été testée avec le logiciel Qgene mais aucun effet d'épistasie n'a été détecté au seuil de 1% au niveau du génome. L'absence d'épistasie avait aussi été mentionnée chez le pêcher (Conner *et al.*, 1998). Par contre chez le saule, des effets significatifs ont été observés entre 2 QTL (Tsarouhas *et al.*, 2002).

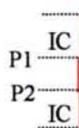
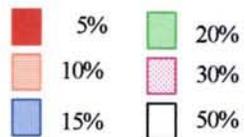
La position des QTL sur la carte (Figure 15) suggère la présence d'un QTL commun aux 4 composantes de la croissance sur le chromosome 10 et commun à 2 composantes sur les chromosomes 11 (entre Ht et N<sub>UC</sub>) et 3 (entre L<sub>UC</sub> et Crois). La cartographie comparée des QTL impliqués dans la croissance détectés chez le chêne pédonculé et chez le châtaignier montre que chez cette deuxième espèce le groupe de liaison 10 présente aussi des QTL impliqués dans la croissance (Manuela Casasoli, communication personnelle). Ce résultat renforce l'importance du groupe 10 dans le déterminisme génétique de la croissance.

**Figure 15** : Distribution sur la carte génétique du chêne pédonculé des QTL impliqués dans la croissance en hauteur.



**Légende :**

Chaque couleur correspond à différents seuils de l'erreur de type I au niveau du génome



P1 : position la plus probable du QTL  
 P2 : position la plus probable du QTL après 1000 ré-échantillonnages réalisés par bootstrap  
 IC : intervalle de confiance du QTL obtenu après 1000 ré-échantillonnages.

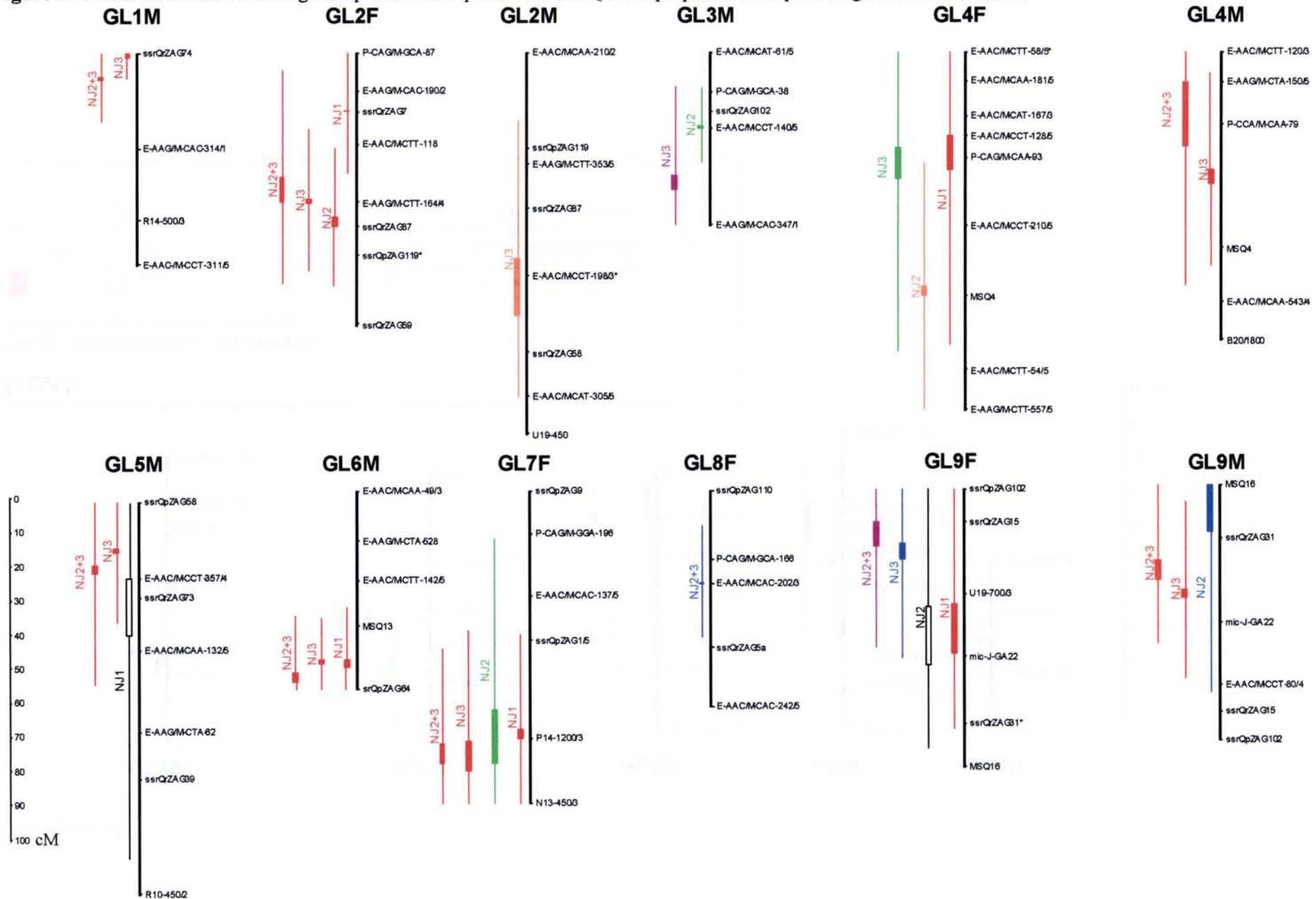
Nuc : le nombre d'unité de croissance (UC)  
 Luc : la longueur moyenne d'une UC  
 Crois : l'accroissement annuel en hauteur (somme des UC)  
 Ht : la hauteur totale

#### C.2.1.4 Conclusion

Le déterminisme génétique des caractères impliqués dans la différenciation phénotypique entre le chêne sessile et le chêne pédonculé diffère suivant les caractères considérés. Pour la morphologie foliaire entre 1 et 5 QTL ont été détectés au seuil de 5% au niveau du chromosome avec des variances de moyenne de clone expliquées qui varient entre 3.7 et 22.9 %. Il semble au vu des résultats que la morphologie foliaire soit contrôlée par quelques gènes ayant des effets faibles à modérés. Aucun gène à effet majeur n'a été mis en évidence.

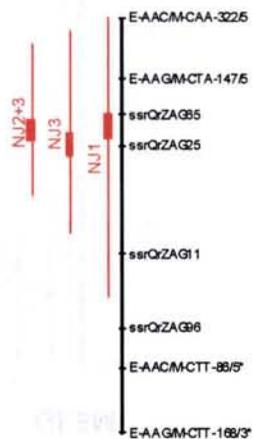
Pour l'efficacité d'utilisation de l'eau 15 QTL ont été détectés qui expliquent entre 4.4 et 34.4% de la variance des moyennes de clones. Il semble que ce caractère soit contrôlé par un gène à effet majeur situé sur le groupe de liaison 11f. La localisation sur ce groupe d'un QTL fortement impliqué dans l'efficacité d'utilisation de l'eau est confirmée par sa présence sur la carte génétique du châtaignier (Manuela Casasoli, communication personnelle). Enfin pour la croissance entre 3 et 4 QTL ont été détectés suivant les caractères avec des variances de moyenne clone expliquées qui varient entre 4.1 et 18.7%. La présence sur le groupe de liaison 10f d'un regroupement de QTL à effets relativement forts (entre 13.7 et 18.7%) est confirmée par la présence sur ce même groupe de liaison d'un QTL impliqué dans la hauteur totale sur la carte génétique du châtaignier (Manuela Casasoli, communication personnelle).

**Figure 17** : Distribution sur la carte génétique du chêne pédonculé des QTL impliqués dans la phénologie du débournement.

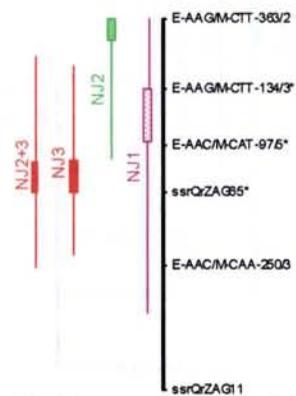


Suite figure 17 :

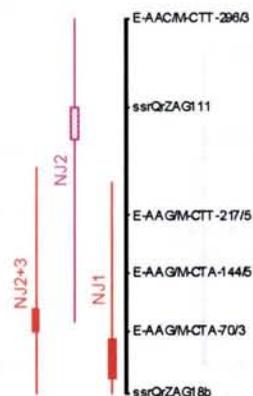
### GL10F



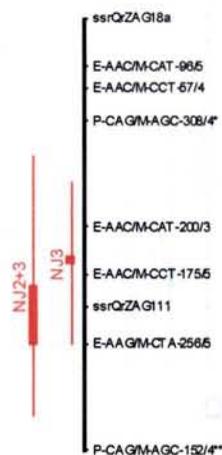
### GL10M



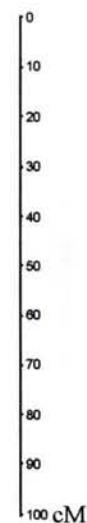
### GL11F



### GL11M

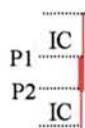
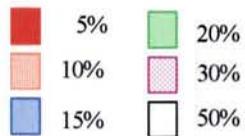


### GL12F



### Légende :

Chaque couleur correspond à différents seuils de l'erreur de type I au niveau du génome



P1 : position la plus probable du QTL  
 P2 : position la plus probable du QTL après 1000 ré-échantillonnages réalisés par bootstrap  
 IC : intervalle de confiance du QTL obtenu après 1000 ré-échantillonnages.

## C.2.2 DETECTION DE QTL IMPLIQUES DANS DES CARACTERES QUI NE DIFFERENCIENT PAS LE CHENE SESSILE ET LE CHENE PEDONCULE

### C.2.2.1 QTL impliqués dans la phénologie du débourrement

#### C.2.2.1.1 Introduction

La phénologie du débourrement est un caractère adaptatif très corrélé avec les mesures de fitness. Il contribue à la viabilité et à la croissance des arbres en évitant des effets néfastes causés par des gels précoces ou tardifs; il contribue sans doute également à la fécondité des arbres dans la mesure où la synchronie de floraison est une condition nécessaire à une fructification efficace. Compte tenu de l'importance de la phénologie du débourrement, il est légitime de se demander à quelle vitesse évolue ce caractère en réponse à de fortes variations des conditions environnementales. Les arbres seront-ils capables par exemple d'évoluer assez rapidement pour répondre à un brusque réchauffement climatique ? La vitesse d'évolution d'un caractère en réponse à des variations de l'environnement est très dépendante de la variance génétique du caractère et du taux de migration des espèces. Ainsi, plus l'héritabilité d'un caractère est élevée plus la réponse des descendants à des pressions de sélection exercée sur une population mère sera rapide (Billington et Pelham, 1991).

L'étude de la génétique quantitative du débourrement dans différentes espèces d'arbres forestiers indique qu'il s'agit d'un caractère à contrôle génétique élevé. Des héritabilités très fortes qui varient entre 94 et 98% ont été obtenues chez le peuplier (Bradshaw *et al.*, 1995, Frewen *et al.*, 2000). Chez le chêne rouge, les héritabilités sont inférieures mais restent fortes (entre 40 et 50%, Ducousso *et al.*, 1997).

Chez le chêne, le débourrement présente des fortes différences entre populations comme l'indiquent les fortes valeurs de différenciation entre populations ( $Q_{ST}$ ) (Kremer *et al.*, 2000) mais aussi intra populations. Par exemple, la date de débourrement peut être décalée de plus d'un mois entre

différentes provenances mais aussi entre les arbres les plus précoces et les plus tardifs à l'intérieur d'une même provenance (Stephan *et al.* 1995 ; Ducouso *et al.*, 1996). Les populations présentent des patterns de variations géographiques de type clinal aussi bien en fonction de la latitude que de l'altitude (Ducouso *et al.*, 1997). Ainsi, le débourrement est un bon marqueur phénotypique de la différenciation entre populations. Par contre, il n'est pas impliqué dans la différenciation phénotypique entre le chêne sessile et le chêne pédonculé. Bacilieri dans sa thèse en observant la date de floraison de tous les arbres adultes du peuplement expérimental de la Petite Charnie (voir paragraphe C 2.1.1.1.1) a remarqué une parfaite synchronie entre les deux espèces.

La date de débourrement et plus généralement la phénologie est un caractère très étudié aujourd'hui, dans la mesure où il sera sans doute soumis à de fortes pressions sélectives si les changements climatiques annoncés se réalisent.

Des récentes études de détection de QTL impliqués dans la phénologie du débourrement donnent des résultats différents suivant les espèces. Chez le pin sylvestre seulement 4 QTL ont été détectés dont 2 présentent des effets majeurs (Hurme *et al.*, 2000) alors que chez le douglas et le peuplier respectivement 33 et 6 QTL à effets mineurs ont été détectés (Jermstad *et al.*, 2001 ; Frewen *et al.*, 2000).

#### **C.2.2.1.2 Mesures du débourrement**

La date de débourrement a été mesurée durant 3 années (1999, 2000, 2002) dans 3 dispositifs différents décrits dans le tableau 4.

- Dispositif 1 : En 1999 le débourrement a été mesuré sur les 278 pieds mères des descendants qui composent le parc à clone de Pierroton. Parmi les 278 clones, ont été exclus 38 individus qui n'avaient pas poussé dans les mêmes conditions (rapatriés du dispositif de Blanquefort 2 ans après leur installation pour cause de forte sécheresse). L'analyse QTL a donc été réalisée sur 240 clones en copie unique.

- Dispositif 2 : En 2000 le débourrement a été mesuré sur 174 descendants répétés en moyenne 6 fois localisés dans le dispositif expérimental de la tranche 1 de Bourran

- Dispositif 3 : En 2002 le débourrement a été mesuré sur 207 descendants répétés en moyenne 10.6 fois dans le dispositif expérimental de la tranche 2 de Bourran.

La date de débourrement a été mesurée régulièrement tous les 2 à 3 jours au mois d'avril en utilisant une échelle de mesure visuelle à 6 classes, allant du bourgeon dormant (stade 0) à l'épanouissement complet des feuilles (stade 5) (0 : bourgeon dormant, 1 : les écailles foliaires du bourgeon apical commencent à s'ouvrir, 2 : apparition de l'ébauche des feuilles, 3 : les premières feuilles sortent du bourgeon apical, 4 : feuilles sorties mais encore flétries, 5 : les feuilles sont complètement étalées).

Deux variables qui présentaient la plus grande variabilité ont été retenues pour l'analyse QTL :

-Le nombre de jours ( $N_j$ ) qui s'écoule entre le premier jour de passage et la notation pour la première fois d'un bourgeon au stade 3 (les premières feuilles sortent du bourgeon apical).

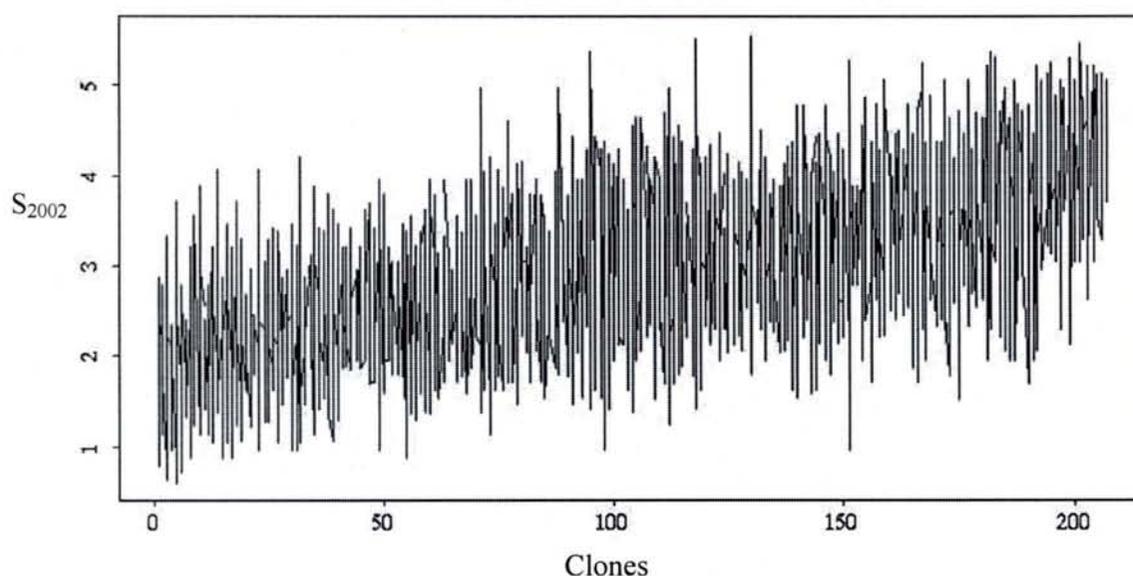
-Le stade de débourrement (S) au milieu de la période d'observation (J14 en 1999, J13 en 2000 et J11 en 2002)

### C.2.2.1.3 Résultats

#### C.2.2.1.3.1 Décomposition de la variance phénotypique

La distribution des valeurs phénotypiques testée par la méthode de Box et Cox (1964) indique que les valeurs sont normalement distribuées. Les valeurs phénotypiques individuelles des caractères mesurés dans les dispositifs 2 et 3 ont été ajustées à l'effet bloc avant d'être utilisées dans le modèle d'ANOVA I. Les composantes de la variance du débourrement mesuré en 1999 dans le parc à pied mère (dispositif 1) n'ont pas pu être estimées à cause de l'absence de répétitions. L'effet clone était très significatif pour l'ensemble des caractères. La variation de la date de débourrement est très importante entre les clones des 3 dispositifs. La durée globale de la période de débourrement à

**Figure 16** : Variabilité intra et inter clones du stade de débourrement observé le 11 ème jour de mesure en 2002 dans la tranche 2 de Bourran.



**Tableau 19** : Décomposition de la variance pour le débourrement

Caractère	Moyenne	Min	Max	CV <sub>G</sub>	CV <sub>R</sub>	R	Héritabilité	
							V <sub>D</sub> =0	V <sub>D</sub> =1/2V <sub>A</sub>
NJ1	13.6	1	21	0.16*				
NJ2	13.8	1	20	0.09	0.21	0.50	0.25	0.15
NJ3	10.0	1	21	0.27	0.37	0.84	0.52	0.32
S1	2.6	1	4	0.32*				
S2	2.7	0	5	0.15	0.33	0.56	0.31	0.18
S3	3.0	1	5	0.18	0.25	0.83	0.51	0.31

CV<sub>G</sub> coefficient de variation génétique  $\frac{\sigma_{clone}}{\mu}$  calculé à partir du modèle I

CV<sub>R</sub> coefficient de variation résiduelle  $\frac{\sigma_{bouture}}{\mu}$  calculé à partir du modèle I

\*Coefficient de variation phénotypique  
R répétabilité des moyennes de clones

**Tableau 20** : Coefficient de corrélation de Pearson.

	NJ1	S1	NJ2	S2	NJ3
S1	<b>-0.84</b>				
NJ2	<b>0.41</b>	<b>-0.35</b>			
S2	<b>-0.41</b>	<b>0.36</b>	<b>-0.94</b>		
NJ3	<b>0.61</b>	<b>-0.59</b>	<b>0.42</b>	<b>-0.47</b>	
S3	<b>-0.61</b>	<b>0.57</b>	<b>-0.4</b>	<b>0.46</b>	<b>-0.98</b>

Calculé à partir des moyennes de clones ajustées à l'effet bloc pour les mesures réalisées dans les dispositifs 2 et 3 et à partir des données individuelles pour les mesures réalisées dans le dispositif 1. Les valeurs en gras correspondent à des corrélations significatives.

l'intérieur de la famille de plein frères est en moyenne de 22 jours depuis le clone le plus précoce jusqu'au plus tardif.

Une illustration de l'effet clonal et de la variation intra clone est donnée pour S dans le dispositif 3 (Figure 16). A la moitié de la période d'observation (11 jours), certains clones sont toujours au stade dormant alors que d'autres présentent déjà des feuilles épanouies.

La répétabilité des moyennes de clones et l'héritabilité du débourrement varient fortement entre les dispositifs 2 et 3 (Tableau 19). Dans le dispositif 2, l'héritabilité est modérée variant entre 0.15 et 0.31 alors que dans le dispositif 3 elle est deux fois plus élevée variant de 0.31 à 0.52. Ces différences sont aussi illustrées par des coefficients de variation génétiques supérieurs dans le dispositif 3.

Cependant, à l'intérieur d'un dispositif, l'héritabilité pour les deux types de mesure du débourrement ( $N_J$  et S) donne des valeurs similaires.

#### C.2.2.1.3.2 Interaction clone\*test

Compte tenu de la proximité des tests clonaux 2 et 3 il nous a semblé intéressant de tester l'interaction entre le dispositif et la performance des clones pour pouvoir, en l'absence d'interaction, cumuler les données pour la détection de QTL et ainsi augmenter l'effectif à 225 clones. L'interaction a été testée en utilisant le modèle d'ANOVA suivant :

$$Y = \mu + clone + site + clone * site$$

$N_J$  et S présentent une interaction *clone\*site* significative entre les deux dispositifs 2 et 3. Cependant, les effets d'interaction ne modifient pas profondément le rang des clones entre les différents tests compte tenu des corrélations significatives obtenues entre les moyennes de clones (Tableau 20). Ainsi pour la détection de QTL l'analyse sera réalisée d'une part, au niveau de chaque dispositif puis, en cumulant l'information des dispositifs 2 et 3.

Les corrélations sont fortement significatives pour un même caractère dans différents tests, variant entre 0.41 et 0.61 pour  $N_J$  et entre 0.35 et 0.57 pour S. Les moyennes de clones du test 2 sont moins corrélées avec les valeurs du test 1 (0.41 pour  $N_J$ , et 0.36 pour S) qu'avec celles du test 3 (0.42 pour

**Tableau 21** : Détection de QTL impliqués dans la phénologie du débourrement

Caractère	GL	N	Lod	$\alpha_{chr}$	$\alpha_G$	P1	P2	IC	D	P	$V_D=0$		$V_D=1/2 V_A$	
											S	PV	S	PV
NJ1	2f	191	3.6	0.001	5%	16.7	16.7	18.5	2.4	5.8				
NJ1	4f	189	3.3	0.002	5%	24.5	34.5	51.2	2.2	5.3				
NJ1	5m	196	1.8	0.044	50%	22.4	39	65.5	-1.8	3.6				
NJ1	6m	216	4.8	0.001	5%	51.5	49.1	15.0	-3.3	10.7				
NJ1	7f	162	4.4	0.001	5%	72.5	69.8	27.6	2.6	7.1				
NJ1	9f	207	3.3	0.001	5%	48.1	33.7	36.3	2.2	5.0				
NJ1	10f	220	6.1	0.001	5%	19.4	24.3	31.9	-3.0	8.7				
NJ1	10m	192	1.8	0.027	30%	13.8	24.8	34.3	-1.9	3.6				
NJ1	11f	207	4.2	0.001	5%	72.5	64.5	31.6	2.6	6.8				
NJ2	2f	148	3.6	0.001	5%	50.8	48	20.0	1.0	9.5	9.6	2.4	16.7	2.5
NJ2	3m	157	2.0	0.013	20%	22.1	21.3	10.9	0.8	5.7	5.7	1.4	10.0	1.5
NJ2	4f	156	2.5	0.010	10%	71.2	68.4	35.9	0.9	5.8	5.8	1.5	10.2	1.5
NJ2	7f	98	2.1	0.023	20%	79.8	64.2	50.2	1.2	8.2	8.3	2.1	14.4	2.2
NJ2	9f	154	1.7	0.042	50%	51.8	34.5	41.6	0.8	4.3	4.3	1.1	7.6	1.1
NJ2	9m	129	2.2	0.010	15%	0	13.7	47.2	1.0	6.8	6.8	1.7	12.0	1.8
NJ2	10m	89	2.0	0.016	20%	0	4.1	24.1	1.1	9.7	9.8	2.5	17.1	2.6
NJ2	11f	155	1.8	0.032	30%	17.7	24.3	36.8	0.8	4.0	4.0	1.0	7.0	1.1
NJ3	1m	156	8.3	0.001	5%	0	1.4	5.8	1.8	10.1	6.0	3.1	10.5	3.4
NJ3	2f	181	6.0	0.001	5%	43.7	42.8	20.6	1.9	10.0	5.9	3.1	10.4	3.3
NJ3	2m	154	2.4	0.011	10%	76.9	60.3	40.4	-1.0	3.2	1.9	1.0	3.3	1.1
NJ3	3m	159	1.9	0.018	30%	40.1	35.8	25.7	-1.0	3.1	1.8	1.0	3.2	1.0
NJ3	4m	168	5.7	0.001	5%	38.4	34.4	28.0	1.7	8.7	5.2	2.7	9.1	2.9
NJ3	4f	185	2.0	0.037	20%	27.7	36.7	51.0	1.2	3.9	2.3	1.2	4.1	1.3
NJ3	5m	174	4.0	0.001	5%	13.5	14.9	20.2	-1.3	4.8	2.9	1.5	5.0	1.6
NJ3	6m	192	5.0	0.001	5%	50.2	49.3	11.9	-1.3	5.2	3.1	1.6	5.4	1.7
NJ3	7f	122	3.5	0.001	5%	81.9	73.2	32.3	1.8	10.2	6.1	3.1	10.6	3.4
NJ3	9f	122	2.2	0.015	15%	15.9	20.6	28.8	1.5	6.1	3.6	1.9	6.3	2.0
NJ3	9m	185	5.0	0.001	5%	33.1	30.9	25.9	1.4	5.7	3.4	1.8	5.9	1.9
NJ3	10f	193	4.1	0.001	5%	27.7	22.9	20.2	-1.6	7.2	4.3	2.2	7.5	2.4
NJ3	10m	172	6.4	0.001	5%	34.8	28.4	19.0	-1.5	6.2	3.7	1.9	6.5	2.1
NJ3	11m	178	8.9	0.001	5%	47.7	49.4	16.2	1.9	9.6	5.7	3.0	10.0	3.2
NJ3	12f	179	2.7	0.003	5%	25.2	26.8	13.2	-1.4	5.3	3.2	1.6	5.5	1.8
NJ2+3	1m	172	6.2	0.001	5%	7.1	7.83	12.3	1.73	13.9	8.4	3.4	14.8	3.7
NJ2+3	2f	196	4.9	0.001	5%	43.6	36.2	31.0	1.33	7.8	4.7	1.9	8.3	2.0
NJ2+3	4m	200	3.8	0.001	5%	8.7	27.4	40.6	1.04	4.8	2.9	1.2	5.1	1.3
NJ2+3	5m	189	3.4	0.002	5%	18.3	20.9	32.3	-1.07	5.0	3.0	1.2	5.3	1.3
NJ2+3	6m	209	3.3	0.002	5%	55.9	52.9	16.4	-0.98	4.2	2.5	1.0	4.5	1.1
NJ2+3	7f	132	4.4	0.001	5%	79.8	73.9	27.6	1.58	10.8	6.5	2.7	11.5	2.8
NJ2+3	8f	163	2.011	0.009	15%	27.3	18.8	16.3	-0.94	3.8	2.3	0.9	4	1
NJ2+3	9f	132	1.7	0.039	30%	9.7	28.7	29.4	0.92	3.7	2.2	0.9	3.9	1
NJ2+3	9m	201	4.4	0.001	5%	27.9	22.1	24.3	1.35	8.0	4.9	2.0	8.5	2.1
NJ2+3	10f	213	5.4	0.001	5%	24.8	20.5	15.1	-1.40	8.5	5.2	2.1	9.0	2.2
NJ2+3	10m	186	3.3	0.002	5%	34.8	28.7	21.4	-1.04	4.5	2.7	1.1	4.8	1.2
NJ2+3	11f	210	2.6	0.004	5%	63	58.5	28.4	0.96	3.9	2.4	1.0	4.1	1.0
NJ2+3	11m	207	5.1	0.001	5%	65.5	53.9	26.3	1.20	6.1	4	1.5	6.4	1.6
Moyenne										6.5	4.5	1.8	7.9	1.9

GL : le groupe de liaison (m = carte mâle et f = carte femelle)

N : le nombre d'individus pris en compte dans l'analyse

Lod : la valeur de Lod maximale obtenue sur le chromosome après 1000 permutations

$\alpha_{chr}$  : L'erreur de type I au niveau du chromosome obtenue après 1000 permutations

$\alpha_G$  : L'erreur de type I en pourcentage au niveau du génome

P1 : La position la plus probable du QTL sur le chromosome d'après la valeur maximale de LOD

P2 : La position du QTL la plus probable après 1000 ré-échantillonnages réalisés par bootstrap

IC : intervalle de confiance du QTL obtenu après 1000 ré-échantillonnages réalisés par bootstrap

D : effet de substitution allélique au QTL

P : le pourcentage de variance des moyennes de clones expliquée par le QTL dans la descendance

S : le pourcentage de variance additive expliquée par le QTL dans la population d'où sont issus les parents

PV : le pourcentage de variance phénotypique expliquée par le QTL dans la population d'où sont issus les parents

$N_j$  et 0.46 pour S.). Cependant, les corrélations sont fortes entre les dispositifs 1 et 3 (0.61 pour  $N_j$  et 0.57 pour S) alors que ces deux dispositifs sont localisés dans des sites écologiques différents séparés de plus de 150 km, comparées aux tests 2 et 3 qui sont adjacents.

#### C.2.2.1.3.3 *Corrélations entre caractères*

A l'intérieur d'un dispositif, les deux mesures  $N_j$  et S du débourrement sont toujours fortement corrélées variant entre  $-0.84$  et  $-0.98$  (Tableau 20).

#### C.2.2.1.3.4 *Détection de QTL*

Tous les QTL détectés pour le stade de débourrement (S) l'étaient aussi pour le nombre de jours ( $N_j$ ), ainsi pour la clarté des résultats nous avons seulement présenté les résultats pour le nombre de jours ( $N_j$ ) (Tableau 21).

Si l'on considère l'analyse réalisée sur les 3 dispositifs pris séparément, 32 QTL (entre 8 et 13 par dispositif) impliqués dans le débourrement ont été détectés en se fixant un seuil de 5% au niveau du chromosome dont 19 sont significatifs à 5% au niveau du génome.

Le pourcentage de moyenne clonale ( $P$ ) expliqué par les QTL dans la descendance est relativement faible (entre 3.1 et 10.7%), le pourcentage de variance additive expliquée (S) par le QTL dans la population d'où sont issus les parents est faible à modéré (entre 1.8 et 17.1%) et le pourcentage de variance phénotypique (PV) expliqué par les QTL dans la population d'où sont issus les parents est faible (entre 1 et 3.4%).

Lorsque la détection de QTL est réalisée sur les moyennes des clones des dispositifs 2 et 3, 13 QTL significatifs à 5% au niveau du chromosome sont obtenus dont 11 sont significatifs à 5% au niveau du génome. On ne retrouve pas le QTL détecté sur le chromosome 3m par contre un QTL supplémentaire est détecté sur le chromosome 8 avec cependant une erreur de type I au niveau de génome égale à 30%.

**Tableau 22** : Erreur de type I des colocalisations entre les QTL impliqués dans le débourement détectés dans plusieurs dispositifs.

*a : Erreur de type I des colocalisations entre les QTL détectés dans 2 dispositifs*

Carte	Dispositifs	GL	N	P	G	C	Pvalue
Femelle	1_2	7 ; 9	135	5	6	2	<b>0.015</b>
Femelle	1_3	4 ; 7 ; 9	135	6	6	3	<b>0.001</b>
Mâle	1_3	6	139	3	9	1	0.172
Femelle	2_3	2 ; 7	135	5	6	2	<b>0.015</b>

*b : Erreur de type I des colocalisations entre les QTL détectés dans le dispositif 1 et les QTL détectés sur les moyennes de clones des dispositifs 2 et 3.*

Carte	Dispositifs	GL	N	P	G	C	Pvalue
Femelle	1_2+3	7 ; 10 ; 11	135	6	6	3	<b>0.001</b>
Mâle	1_2+3	6 ; 10	139	3	7	2	<b>0.006</b>

Avec

GL : les groupes de liaison où une colocalisation est observée

N : le nombre d'intervalle comparés. Correspond au rapport entre la taille de la carte en cM et l'intervalle moyen de la position la plus probable du QTL (moyenne de  $|P1-P2|$  dans le tableau 21).

P : le nombre de QTL détecté dans le dispositif qui donne le plus petit nombre de QTL

G : le nombre de QTL détecté dans le dispositif qui donne le plus grand nombre de QTL

C : le nombre de colocalisations observé entre deux dispositifs

Pvalue : Probabilité d'observer C colocalisations par chance entre deux dispositifs

$$Pvalue = \frac{C_G^C \cdot C_{N-G}^{P-C}}{C_N^P}$$

Etant donné qu'un grand nombre de QTL a été détecté pour le débourrement nous avons testé l'épistasie par ANOVA à deux facteurs en utilisant le logiciel QGene. Sur les 5400 tests réalisés, aucun n'était significatif au seuil de 1% au niveau du génome, indiquant ainsi l'absence d'interaction entre allèles de locus différents.

#### *C.2.2.1.3.4.1 Signification statistique des colocalisations entre les QTL détectés dans deux dispositifs*

Pour tester si le chevauchement des QTL détectés dans les différents dispositifs est dû au hasard ou pas, nous avons calculé la probabilité de co-occurrence des QTL dans les mêmes régions génomiques sous l'hypothèse nulle de distribution aléatoire (voir C.1.3.3).

La distance séparant les deux estimations de la localisation du QTL (celle qui correspond au Lod score maximum et celle qui correspond à la moyenne des 1000 ré-échantillonnages bootstraps) est en moyenne de 6.7 cM. Ainsi dans la carte femelle 135 intervalles (902cM/6.7) sont comparés et 139 (933cM/6.7) chez le mâle.

Si l'on considère l'analyse réalisée sur chaque dispositif, 8 groupes de liaisons présentent des QTL à la même position pour au moins deux dispositifs (Figure 17). Les colocalisations entre les dispositifs 1 et 3 étant les plus fréquentes (3) comme le suggéraient les meilleures corrélations entre moyenne de clones observées entre ces deux dispositifs (Tableau 20). Les colocalisations observées sur la carte femelle sont toutes significatives (Tableau 22) alors que l'unique colocalisation observée sur le groupe 6m entre le dispositif 1 et 3 ne l'est pas. Si l'on considère maintenant les colocalisations entre le dispositif 1 et les dispositifs 2 et 3 cumulés, l'ensemble des colocalisations observées sur les deux cartes mâle et femelle sont significatives. Ces résultats confirment la stabilité des QTL dans l'espace.

#### C.2.2.1.4 Conclusion

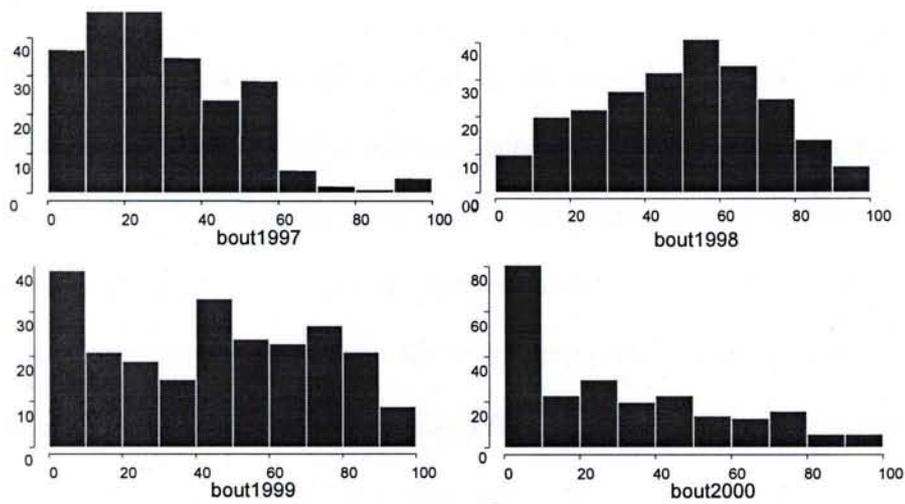
La phénologie du débourrement chez le chêne est modérément héritable (entre 0.15 et 0.52) comparé aux très fortes héritabilités obtenues chez le peuplier qui atteignent plus de 90% (Bradshaw *et al.*, 1995 ; Frewen *et al.*, 2000). Les pourcentages de moyenne de clones expliqués par les QTL sont en moyenne de 6.5% variant entre 3.1 et 13.8%. L'architecture génétique du débourrement implique au moins 8 QTL uniques répartis sur 8 chromosomes, qui sont stables entre les dispositifs, associés à au moins 18 QTL "dispositifs spécifiques". Les colocalisations observées sur les groupes de liaison 9m et 9f sont renforcées par la présence sur ce même groupe de liaison d'un QTL impliqué dans le débourrement chez le châtaignier (Manuela Casasoli, communication personnelle).

### C.2.2.2 QTL impliqués dans l'aptitude au bouturage

La multiplication végétative des plantes est possible grâce à la totipotence de la cellule végétale, c'est à dire la possibilité qu'a potentiellement n'importe quelle cellule végétale de se différencier pour se redifférencier ensuite et donner un nouvel organisme. Ainsi il est possible de reconstruire une plante entière à partir d'un morceau de tige en le stimulant avec des hormones de rhizogénèse qui favorisent la production de racines adventives. La multiplication clonale des individus a plusieurs avantages. Un premier avantage, exploité dans la détection de QTL, est la possibilité de tester dans différents environnements, la performance des individus et d'en estimer la variance génétique. Un deuxième avantage est qu'elle permet d'accélérer les programmes d'amélioration en "immortalisant" la supériorité génétique des individus sélectionnés, et accélérer ainsi les sorties variétales des espèces économiquement intéressantes. En horticulture, la propagation végétative de plantes au phénotype désiré est largement employée alors que chez les arbres forestiers son utilisation dans les programmes de plantation reste limitée au peuplier, au saule, à l'eucalyptus, au merisier, à l'épicéa et au cryptomeria. Une des raisons majeures est que le bouturage d'arbres physiologiquement matures est difficile et parfois impossible chez certaines espèces comme le pin maritime ou le chêne.

L'aptitude au bouturage a largement été étudiée chez le peuplier et l'eucalyptus dans des perspectives de sélection précoce assistée par marqueur. C'est un caractère qui présente des fortes héritabilités (0.52 chez l'eucalyptus, Grattapaglia *et al.*, 1995 ; 0.72 chez le peuplier, Han *et al.*, 1995) et qui semble être contrôlé par un petit nombre de gènes : chez *Eucalyptus grandis* un seul QTL a été détecté qui explique 8% de la variance phénotypique, et chez *E. urophylla* 3 QTL expliquent au total 28% de la variance phénotypique (Grattapaglia *et al.*, 1995). Similairement chez le peuplier, un seul QTL a été détecté qui explique 37% de la variance phénotypique du taux de rhizogénèse in vitro des tiges. Une étude récente sur l'eucalyptus réalisée durant deux années consécutives indique que la plupart des QTL sont stables entre les années (Marques *et al.*, 1999).

**Figure 18** : Histogramme de distribution de l'aptitude au bouturage.



**Tableau 23** : Moyennes et coefficients de variation phénotypique de l'aptitude au bouturage chez le chêne pédonculé selon les années de mesures.

	1997	1998	1999	2000
Moyenne	30.41	51.73	45.78	30.06
Ecart type	20.61	22.94	28.43	27.51
CV <sub>P</sub>	0.68	0.446	0.62	0.92

**Tableau 24**: Corrélations de Pearson entre 4 années de mesures de l'aptitude au bouturage

	1997	1998	1999
1998	<b>0.19</b>		
1999	<b>0.14</b>	<b>0.36</b>	
2000	0.099	<b>0.37</b>	<b>0.58</b>

En gras les corrélations sont significatives

#### C.2.2.2.1 Matériel végétal

La technique de bouturage chez le chêne est présentée dans le chapitre C.1.2.3.1.1. ainsi que dans l'article en l'annexe 1. Depuis 1997, chaque année les pieds mères des descendants de la population de cartographie sont bouturés en moyenne 20 fois en vue d'analyse QTL comparative.

Dans notre étude, l'aptitude au bouturage qui correspond au pourcentage de boutures viables (bout) a été mesurée sur 232 descendants de la population de cartographie durant 4 années consécutives.

#### C.2.2.2.2 Résultats

L'analyse préliminaire des données indique un écart à la normalité de la distribution de l'aptitude au bouturage mesuré en 1997, 1999 et 2000 (Figure 18) . Une transformation  $Y' = \sqrt{Y}$  a été réalisée pour la détection de QTL.

En moyenne sur les 4 années, le taux de réussite du bouturage est de 40 % et le coefficient de variation phénotypique est de 0.67 (Tableau 23). En cumulant les 4 années de mesure, nous avons testé l'effet année et l'effet clone par ANOVA qui se sont révélés être hautement significatifs. L'absence de répétition des mesures à l'intérieur de chaque année n'a pas permis de mesurer les interactions entre clones et années. Cependant les corrélations phénotypiques entre les années de mesures étant fortement hétérogènes (Tableau 24), l'hypothèse d'interactions ne doit pas être rejetée. L'année 1997 est par exemple faiblement corrélée à toutes les autres ( $r$  entre 0.09 et 0.19), alors qu'une forte corrélation est observée entre 1999 et 2000 ( $r=0.58$ ). Nous avons donc préféré réaliser la détection de QTL pour chaque année plutôt que de cumuler les 4 années de mesure. L'absence de répétition ne permettant pas l'estimation de la variance génétique additive et de l'héritabilité, l'effet des QTL correspond à la part de variance phénotypique expliquée dans la descendance.

18 QTL ont été détectés au seuil de 5% au niveau du chromosome dont 10 sont significatifs à 5% au niveau du génome qui expliquent entre 3.6 et 13.8 % de la variance phénotypique (Tableau 25). Parmi les 18 QTL détectés, 9 QTL sont année-spécifiques. Des colocalisations sont observées uniquement

**Tableau 25** : Résultat de la détection de QTL impliqués dans l'aptitude au bouturage au cours de 4 années de mesure : 1997, 1998, 1999, 2000.

caractère	GL	N	Lod	$\alpha_{chr}$	$\alpha_G$	D	PEV	P <sub>1</sub>	P <sub>2</sub>	IC
bout1997	8m	175	1.6	0.04	50	-0.9	4	60.6	64.4	35.9
bout1998	7f	138	3.0	0.002	5	-15.4	10.6	87.6	84.9	24.1
bout1998	1m	148	2.7	0.002	5	11.5	6.4	61.6	58.6	15.4
bout1998	3m	213	2.7	0.002	5	9.6	4.3	22.1	23.8	16.8
bout1998	4m	149	1.8	0.032	50	-9.6	4.3	82.2	63.3	51.9
bout1998	8m	173	2.3	0.012	15	-10.7	4.9	6.7	13.8	36.8
bout1998	12m	196	4.1	0.001	5	14.8	9.9	55.6	54.9	15.9
bout1999	9f	206	1.7	0.04	30	1.0	3.6	48.1	35.7	42.1
bout1999	11f	206	2.2	0.01	15	1.2	4.6	75.5	66.0	44.9
bout1999	3m	177	1.5	0.044	50	1.1	3.8	33.8	25.7	25.3
bout1999	6m	199	5.0	0.001	5	-1.6	8.9	14.8	18.0	20.0
bout1999	8m	179	4.8	0.001	5	-1.7	10.7	17.3	16.7	18.4
bout1999	12m	175	3.0	0.002	5	1.3	5.5	38.1	41.6	24.7
bout2000	6f	200	2.1	0.013	15	1.3	4.5	39.4	40.2	23.1
bout2000	9f	190	1.6	0.039	30	1.1	3.5	0.0	23.8	51.7
bout2000	6m	185	3.2	0.001	5	-1.6	6.6	9.8	21.5	32.9
bout2000	8m	173	6.2	0.001	5	-2.3	13.8	3.3	12.2	25.3
bout2000	12m	196	5.0	0.001	5	1.8	8.8	65.3	59.8	22.7
Moyenne							6.6			

GL : groupe de liaison (m = carte mâle et f = carte femelle)

N : nombre d'individus pris en compte dans l'analyse

Lod : valeur de Lod maximale obtenue sur le chromosome après 1000 permutations

$\alpha_{chr}$  : erreur de type I au niveau du chromosome obtenue après 1000 permutations

$\alpha_G$  : erreur de type I au niveau du génome en pourcentage

D : effet de substitution allélique au QTL

PEV : pourcentage de variance phénotypique expliquée par le QTL

P<sub>1</sub> : position la plus probable du QTL d'après la valeur maximale du LOD

P<sub>2</sub> : position du QTL la plus probable après 1000 ré-échantillonnage réalisés par bootstrap

IC : intervalle de confiance du QTL obtenu après 1000 ré-échantillonnage réalisés par bootstrap

Remarque : les données brutes de bout1997, bout1999 et bout2000 ont été transformées ( $Y' = \sqrt{Y}$ )

**Tableau 26** : Erreur de type I des colocalisations entre les QTL impliqués dans l'aptitude au bouturage détectés sur 2 années.

	Années	GL	N	P	G	C	Pvalue
Carte mâle	1998-1999	3 ; 8	133	4	5	2	0.006
Carte mâle	1998-2000	8 ; 12	133	3	5	2	0.003
Carte mâle	1999-2000	6 ; 8	133	3	4	2	0.002

GL : les groupes de liaison où une colocalisation est observée

N : le nombre d'intervalle comparés. Correspond au rapport entre la taille de la carte en cM et l'intervalle moyen de la position la plus probable du QTL (moyenne de  $|P_1 - P_2|$  dans le tableau 24)

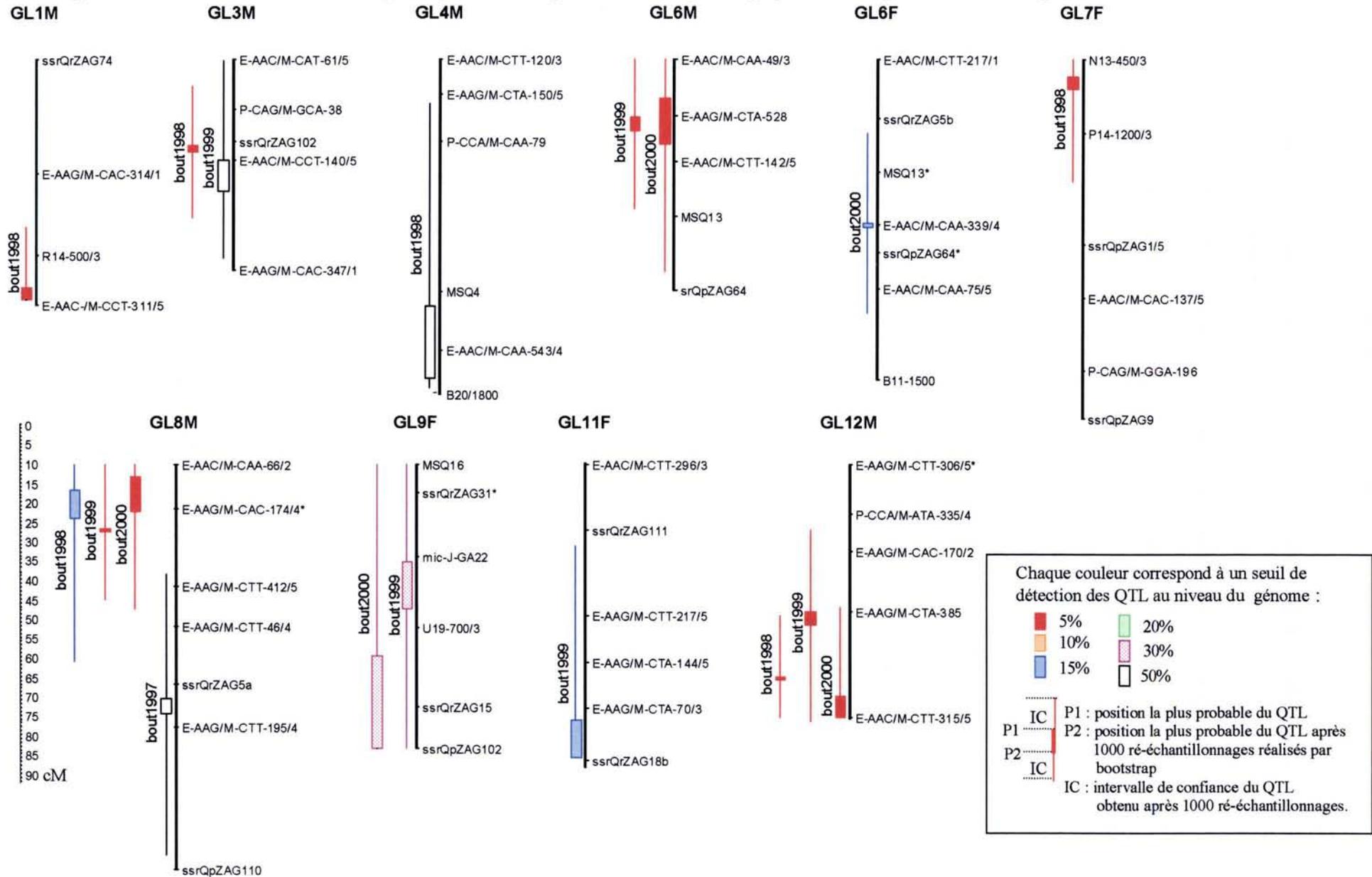
P : nombre de QTL l'année 2 (année qui donne le plus petit nombre de QTL)

G : nombre de QTL détecté l'année 1 (année qui donne le plus grand nombre de QTL)

C : nombre de colocalisations observées entre les années 1 et 2

Pvalue : probabilité d'observer C colocalisations par chance entre les années 1 et 2

**Figure 19 :** Distribution sur la carte génétique du chêne pédonculé des QTL impliqués dans l'aptitude au bouturage



**Tableau 27** : Gènes impliqués chez Arabidopsis dans les différentes étapes de la formation des racines latérales (modifié d'après Malamy et Benfey (1997)).

Événements morphogénétiques	Gènes impliqués chez Arabidopsis
Transport et perception des signaux auxiniques	<i>alf 4, aux 1, axr 1, axr 4</i>
Entrée en division des cellules en situation profonde à l'origine de la racine	<i>cyc, cdc2a</i>
Formation du primordium racinaire et spécification cellulaire	<i>scr</i>
Accroissement cellulaire et émergence du primordium	?
Production locale d'auxine	<i>alf3</i>
Activation du méristème	<i>rml 1, rml 2, cyc</i>

chez le mâle, sur les groupes de liaison 3m, 6m, 12m pour deux années de mesure, et sur le groupe 8m pour trois années (Figure 19). Le test de colocalisation par paire d'années indique qu'elles sont toutes significatives au seuil de 5% (Tableau 26). Ainsi, sur l'ensemble des QTL détectés, 9 QTL seraient année-spécifique et 4 seraient stables entre les années.

#### **C.2.2.2.3 Conclusion**

L'aptitude à l'enracinement chez le chêne est contrôlée par plusieurs QTL plus ou moins stables entre les années qui ont un effet faible à modéré. La production de racines adventives par les boutures ou tiges néoformées a été étudiée chez *Arabidopsis* (Malamy et Benfey, 1997). L'organogenèse des racines adventives implique plusieurs événements contrôlés génétiquement (voir Tableau 27). Il serait intéressant dans une étude future de rechercher des gènes candidats de l'aptitude à la formation des racines adventives chez le chêne en se basant sur les résultats obtenus chez *Arabidopsis* et regarder si ces gènes colocalisent avec les QTL. Par ailleurs, l'équipe d'Ecophysiologie de l'INRA de Nancy a entrepris une étude de détection de QTL impliqués dans la formation de racines adventives en condition de stress anoxique qui permettra de voir s'il existe des régions génomiques communes avec l'aptitude à l'enracinement.

### **C.2.3 COMPARAISON DES QTL DE CARACTERES DISCRIMINANTS ET NON DISCRIMINANTS.**

Dans la partie qui suit nous allons comparer l'effet des QTL ainsi que leur distribution entre les caractères discriminants et non discriminants. Les caractères discriminants et non discriminants ont été répartis de la façon suivante :

#### Caractères discriminants :

- Sept caractères morphologiques foliaires (PL, SW, NL, NV, BS, PU<sub>v</sub>, PUI),
- La discrimination isotopique du carbone ( $\delta^{13}\text{C}$ )

**Tableau 28 :** Nombre de QTL détectés par caractère en considérant séparément la carte mâle, la carte femelle et les différentes années de mesure (ou dispositifs).

Discriminant	Femelle	Mâle	Non discriminant	Femelle	Mâle
Ht	1	3	NJ_1	6	3
L <sub>UC</sub>	1	2	NJ_2	5	3
N <sub>UC</sub>	3	1	NJ_3	6	8
NL_1	1	1	NJ_2+3	6	7
PL_1	1	1	bout_1997	0	1
NV_1	2	0	bout_1998	1	5
SW_1	2	3	bout_1999	2	4
Pul_1	3	1	bout_2000	1	3
Puv_1	3	0	LL_1	2	0
BS_1	0	0	LW_1	1	0
NL_2	0	1	WP_1	1	1
PL_2	3	0	LL_2	1	1
NV_2	3	0	LW_2	1	0
SW_2	1	1	WP2_2	0	0
Pul_2	0	2			
PUv_2	3	1			
BS_2	2	3			
$\delta^{13}\text{C}_1$	1	3			
$\delta^{13}\text{C}_2$	1	3			
$\delta^{13}\text{C}_3$	2	5			
Somme	33	31		33	36
Moyenne	1.65	1.55		2.36	2.57

**Tableau 29 :** Comparaison entre le nombre moyen de QTL détecté pour les caractères discriminants (D) et le nombre moyen de QTL détecté pour les caractères non discriminants (ND).

	Femelle	Mâle
Moyenne D	1.65	1.55
Moyenne ND	2.36	2.57
Test Student (P.value)	-1.20 (0.24)	-1.46 (0.15)

- Trois caractères de croissance en hauteur ( $H_t$ ,  $L_{UC}$ ,  $N_{UC}$ ). L'accroissement annuel en hauteur (Crois) a été exclu car cette variable est dépendante de  $N_{UC}$  et  $L_{UC}$ .

Caractères non discriminants :

- Trois caractères de morphologie foliaire (LL, LW, WP),
- La phénologie du débourrement (deb)
- L'aptitude au bouturage (bout)

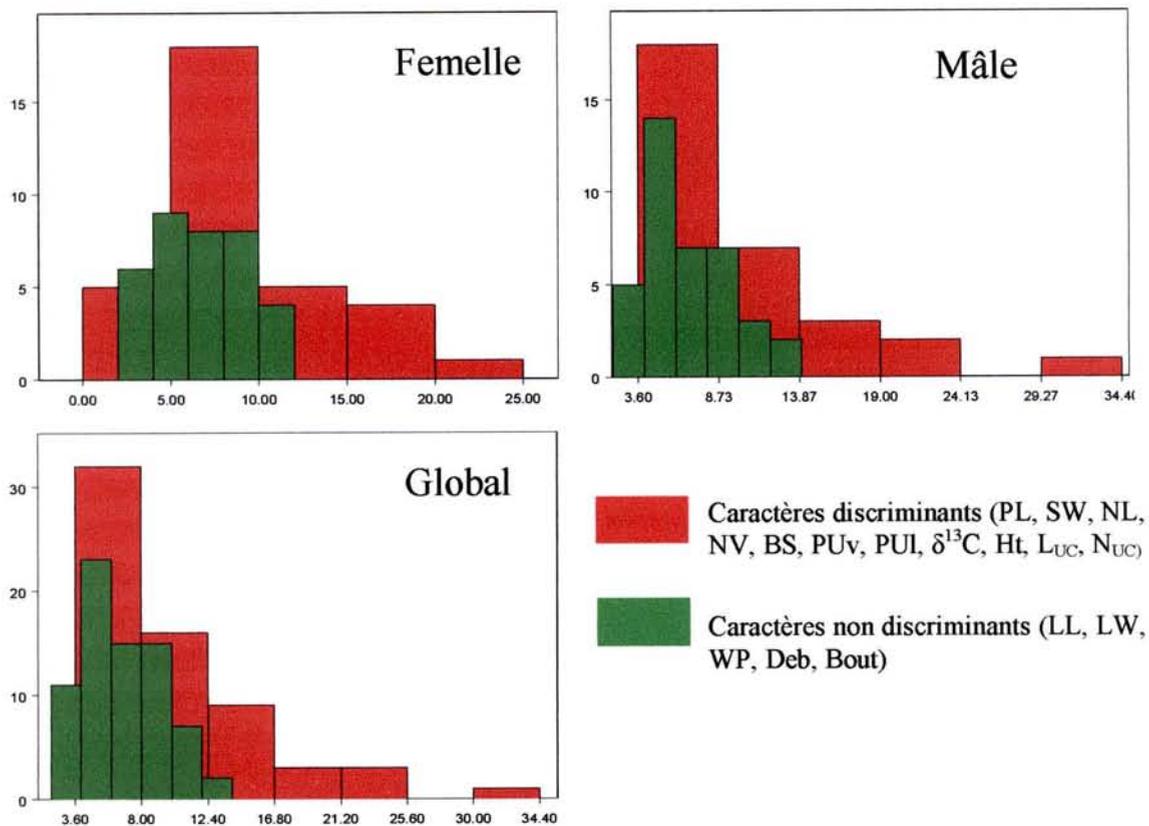
L'analyse porte sur le nombre total de QTL détectés avec un seuil de 5% au niveau du chromosome. Ce qui correspond respectivement chez la femelle, le mâle, et la carte consensus à 29, 27 et 50 QTL pour les caractères discriminants et 24, 27 et 40 QTL pour les caractères non discriminants. Ces deux catégories de caractères seront comparées pour la distribution spatiale des QTL vis à vis de l'hypothèse nulle  $H_0$  de distribution aléatoire des QTL. Nous avons donc omis dans chaque catégorie les caractères qui étaient " liés par construction" (par exemple Crois=  $N_{UC}$  et  $L_{UC}$ ) pour lesquels il y avait des raisons évidentes de regroupement dues à de la pléiotropie.

### **C.2.3.1 Nombre de QTL détectés et distribution de la variance expliquée par les QTL**

Si l'on considère séparément chaque caractère, chaque carte (mâle et femelle) et chaque année (ou dispositif de mesure) entre 0 et 8 QTL ont été détectés. Pour les caractères discriminants, le nombre de QTL détectés par caractère varie entre 0 et 5 avec en moyenne, respectivement chez la femelle et le mâle, 1.65 et 1.55 QTL détectés. Pour les caractères non discriminants, le nombre de QTL détectés par caractère varie entre 0 et 8 avec en moyenne, respectivement chez la femelle et le mâle, 2.36 et 2.57 QTL détectés (Tableau 28). La différence entre le nombre moyen de QTL détecté pour les caractères discriminants et le nombre moyen de QTL détectés pour les caractères non discriminants n'est cependant pas significatif (Tableau 29).

La distribution des pourcentages de variance expliquée (PVE) par les QTL est donnée figure 20.

**Figure 20** : Distribution des pourcentages de variance phénotypique expliquée par les QTL en fonction du pouvoir discriminant des caractères.



**Tableau 30** : Pourcentage de variance phénotypique expliquée par les QTL.

		Min	Max	Moyenne	Mediane
Femelle	D	3.7	22.9	9.09	7.90
	ND	3.5	10.8	6.66	6.30
Mâle	D	3.6	34.4	9.98	8.00
	ND	3.1	13.9	6.67	5.85
Global	D	3.6	34.4	9.52	7.95
	ND	3.1	13.9	6.67	6.10

N : Caractères discriminants (PL, SW, NL, NV, BS, PUv, PUI,  $\delta^{13}C$ , Ht, L<sub>UC</sub>, N<sub>UC</sub>)

ND : Caractères non discriminants (LL, LW, WP, Deb, Bout)

**Tableau 31** : Comparaison entre les caractères discriminants et non discriminants des moyennes et des médianes des pourcentages de variance phénotypique expliquée par les QTL.

Carte	t de Student	W de Wilcoxon
Femelle	2.57 (0.008)	1303 (0.24)
Mâle	2.72 (0.008)	1249 (0.33)
Global (mâle et femelle cumulés)	3.81 (0.0002)	5073 (0.87)

Entre parenthèse est indiquée la Pvalue.

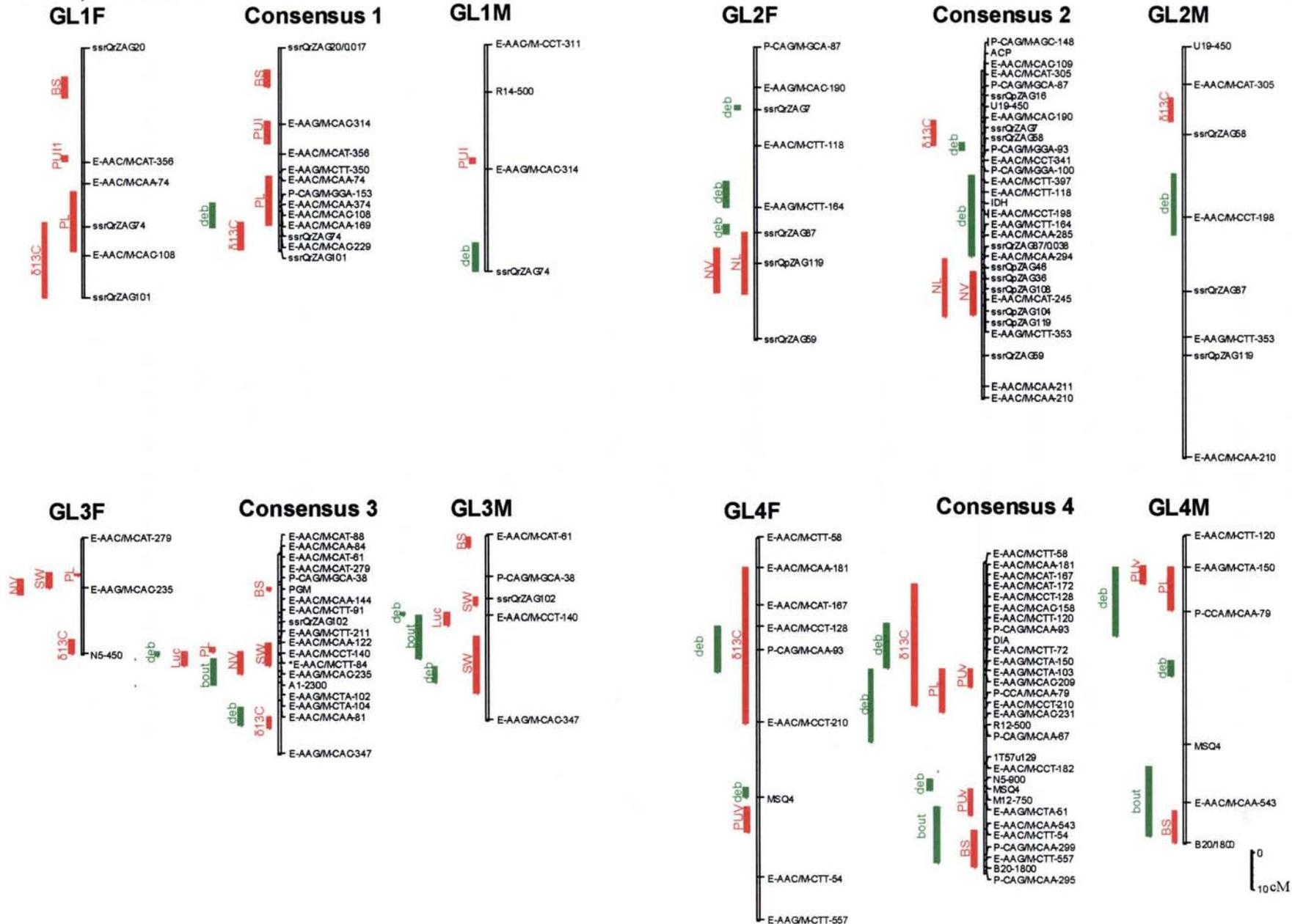
Si l'on considère les QTL détectés sur les deux cartes mâle et femelle, les PVE pour les caractères discriminants présentent une distribution en L avec plus de 30 QTL à effets faibles (inférieurs à 8%) et 11 QTL à effet forts (supérieurs à 15%) alors que les caractères non discriminants présentent une distribution qui tend vers la normalité avec 50 QTL à effet faibles et aucun QTL à effet fort. En moyenne les PVE des QTL discriminants sont donc supérieurs à ceux des QTL non discriminants (9.5 contre 6.8% si l'on considère les deux cartes mâle et femelle)(Tableau 30). Compte tenu de l'écart à la normalité des PVE pour les QTL détectés chez le mâle et globalement (cartes mâle et femelle) il est difficile de tester les différences entre les moyennes par un test t de Student. Les résultats sont tout de même fournis dans le tableau 31 accompagnés des résultats du test de Wilcoxon basé sur l'hypothèse nulle d'égalité des médianes. Les tests de Student sont tous significatifs alors que les tests de Wilcoxon ne le sont pas. Cependant un test de Wilcoxon non significatif ne peut pas se traduire par l'absence de différence entre les moyennes.

### **C.2.3.2 Répartition spatiale dans le génome**

La distribution sur le génome des QTL impliqués dans des caractères discriminants et non discriminants a été comparée à l'intérieur de chaque carte parentale (mâle et femelle) ainsi qu'en considérant l'ensemble des QTL détectés en les repositionnant sur une carte consensus qui sera décrite dans le prochain chapitre. La localisation des QTL sur les trois cartes est présentée dans la figure 21.

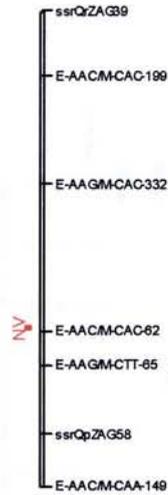
En moyenne l'intervalle de la position du QTL est de 4.8 cM pour les 3 cartes (femelle, mâle et consensus). Nous avons donc réalisé les tests de colocalisation en divisant le génome en intervalle de 5cM d'une part et en intervalles 10 cM d'autre part. Pour l'ensemble des 3 cartes tous les QTL significatifs au seuil de 5% au niveau du chromosome ont été considérés. Cependant, lorsque plusieurs QTL impliqués dans le même caractère se situaient dans le même intervalle un seul QTL a été considéré. La distribution des QTL le long du génome a été testée en considérant l'hypothèse nulle de distribution aléatoire des QTL selon la loi de Poisson décrite dans le chapitre C.1.3.1. Les résultats

**Figure 21** : Distribution des QTL impliqués dans des caractères discriminants (rouge) et non discriminants (vert) sur les cartes génétiques femelle, mâle et consensus.

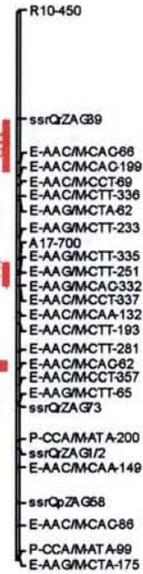


Suite figure 21 :

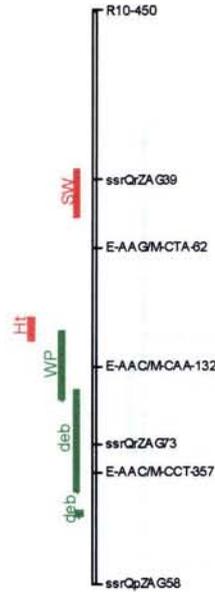
**GL5F**



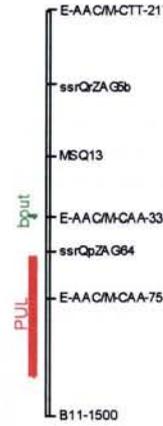
**Consensus 5**



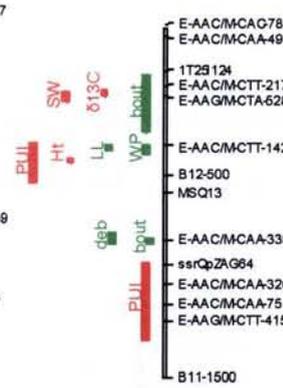
**GL5M**



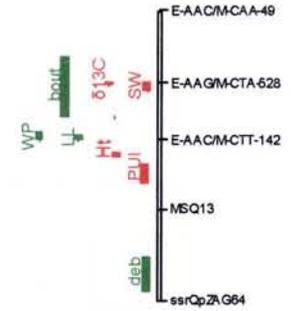
**GL6F**



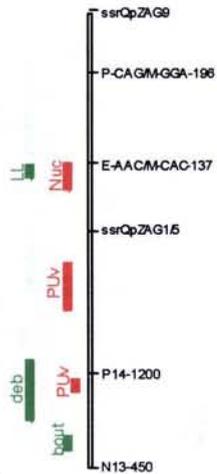
**Consensus 6**



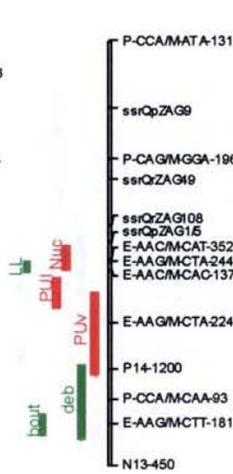
**GL6M**



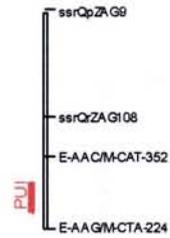
**GL7F**



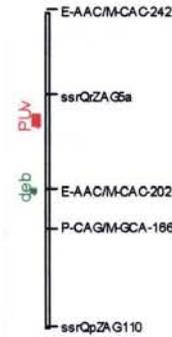
**Consensus 7**



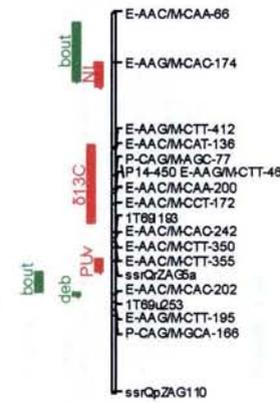
**GL7M**



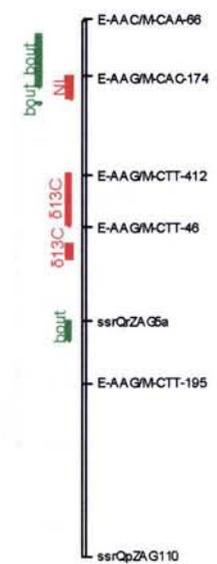
**GL8F**



**Consensus 8**

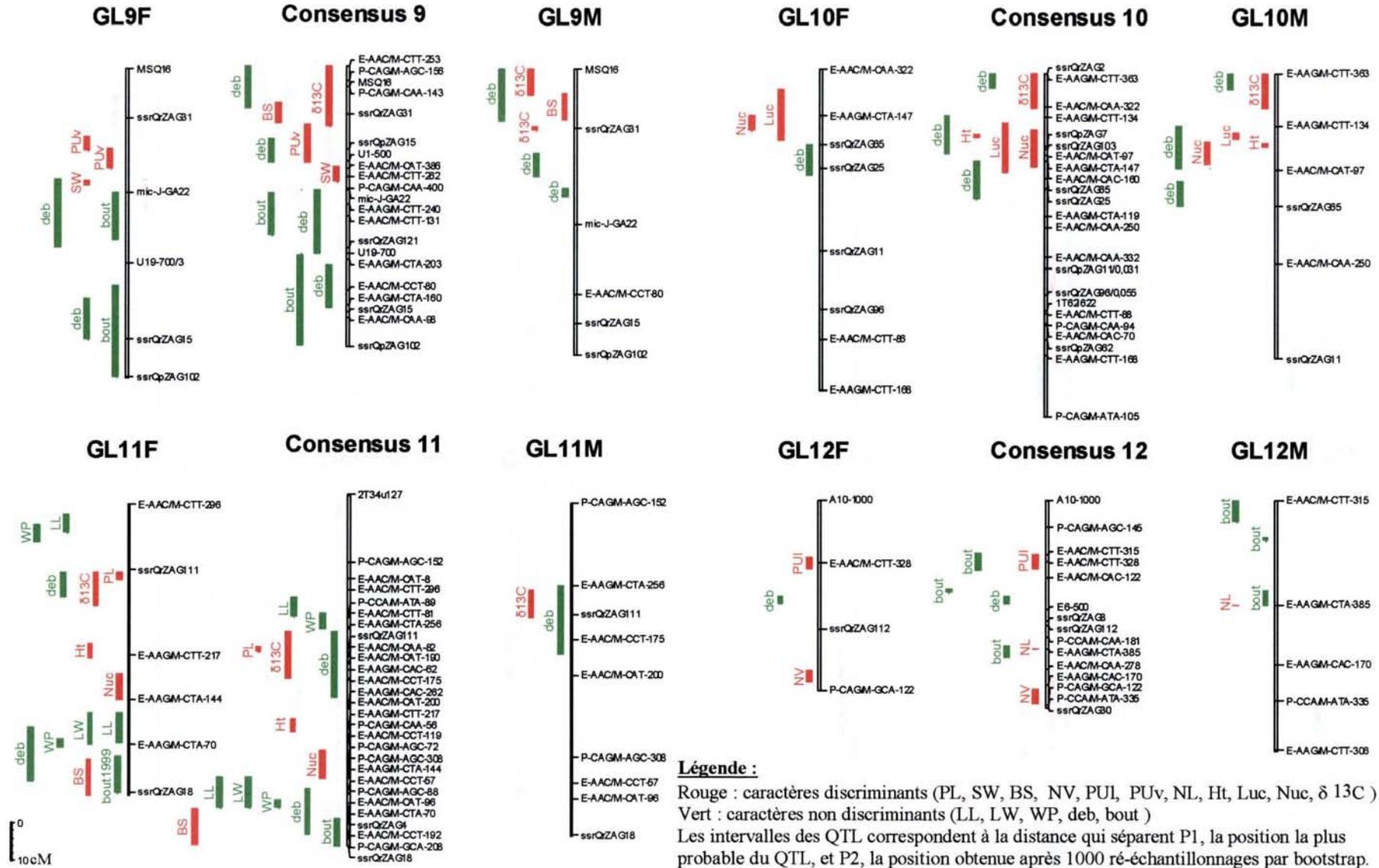


**GL8M**



0  
10cM

Suite figure 21 :



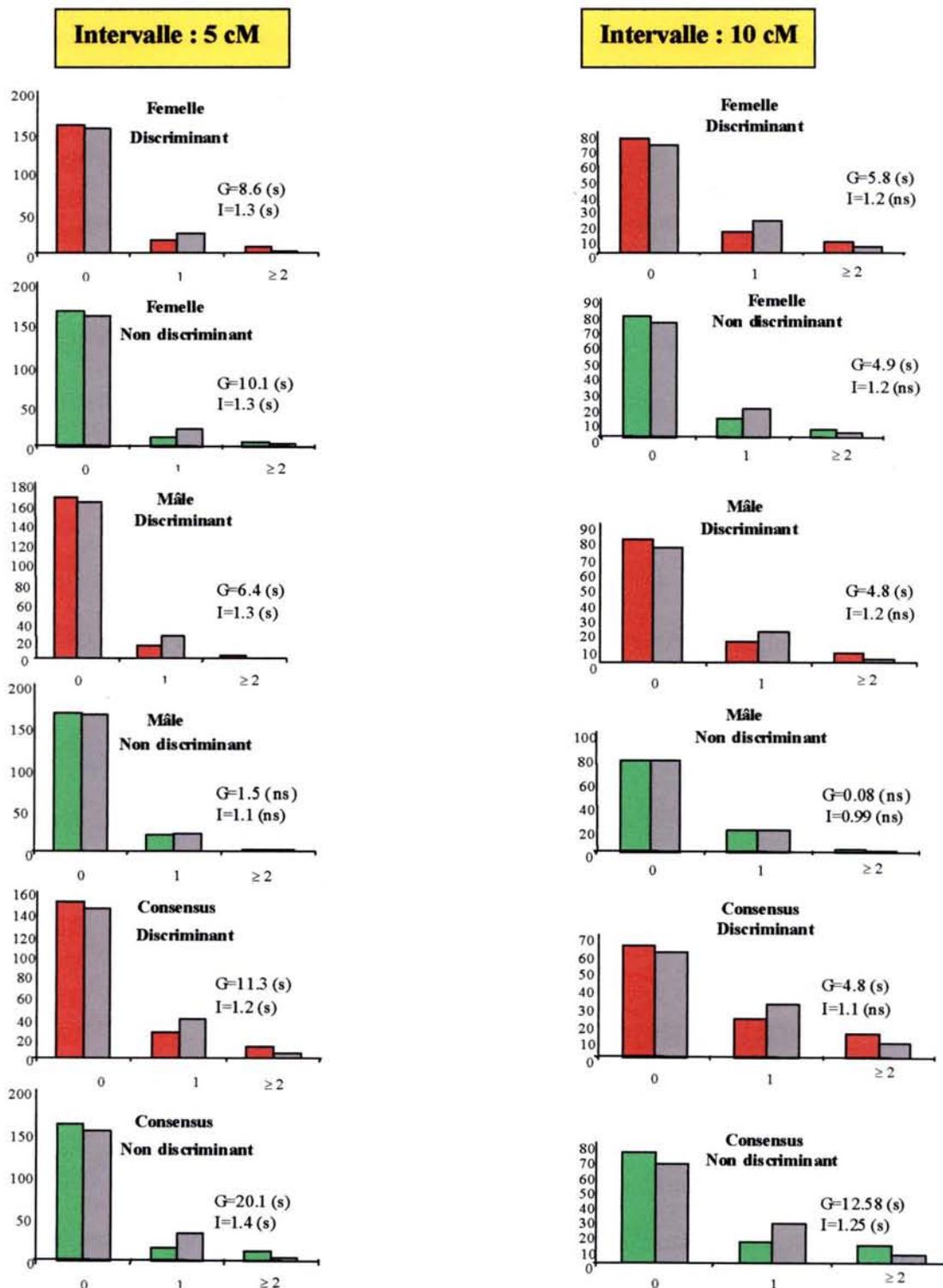
du test sont donnés en figure 22. On peut voir que sur les 3 cartes et en considérant les deux classes d'intervalle (5 et 10 cM), le nombre d'intervalles qui présentent une colocalisation entre au moins 2 QTL impliqués dans des caractères discriminants est supérieur à celui attendu sous l'hypothèse nulle de distribution aléatoire des QTL. Pour les caractères non discriminants le test est aussi significatif chez la femelle et la carte consensus pour les deux classes d'intervalle mais non significatif chez le mâle. L'indice de dispersion  $I$  est significativement supérieur à 1 pour l'ensemble des tests réalisés en considérant des intervalles de 5 cM (excepté pour les caractères non discriminants chez le mâle) ce qui traduit une distribution agrégée des QTL. Par contre, les indices de dispersion ne sont plus significatifs si l'on considère des intervalles de 10 cM (excepté pour les caractères non discriminants sur la carte consensus) ce qui traduit une distribution aléatoire des QTL.

Les colocalisations entre les QTL discriminants impliquent principalement des caractères morphologiques foliaires (NV et NL sur GL2, NV, PL et SW sur GL3, PL, PUv sur GL4 et PUv, PUI sur LG7), des caractères de croissance en hauteur (NUC, LUC, Ht sur LG10) mais aussi des caractères morphologiques foliaires et l'efficacité d'utilisation de l'eau (PL, PUv et  $\delta^{13}\text{C}$  sur LG4, BS,  $\delta^{13}\text{C}$  sur LG9 et PL,  $\delta^{13}\text{C}$  sur LG11). Les corrélations de Pearson entre  $\delta^{13}\text{C}$  et ces caractères foliaires ont été par ailleurs calculées et ne sont significatives qu'entre PUv mesurée dans le dispositif 2 et le  $\delta^{13}\text{C}$  mesuré en 2001 ( $r=-0.22$ )(résultats non présentés).

#### C.2.4 CONCLUSION GENERALE

L'approche QTL a permis de localiser sur le génome du chêne des régions impliquées dans la différenciation interspécifique. La répétition des mesures sur plusieurs années et dans différents dispositifs a permis de montrer une grande stabilité de l'expression de certains gènes dans l'espace et le temps mais aussi entre différents genres de la même famille puisque la localisation de certains QTL est confirmée chez le châtaignier (Manuela Casasoli, communication personnelle). La comparaison entre les caractères discriminants et non discriminants du nombre et de l'effet des QTL indique que le nombre moyen de QTL détectés pour les caractères discriminants est inférieur avec des PEV plus

**Figure 22** : Test de la structuration le long du génome de la distribution des QTL impliqués dans les caractères discriminants (rouge) et non discriminants (vert) sous l'hypothèse de distribution totalement aléatoire (processus poissonien) (gris).



**Légende :**

G : valeur de la statistique G, I : indice de dispersion. Entre parenthèses est indiqué si le test est significatif (s) ou pas (ns).

Vert : Nombre d'intervalles où 0, 1 ou au moins 2 QTL ont été détectés pour des caractères non discriminants (LL, LW, WP, deb, bout)

Rouge : Nombre d'intervalles où 0, 1 ou au moins 2 QTL ont été détectés pour des caractères discriminants (PL, SW, BS, NV, PUI, PU-NL, Ht, L<sub>UC</sub>, N<sub>UC</sub>, δ<sup>13</sup>C)

Gris : Nombre d'intervalles attendus sous l'hypothèse nulle de distribution aléatoire des QTL

fortes. Cependant, ces différences ne sont pas significatives. La répartition des QTL sur le génome ne semble pas totalement aléatoire. Les caractères discriminants tendent à se regrouper entre eux sur les cartes mâle et femelle lorsque l'on considère des intervalles de 5 cM, par contre, ils ne sont plus significatifs si l'on considère des intervalles de 10 cM.

Si la détection de QTL a permis de mettre en évidence des régions fortement impliquées dans la différenciation il ne faut cependant pas perdre de vue que cette stratégie ne permet pas une recherche exhaustive de régions discriminantes.

Tout d'abord l'étude est limitée par des contraintes expérimentales. Notre population de cartographie est une famille intraspécifique de plein frères qui réduit les chances de détecter des QTL informatifs. En effet, certains locus fortement discriminants pourraient être fixés à l'intérieur des deux espèces au niveau d'allèles différents et ne pas ségréger dans la descendance intraspécifique. La population idéale pour détecter ces QTL serait des familles interspécifiques  $F_2$  ou backcross qui ne sont pas disponibles à l'heure actuelle compte tenu de la maturité sexuelle tardive des chênes. Nous disposons néanmoins d'une famille interspécifique  $F_1$  de plein frère qui pourra faire l'objet d'une étude de QTL lorsque son effectif sera suffisamment grand. Deuxièmement, des loci peuvent ne pas être détectés car non polymorphes entre les deux parents de la famille. Pour le vérifier, il faudrait faire la détection de QTL dans des familles issues d'autres parents. Enfin, il est très probable que certains QTL, notamment ceux qui présentent les plus faibles effets, ne soient pas détectés en raison de l'effectif trop faible de la population. Une autre conséquence du faible effectif est une sur-estimation de l'effet des QTL détectés (Beavis, 1995).

## Chapitre D

# DISTRIBUTION DES INDICES DE DIFFERENCIATION GENETIQUE SUR LE GENOME

## **CHAPITRE D DISTRIBUTION DES INDICES DE DIFFERENCIATION GENETIQUE SUR LE GENOME**

Au cours du chapitre précédent, la stratégie suivie pour identifier des régions impliquées dans la différenciation génétique entre le chêne sessile et le chêne pédonculé était la recherche de relations entre la variabilité des caractères impliqués dans la différenciation phénotypique entre les deux espèces et le polymorphisme de marqueurs neutres à l'intérieur d'une population de cartographie intraspécifique de chêne pédonculé. Dans ce chapitre, la stratégie est différente mais complémentaire puisque nous allons associer aux marqueurs cartographiés dans la population intraspécifique une valeur de différenciation génétique interspécifique estimée en population naturelle. Nous réaliserons ainsi une cartographie le long des marqueurs de ces valeurs de différenciation. Le but final étant d'étudier la colocalisation entre les QTL et les valeurs de différenciation génétique interspécifique.

La distribution des indices de différenciation génétique interspécifique sur le génome a pu être réalisée en associant aux résultats de cette thèse ceux provenant d'autres études. Ils regroupent les résultats de partenaires européens ainsi que des résultats d'étudiants en thèse au laboratoire (Bodénès, 1996c ; Zanetto, 1998 ; Mariette, 2001). Au cours de ces 10 ans de recherche, les méthodes d'échantillonnage des populations naturelles ainsi que les techniques de marquage de l'ADN ont évolué. Pour être aussi exhaustif que possible nous avons réalisé une synthèse de toutes les données accumulées sur la différenciation obtenues à l'aide de marqueurs différents. Cette synthèse prend en compte bien évidemment les difficultés d'analyse liées à l'hétérogénéité des données.

## D.1 MATERIEL ET METHODES

### D.1.1 ECHANTILLONNAGE DES POPULATIONS NATURELLES

Les comparaisons de fréquences alléliques entre les deux espèces ont toujours été réalisées sur la base de paires de populations (sessile-pédonculé) autant que possible récoltées dans différentes régions de l'aire naturelle. En règle générale, au moins une demi-douzaine de paires ont été considérées dans chaque étude. Cependant, rétrospectivement, deux générations de dispositifs peuvent être considérées. Les premières études qui firent l'objet de 2 thèses (Bodénès, 1996c ; Zanetto, 1998) ont été menées sur des plantules élevées en pépinière issues de glands prélevés dans des chênaies correspondant à chaque paire, par notre équipe ou par des collègues d'autres pays européens. Dans ces premières études, une paire était constituée d'un peuplement de chênes sessiles et de chênes pédonculés aussi purs que possible et séparés au maximum par 100 km (à quelques exceptions près) ; ces paires provenant de régions diversifiées de l'aire naturelle (depuis l'Espagne jusqu'à la Scandinavie). Bien évidemment il n'y avait pas d'assurance absolue sur le degré de pureté spécifique des lots de graines en dehors de la confiance réciproque entre collègues. Bien qu'il soit difficile de distinguer individuellement des graines d'une espèce au sein d'un lot de graines d'une autre espèce, il ne semble pas que les lots de glands qui ont servi à ces expériences aient eu une composition mixte ou aient été pollués. En effet, le statut taxonomique phénotypique des semis, pour autant que l'on puisse en juger sur la base des caractères morphologiques au stade juvénile, était conforme à la dénomination du lot de graines. Connaissant par ailleurs la très faible différenciation génétique entre populations d'une même espèce, l'effet de l'éloignement des populations sessile et pédonculé d'une même paire sur la différenciation génétique interspécifique reste sans doute négligeable.

Grâce au soutien de l'union européenne, un échantillonnage plus précis des populations a pu être réalisé. Deux projets européens (FAIROAK et OAKFLOW) sont à l'origine de la caractérisation de populations de référence (ISP, Intensive Study Plots) dans différents pays européens pour l'étude de la

**Figure 23** : Peuplements de chênes échantillonnés pour l'évaluation de la différenciation génétique interspécifique.



\*Population en commun avec Mariette (2001).

La population espagnole *Salinasco Mendia* utilisée par Mariette (2001) dans le projet Fair Oak a été remplacée par la population *Arlaban* dans le projet Oak Flow.

**Tableau 32** : Récapitulatif des études menées sur la différenciation interspécifique entre le chêne sessile et le chêne pédonculé et nombre de marqueurs polymorphes dans la population de cartographie.

Type de marqueur	Paires de population	Taille moyenne de l'échantillon	Marqueurs étudiés	Références
Isozyme	7	120	11	Zanetto <i>et al.</i> 1994
SCARs	8	45	9	Bodénès <i>et al.</i> 1997
Microsatellites <sup>a</sup>	5	16	20	Muir <i>et al.</i> 2000
Microsatellites <sup>a</sup>	7	170	6	Mariette <i>et al.</i> 2002
Microsatellites <sup>a</sup>	11*	10	14	Goicoechea <i>et al.</i> Non publié
SNP	11*	10	8	Porth <i>et al.</i> non publié
AFLP	7	45	155	Mariette <i>et al.</i> 2002
AFLP	11*	10	218	Cette étude

\*deux populations (Suède et Danemark) contiennent une seule des deux espèces.

<sup>a</sup> Les 3 études regroupent 30 microsatellites (10 ont été analysés 2 fois).

diversité et de la différenciation entre espèces de chênes. Une population de référence consiste en un peuplement naturel d'arbres adultes composé de chênes sessiles et de chênes pédonculés en mélange en proportion à peu près égale (proche de 200 à chaque fois que c'est possible). Au plan de la répartition spatiale des espèces, un ISP devait être constitué de 3 zones : deux zones monospécifiques et une zone où les espèces sont en mélange pied à pied. Par ailleurs, l'échantillonnage était exhaustif sur une surface donnée variant de 5 à 20 hectares selon la nature du peuplement. Un protocole standard basé sur l'analyse en composante principale de 14 caractères morphologiques foliaires a été utilisé par tous les partenaires européens pour distinguer les deux espèces (Kremer *et al.*, 2002a). Un ISP est donc l'équivalent de la paire de populations étudiées dans la première phase, avec la différence que le statut taxonomique de tous les arbres d'un ISP était attribué de manière objective et qu'aucune sélection n'avait été faite au préalable (échantillonnage exhaustif). La thèse de Mariette s'inscrit dans le projet européen FAIROAK (FAIR1 PL95-0297, de 1996 à 1999) qui avait pour but de dresser une description détaillée des ressources génétiques du complexe d'espèces de chênes en Europe. Neuf ISP avaient été installés dans 8 pays différents qui ont été utilisés pour l'estimation de la différenciation interspécifique.

Ma thèse s'inscrit dans le projet Européen OAKFLOW (QLK5-2000-00960, 2001-2004) (<http://www.pierroton.inra.fr/Oakflow/>) qui a trois objectifs principaux : retracer et quantifier le flux de gènes et l'hybridation, évaluer les conséquences écologiques et génétiques du flux de gènes et de l'hybridation sur l'adaptation des peuplements de chênes, évaluer l'impact du flux de gènes sur la gestion et les régimes de sylviculture des peuplements de chênes. Dans le cadre de cette thèse, 2 ISP supplémentaires ont été ajoutés qui complètent donc les 9 étudiés par Mariette (Figure 23).

### **D.1.2 TYPES DE MARQUEURS UTILISES POUR L'EVALUATION DE LA DIFFERENCIATION GENETIQUE**

Les différents types de marqueurs utilisés pour étudier la différenciation génétique interspécifique entre le chêne sessile et le chêne pédonculé sont indiqués dans le tableau 32.

### D.1.2.1 Marqueurs isozymes

Les premières études menées sur la différenciation génétique entre le chêne sessile et le chêne pédonculé ont été réalisées avec des marqueurs isozymes. Les enzymes étant des produits primaires de l'expression des gènes, leur variabilité reflète directement la variabilité au sein de l'ADN. La variabilité est cependant limitée par la dégénérescence du code génétique qui n'implique pas forcément un changement d'acide aminé à chaque substitution de base au niveau de l'ADN. Cet outil est aussi limité par le nombre de locus révélés (50 au maximum) et par une certaine spécificité d'organe.

Parmi les 11 marqueurs isozymes utilisés par Zanetto *et al.* (1994) pour étudier la diversité et la différenciation de 7 paires de populations de chênes réparties du sud de la France à la Roumanie, 4 sont polymorphes dans la descendance de 3P et A4 et ont été cartographiés.

### D.1.2.2 Marqueurs RAPD transformés en SCAR

Afin de réaliser une recherche exhaustive des fragments qui discriminent les deux espèces Bodénès *et al.* (1996b) ont utilisé la technique RAPD basée sur la PCR qui permet d'amplifier un grand nombre de marqueurs rapidement. Le polymorphisme observé est de type présence-absence d'une bande (marqueur dominant) qui résulte de la substitution d'au moins une base dans l'un des sites de reconnaissance de l'amorce PCR. La méthode est caractérisée par une faible reproductibilité, et elle ne produit principalement que des marqueurs dominants. Elle a cependant l'avantage de « prospecter » rapidement le génome. Et c'est dans cet objectif qu'elle fut utilisée par Bodénès pour repérer des régions du génome différenciant les deux espèces de chênes. Parmi les 2800 fragments analysés par Bodénès *et al.* (1996a) seuls 36 présentaient des fréquences d'apparition significativement différentes entre les 2 espèces (soit 2%) dont l'isolement, le clonage et le séquençage ont permis la synthèse de 23 paires d'amorces spécifiques des régions discriminantes (SCAR). Neuf de ces amorces, dont 3 fournissent des marqueurs codominants révélés par SSCP, ont

ensuite été utilisées pour estimer l'indice de différenciation interspécifique dans 8 paires de populations européennes réparties du Nord de l'Espagne à la Pologne (Bodénès *et al.*, 1997). Parmi les 9 SCAR étudiés en population naturelle, 8 sont polymorphes dans la population de cartographie.

### D.1.2.3 Marqueurs microsatellites

Les marqueurs microsatellites ou SSRs pour " Simple Sequence Repeats " sont des fragments d'ADN constitués de séquences répétées en tandem de 2 à 3 paires de base. Le polymorphisme observé est lié à des variations de longueur des fragments en raison de changement du nombre d'unités de répétition. Ces marqueurs ont l'avantage d'être codominants, d'être repartis sur l'ensemble du génome et d'être reproductibles.

La mise au point de banques microsatellites chez le chêne (Kampfer *et al.*, 1998 ; Steinkellner *et al.*, 1997) a permis d'utiliser ces marqueurs pour étudier la différenciation génétique entre le chêne sessile et le chêne pédonculé. Tout d'abord, dans le cadre du Projet européen FAIROAK où six microsatellites ont été utilisés (Mariette *et al.*, 2002) puis dans le cadre du projet OAKFLOW où 14 microsatellites supplémentaires ont été analysés par une équipe Espagnole. Enfin, la collaboration avec une équipe autrichienne dans le cadre d'un troisième projet européen BIOTECH2 (BIO4 CT96-0706) a permis d'accéder à la valeur de différenciation de 10 microsatellites supplémentaires. Dans le cadre de ce projet, nous avons mis à disposition les paires de population analysées par Bodénès à l'équipe de Schloetterer pour une étude de diversité des microsatellites au sein du complexe chêne sessile-chêne pédonculé (Muir *et al.*, 2000). Au total, nous disposons donc de 30 microsatellites, pour lesquels des valeurs de différenciation ont été estimées, qui sont cartographiés sur la carte génétique du chêne.

#### **D.1.2.4 Marqueurs SNP**

Les SNP (Single Nucleotide Polymorphism) correspondent à des mutations ponctuelles isolées. Le polymorphisme est donc au niveau d'un seul nucléotide. La variation de la séquence d'ADN génomique est stable, portant sur une seule base, toutes les 100 à 300 bases environ du génome. Ainsi, beaucoup de SNP n'ont pas d'implications fonctionnelles. Cependant ceux se trouvant dans les régions codantes (SNPc) et dans les régions régulatrices des gènes, seront particulièrement intéressants pour réaliser la cartographie des maladies multifactorielles chez l'homme (étude d'association de gènes candidats impliqués dans ces maladies).

Dans le cadre du projet OAKFLOW, l'équipe autrichienne de Kornel Burg (ARCS) a utilisé la méthode de "screening différentiel" pour identifier, à partir d'ARN extraits de feuilles, tiges et racines, des gènes impliqués dans la différenciation entre le chêne sessile et le chêne pédonculé et exprimés dans différentes conditions de stress osmotique (Thèse d'Ilga Porth). Le séquençage de chaque gène a permis d'identifier plusieurs variants SNPc qui ont ensuite été utilisés en population naturelle pour estimer leur indice de différenciation interspécifique. Parmi les 24 SNP étudiés qui correspondent à 8 gènes, 6 ont pu être cartographiés sur la carte génétique.

#### **D.1.2.5 Marqueurs AFLP**

La technique AFLP (Amplified Fragment Length Polymorphisms) consiste à amplifier sélectivement des fragments d'ADN issus d'une digestion de l'ADN génomique par des enzymes de restriction (Vos *et al.*, 1995). Le polymorphisme observé correspond à une ou plusieurs mutations aux sites de restriction des enzymes ou à des mutations au niveau des nucléotides sélectifs. Comme les marqueurs RAPD, cette technique fournit rapidement un grand nombre de marqueurs répartis sur l'ensemble du génome. Elle présente en plus l'avantage d'être nettement plus reproductible (Bleas *et al.*, 1998).

La mise au point de cette technique au laboratoire a permis de l'utiliser en routine pour l'étude de la diversité et de la différenciation entre le chêne sessile et le chêne pédonculé.

**Tableau 33:** Populations de chênes échantillonnées (ISP) pour estimer l'indice de différenciation génétique interspécifique.

Pays	Nom de l'ISP	Composition	Nombre d'arbres génotypés
France	Petite Charnie	<i>Q. petraea</i> x <i>Q. robur</i>	20
The Netherlands	De Meinweg	<i>Q. petraea</i> x <i>Q. robur</i>	20
Spain	Arlaban	<i>Q. petraea</i> x <i>Q. robur</i>	20
United Kingdom	Dalkeith	<i>Q. petraea</i> x <i>Q. robur</i>	20
Austria	Sigmundsherdberg	<i>Q. petraea</i> x <i>Q. robur</i>	20
United Kingdom	Roudsea Wood	<i>Q. petraea</i> x <i>Q. robur</i>	20
Switzerland	Büren	<i>Q. petraea</i> x <i>Q. robur</i>	20
Germany	Escherode	<i>Q. petraea</i> x <i>Q. robur</i>	20
Hungary	Sopron	<i>Q. petraea</i> x <i>Q. robur</i>	20
Denmark	Hald Ege	<i>Q. petraea</i>	10
Sweden	Gysinge	<i>Q. robur</i>	10

**Tableau 34 :** Combinaisons AFLP cartographiées lues en population naturelle pour l'estimation de l'indice de différenciation interspécifique.

E-AAC/M-CAC
E-AAC/M-CAA
E-AAC/M-CAT
E-AAC/M-CCT
E-AAC/M-CTT
E-AAG/M-CAC
E-AAG/M-CTA
E-AAG/M-CTT
P-CAG/M-AGC
P-CAG/M-CAA
P-CCA/M-ATA

**D.1.2.5.1 AFLP étudiés au cours de la thèse de S. Mariette**

45 individus par espèce et par ISP (Tableau 32) ont été analysés dans le but d'étudier la diversité intra population, la différenciation entre populations à l'intérieur d'une espèce et la différenciation entre les deux espèces (soit de l'ordre de 300 à 350 individus par espèce). Quatre combinaisons AFLP ont été utilisées qui ont fourni 155 marqueurs dont 12 sont polymorphes dans la population de cartographie.

**D.1.2.5.2 AFLP étudiés au cours de cette thèse**

L'ensemble des marqueurs, à la fois cartographiés et dont des valeurs de  $G_{ST}$  étaient disponibles restait toutefois assez faible pour réaliser une véritable exploration du génome pour la différenciation interspécifique. Ma contribution à ce travail a donc consisté à compléter la carte génétique pour des marqueurs dont les fréquences alléliques dans les ISP pouvaient également être obtenues. Nous avons donc mis au point une procédure particulièrement adaptée à l'objectif : (1) tout d'abord, cartographie génétique des marqueurs puis, (2) analyse en populations naturelles des marqueurs qui ségrégeaient dans la famille de cartographie. Pour les marqueurs précédemment étudiés ces deux étapes avaient généralement été inversées, si bien que l'on se retrouvait souvent avec un excédent de marqueurs de différenciation interspécifique connue que l'on ne pouvait pas cartographier car non polymorphes dans le croisement. La procédure mise en place dans le cadre de ma thèse devait rétablir cette situation. Un protocole d'échantillonnage particulier a également été utilisé pour l'estimation des  $G_{ST}$  interspécifiques. Etant donné que le but de cette thèse était l'estimation de la différenciation génétique interspécifique, un échantillonnage de seulement 10 individus par espèce et par ISP a été réalisé. Soit un total de 200 individus analysés (100 sessiles et 100 pédonculés). Le détail de l'échantillonnage est présenté dans le tableau 33.

Les combinaisons AFLP sélectionnées pour l'estimation de la différenciation interspécifique sont 11 combinaisons d'amorces utilisées pour la construction de la carte génétique du chêne pédonculé : 3

combinaisons Pst1+Mse1 et 8 combinaisons EcoR1+Mse1 (Tableau 34). Le protocole de la technique AFLP est décrit sur le site web de Pierroton <http://www.pierroton.inra.fr/Oakflow/>.

Afin de se recaler avec les précédentes lectures réalisées en population de cartographie, les deux parents du croisement intraspécifique pédonculé (3P et A4) ont été placés sur chaque gel ainsi que les 2 parents de la population de cartographie interspécifique (11P et QS29) qui est en cours de construction. Le logiciel Saga<sup>TM</sup> LI COR a été utilisé pour la lecture des gels. A la différence de S. Mariette, la lecture des fragments ne fut pas exhaustive : nous n'avons pas lu les fragments monomorphes.

Ainsi 4 types de marqueurs ont été utilisés pour cartographier les indices de différenciation dans le génome du chêne que l'on peut classer en deux catégories : des marqueurs codominants (microsatellites, isozymes, et quelques SCAR) et des marqueurs dominants (SCAR, AFLP).

### **D.1.3 ESTIMATION DE LA DIFFERENCIATION GENETIQUE INTERSPECIFIQUE**

#### **D.1.3.1 Indices de différenciation : $G_{ST}$ de Nei ; $F_{ST}$ de Weir et Cockerham**

La structuration génétique au sein d'un ensemble de populations a d'abord été décrite par les F statistiques de Wright (1969) au niveau allélique puis généralisée au niveau d'un et de plusieurs locus par Nei (1973, 1987) et Weir et Cockerham (1984).

##### **D.1.3.1.1 $G_{ST}$ de Nei**

L'approche de Nei est décrite dans l'encadré 4. Elle définit l'indice de différenciation génétique entre populations ( $G_{ST}$ ) de la manière suivante au niveau d'un locus :

$$G_{ST} = \frac{H_T - H_S}{H_T} = 1 - \frac{H_S}{H_T}$$

## Encadré 4 : Différenciation génétique par l'approche de Nei (1973)

La stratégie suivie pour estimer la différenciation entre les populations consiste à décomposer la diversité génétique en deux niveaux hiérarchiques : populations individuelles (sous-populations) et ensemble des populations. Pour caractériser la diversité génétique de ces niveaux, Nei (1973) considère un locus multiallélique :

Soient  $a$  le nombre d'allèles au niveau d'un locus,  $l$  le nombre de sous-populations et  $t$  la population globale virtuelle qui englobe les  $l$  sous-populations. Il est alors possible d'estimer les paramètres de diversité suivants :

### ○ Diversité dans une sous population : $h$

Elle correspond à la probabilité pour que deux allèles tirés au hasard avec remise dans la sous population soient différents

$$h = 1 - \sum_{i=1}^a p_i^2 \quad (1)$$

avec  $p_i$  la fréquence de l'allèle  $i$  au locus considéré et  $a$  le nombre d'allèles à ce locus.

### ○ Diversité dans les $l$ sous populations : $H_S$

Elle correspond à la diversité moyenne au sein des  $l$  sous populations

$$H_S = 1 - \sum_1 \overline{p_i^2} \quad (2)$$

avec  $\overline{p_i^2}$  la moyenne sur les  $l$  sous populations des carrés des fréquences de l'allèle  $i$ .

### ○ Diversité sur l'ensemble des populations considérées comme une seule population globale : $H_T$

Elle est calculée en prenant pour fréquences dans l'équation 1 les fréquences moyennes sur toutes les populations :

$$H_T = 1 - \sum_1 \overline{p_i}^2 \quad (3)$$

avec  $\overline{p_i}$  la moyenne pondérée de la fréquence de l'allèle  $i$  dans les  $l$  sous-populations

### ○ Différenciation globale entre les $l$ sous populations

La différenciation globale entre les sous-populations ( $d_{st}$ ) est définie par Nei comme la différence entre la diversité totale et la diversité moyenne dans les  $l$  sous-populations :

$$d_{st} = H_T - H_S \quad (4)$$

### ○ Définition de l'indice de différenciation génétique de Nei ( $G_{ST}$ )

Nei exprime la différenciation globale entre les sous-populations sous forme relative en la rapportant à la diversité totale :

$$G_{ST} = \frac{H_T - H_S}{H_T} = \frac{d_{st}}{H_T} \quad (5)$$

avec  $H_S$  la moyenne sur toutes les populations des diversités génétiques intra population et  $H_T$  la diversité génétique sur l'ensemble des populations considérées comme une population unique (diversité totale).

Le  $G_{ST}$  a été utilisé pour estimer la différenciation interspécifique au niveau des marqueurs AFLP (Mariette, 2001 ; travail de thèse), Isozymes (Zanetto *et al.*, 1994), SCAR (Bodénès *et al.*, 1997), et certains microsatellites (6) (Mariette *et al.*, 2002).

#### D.1.3.1.2 Estimation du $G_{ST}$ dans le cas de marqueurs dominants

La mesure de la diversité allélique pour des marqueurs dominants (type RAPD, AFLP) requiert que les fréquences des allèles 1 (présence du fragment) et 0 (absence du fragment) soient connues. Dans le cas où la population est à l'équilibre de Hardy-Weinberg, la fréquence  $q$  de l'allèle nul peut être déduite de la fréquence  $Q$  des individus dont le profil électrophorétique ne montre pas ce fragment :  $q = \sqrt{Q}$ . Dans le cas où la population n'est pas à l'équilibre de H-W, la fréquence  $q$  devient (Le Corre *et al.*, 1997) :

$$q = \left( \frac{-1}{2(1-F)} \right) \left( F - \sqrt{F^2 + 4(1-F)Q} \right)$$

où  $F$  est l'indice de fixation, qui mesure l'écart entre l'hétérozygotie théorique à l'équilibre ( $H_e$ ) et l'hétérozygotie observée ( $H_0$ ) ( $F = 1 - \frac{H_0}{H_e}$ ). Dans le cas de marqueurs neutres,  $F$  ne dépend que du régime de reproduction et de la structure de la population.

L'introduction d'une valeur de  $F$  estimée à partir des données microsatellites dans l'analyse génotypique de Mariette *et al.*, (2002) ( $F_{IS}$  moyen sur 6 microsatellites=0.08) donne une valeur de différenciation inférieure mais la plupart du temps non significativement différente de la valeur obtenue à partir d'une analyse phénotypique ne prenant pas en compte ce paramètre. L'évaluation du  $G_{ST}$  dans notre étude a été réalisée en faisant l'hypothèse de  $F_{IS} \approx 0$  et nous avons ré-analysé les données de Mariette de la même manière avec  $F_{IS} \approx 0$ .

Lynch et Milligan (1994) insistent sur le biais très important qui peut affecter l'estimation des paramètres génétiques faite à partir des marqueurs dominants dans le cas de marqueurs très peu polymorphes (dont la fréquence de présence de bandes est supérieure à  $1-3/N$ , où  $N$  est la taille de l'échantillon). Nous n'avons pas appliqué cette restriction pour les raisons suivantes :

- Compte tenu des tailles d'échantillon ( $N$  la plupart du temps proche de 100 individus par espèce), la proportion de marqueurs susceptibles d'être éliminés (marqueurs dont la fréquence de bandes était supérieure à 97 %) restait très modeste.
- Dans la suite, ce n'est pas la valeur du paramètre ( $G_{ST}$ ) qui sera prise en compte mais la signification statistique de la différenciation entre les deux espèces. Dans tous les cas, le test statistique appliqué pour vérifier la différenciation, est indépendant de la fréquence de la présence de bandes AFLP.

#### D.1.3.1.3 $F_{ST}$ (ou theta) de Weir et Cockerham

Dans cette approche les populations sont organisées de façon hiérarchique. On considère des allèles à l'intérieur d'individus, des individus à l'intérieur de sous-populations et des sous-populations à l'intérieur de la population globale.

Dans le cas simplifié d'un locus à 2 allèles  $A$  et  $a$ , de fréquences  $p$  et  $q=1-p$  respectivement, le modèle d'ANOVA suivant permet d'estimer les différentes composantes de la variance :

$$Y_{ijk} = p + a_k + b_{ik} + w_{ijk}$$

Avec  $Y_{ijk}$  la valeur du  $j^{\text{ième}}$  allèle (1 ou 0) pour le  $i^{\text{ième}}$  individu de la  $k^{\text{ième}}$  population,  $a_k$  l'effet population qui rend compte de la structuration entre populations,  $b_{ik}$  l'effet individu intra population qui rend compte de la répartition des individus dans les populations,  $w_{ijk}$  la résiduelle qui rend compte de la répartition des allèles à l'intérieur des individus.

Il est alors possible d'estimer l'indice de différenciation entre populations  $F_{ST}$  (aussi appelé théta) comme le rapport entre la variation entre les sous-populations ( $V_a$ ) et la variation totale ( $V_t$ )

$$F_{ST} = \theta = \frac{V_a}{V_t} = \frac{V_a}{V_a + V_b + V_w}$$

Avec  $V_a$  la variance entre sous-populations,  $V_b$  la variance entre individus appartenant à la même sous-population, et  $V_w$  la variance à l'intérieur des individus. Le  $F_{ST}$  est avant tout un paramètre de différenciation relatif à un allèle donné (dans le cas particulier de l'allèle A). Une estimation pluri-allélique est donnée dans Weir et Cockerham 1984 en sommant les variances associées aux différents allèles au numérateur et dénominateur de l'expression précédente.

C'est cette approche que les équipes allemandes et espagnoles ont suivie pour estimer les indices de différenciation interspécifique au niveau de 24 microsatellites.

#### D.1.3.1.4 Quelles différences entre $G_{ST}$ et $F_{ST}$ ?

L'estimation du  $G_{ST}$  de Nei et du  $F_{ST}$  par le  $\theta$  de Weir et Cockerham utilise des modèles qui diffèrent par leur hypothèse sous-jacente. Alors que Weir et Cockerham font l'hypothèse que les populations dérivent d'une même population ancêtre et qu'elles ont évolué sans migration, mutation, sélection, ce qui suppose une différenciation par dérive génétique, Nei ne fait aucune hypothèse sur l'histoire des populations (Kremer, 1997). Une autre différence est liée à la méthode d'estimation des paramètres. Par ailleurs  $G_{ST}$  est une statistique pluri-allélique alors que  $F_{ST}$  est un paramètre relatif à un allèle. Nei montre que les estimateurs des deux paramètres convergent quand le nombre de populations et le nombre d'individus par population sont élevés. Quand le nombre de populations est faible, comme dans notre cas où il s'agit de différenciation interspécifique, la convergence n'est observée que pour  $G_{ST}'$  et  $F_{ST}$ :

$$G_{ST}' = \frac{D_{ST}'}{H_T'}$$

$$\text{avec } D_{ST}' = \frac{N}{N-1} D_{ST} \quad \text{et} \quad H_T' = H_S - D_{ST}'$$

où  $N$  est le nombre de populations analysées.

Dans notre cas où  $N=2$ ,  $G_{ST}'$  s'écrit donc :

$$G_{ST}' = \frac{2D_{ST}}{H_S + 2D_{ST}} \quad \text{à comparer avec } G_{ST} = \frac{D_{ST}}{H_S + D_{ST}}$$

Dans le cas où  $G_{ST}$  est faible (donc  $D_{ST}$  faible par rapport à  $H_S$ ) on peut faire l'approximation suivante :

$$G_{ST} \approx \frac{1}{2} G_{ST}' = \frac{1}{2} F_{ST}$$

A titre d'exemple si  $G_{ST}$  est de l'ordre de 0.10, l'estimation de  $G_{ST}$  par la formule précédente est 0.091. De la même manière si  $G_{ST}$  est de 0.05, l'estimation de  $G_{ST}$  par la formule précédente est de 0.047.

### D.1.3.2 Tests des indices de différenciation

Trois types de tests ont été réalisés en fixant un seuil de rejet de l'hypothèse nulle (absence de différenciation entre les populations) à 5%.

- Pour les marqueurs codominants (certains SCAR (Bodénès *et al.*, 1997), isozymes (Zanetto *et al.*, 1994) et microsatellites étudiés au laboratoire (Mariette *et al.*, 2002), une table de contingence (espèce, allèle) a été construite et un test de  $\chi^2$  a été utilisé (logiciel DIPLOID écrit par A. Kremer).
- Pour les marqueurs dominants (AFLP et certains SCAR), une table de contingence (espèce, présence de bandes) a été construite et l'homogénéité des fréquences de bandes entre les deux espèces a été testée par un test de permutations (1000 permutations, logiciel SPAGEDI, Hardy et Vekemans, 2002).
- Pour les autres marqueurs codominants non étudiés au laboratoire (microsatellites (Muir *et al.*, 2000) et Goicoechea *et al.* (données non publiées)), la différenciation entre espèces a aussi été testée en faisant des tests de permutations (au niveau des allèles) (Goudet *et al.*, 1996, logiciel FSTAT).

### D.1.3.3 Comparaison du niveau de différenciation estimé avec différents types de marqueurs

Dans le cas d'un modèle simplifié d'évolution des populations (sans sélection), où des populations de même taille échangent toutes un nombre égal de gènes (modèle en îles de Wright, 1969), la valeur attendue à l'équilibre entre migration et dérive s'écrit (Nei, 1987) :

$$G_{ST} = \frac{1}{1 + (4N(\frac{s}{s-1})(\mu + m))}$$

Avec  $s$ , le nombre de populations,  $N$  la taille de la population,  $\mu$  le taux de mutation et  $m$  le taux de migration.

Si le taux de mutation  $\mu$  est très inférieur au taux de migration  $m$ , et que le nombre de populations  $s$  échantillonnées est très élevé, cette relation prend la forme la plus connue (Hartl et Clark, 1989) :

$$G_{ST} \approx \frac{1}{1 + 4Nm}$$

En théorie, la différenciation est relativement indépendante des marqueurs considérés, pour autant que ceux-ci soient neutres. En effet, sa valeur à l'équilibre n'est pas affectée par le taux de mutation, quand celui-ci est négligeable devant le taux de migration : elle dépend principalement du taux de migration et de la taille des populations qui constituent des facteurs agissant dans le même sens et avec la même amplitude sur les différents marqueurs (le nombre de populations  $s$  a moins d'influence).

Cependant, pour des marqueurs très polymorphes comme les microsatellites, on peut s'attendre à observer une différenciation inférieure aux autres marqueurs car le taux de mutation est très élevé ( $\approx 10^{-3}$ ). Une analyse comparative de la différenciation génétique entre populations estimée à partir de différents types de marqueurs a été réalisée par Mariette (2001) en s'appuyant sur des résultats recensés dans la littérature ainsi que sur ses propres résultats expérimentaux. La synthèse bibliographique de 18 articles dans lesquels la différenciation génétique a été estimée avec différents

types de marqueurs (notamment isozymes, RAPD, microsatellites), indique que seuls 28% (5 cas dont 2 n'incluent pas de microsatellites) présentent des différences significatives dans l'estimation de la différenciation. Pour les 8 cas restants (72%) dont 4 comparent des isozymes à des microsatellites (pourvus de taux de mutations différents), les différences ne sont pas significatives. Dans son étude expérimentale qui compare le niveau de différenciation estimé à partir de marqueurs microsatellites avec celui estimé à partir de marqueurs AFLP, elle montre cependant que la différenciation génétique interspécifique entre le chêne sessile et le chêne pédonculé est significativement supérieure pour les marqueurs AFLP ( $G_{ST}=0.03$ ) que pour les microsatellites ( $G_{ST}=0.013$ ).

#### D.1.4 DISTRIBUTION DES VALEURS DE $G_{ST}$

Les histogrammes de distribution des valeurs de  $G_{ST}$  pour les marqueurs AFLP et microsatellites ont été réalisés en répartissant les  $G_{ST}$  dans 9 classes de valeurs. La première classe étant :  $G_{ST}<0.005$  et la dernière  $G_{ST}> 0.1$ . Les marqueurs AFLP Eco-Mse et Pst-Mse ont tout d'abord été traités séparément puis globalement. La principale différence entre les combinaisons Pst-Mse et Eco-Mse réside dans la sensibilité à la méthylation des enzymes Pst qui couperont préférentiellement dans les zones non méthylées du génome (voir par exemple Vuylsteke *et al.*, 1999).

La distribution des classes de  $G_{ST}$  en fonction des types de combinaisons AFLP utilisés (Eco-Mse / Pst-Mse) ainsi qu'en fonction des types de marqueurs (AFLP / microsatellites) a été testée en utilisant l'option différenciation génique entre population du logiciel GenePop (Raymond et Rousset, 1995a). Par analogie, les classes de  $G_{ST}$  correspondent aux allèles d'un locus et les types de marqueurs (ou combinaisons AFLP) à des sous-populations. Dans une première étape, un tableau de contingence est construit sur lequel est ensuite réalisé un test de Fisher exact selon la méthode décrite par Raymond et Rousset (1995b). L'hypothèse nulle  $H_0$  testée est : la distribution des classes de  $G_{ST}$  est identique entre les types de marqueurs (ou combinaisons AFLP).

La moyenne des valeurs de  $G_{ST}$  entre les différents types de marqueurs (ou combinaisons AFLP) a ensuite été testée en utilisant le test non paramétrique des sommes de rangs de Wilcoxon qui, à la

différence du test t de Student, est indépendant de la taille de l'échantillon et de la distribution de la variable mesurée. Le test des sommes des rangs de Wilcoxon consiste à ranger les observations provenant des 2 échantillons de façon croissante et à comptabiliser les rangs du plus petit échantillon. La valeur obtenue (W) est la statistique de Wilcoxon qui permet en fonction de l'effectif des deux échantillons de déterminer le niveau de signification. Par abus de langage, il est dit que ce test permet de comparer deux moyennes, en réalité, l'hypothèse nulle  $H_0$  testée est : les deux échantillons possèdent la même médiane (valeur pour laquelle 50% des mesures sont plus grandes et 50% plus petites).

### **D.1.5 CONSTRUCTION DE LA CARTE GENETIQUE ET POSITIONNEMENT DES QTL**

Nous avons sélectionné pour la construction de la carte génétique :

- Tous les marqueurs pour lesquels une valeur de  $G_{ST}$  interspécifique était disponible.
- Tous les marqueurs microsatellites car ils permettent de faire la jonction entre la carte mâle et la carte femelle.
- Tous les marqueurs AFLP analysés sur les 275 descendants de la population de cartographie (tous n'avaient pas été sélectionnés pour la construction de la carte jalon utilisée pour la détection QTL, notamment les marqueurs qui ségrégeaient en proportion 3 :1.).

La carte génétique réalisée pour la cartographie des  $G_{ST}$  est une carte consensus qui réunit les deux cartes parentales, construite avec le logiciel Join map version 3.0 (Van Ooijen et Voorrips, 1993). Dans une première étape, nous avons construit les deux cartes parentales en utilisant la stratégie double pseudo testcross décrite dans le chapitre C.1.1.2. Les marqueurs ont tout d'abord été placés dans les groupes de liaison en se fixant un seuil de LOD >5 puis ordonnés à l'intérieur des groupes en se fixant un LOD >3. Dans une seconde étape, les marqueurs communs aux 2 cartes parentales ont permis la construction d'une carte consensus. Avant l'intégration des deux cartes, réalisée en se fixant un seuil de LOD >2, l'homogénéité de recombinaison entre les paires de marqueurs a été testée en

utilisant un test G. Lorsque le test révélait une hétérogénéité de recombinaison supérieure à 1% les marqueurs étaient exclus de la carte consensus.

Afin de pouvoir comparer la distribution des QTL détectés dans le chapitre C avec la distribution des  $G_{ST}$  le long du génome, nous avons positionné sur la carte consensus les QTL détectés sur les cartes jalons mâle et femelle. Pour cela nous avons considéré l'intervalle de la position du QTL qui correspond à la distance entre P1, la position au niveau du LOD maximal et P2 la position obtenue après 1000 ré-échantillonnages par bootstrap. Nous avons ensuite considéré sur les cartes jalons la distance qui sépare chaque extrémité des intervalles avec les marqueurs les plus proches afin de la reporter sur la carte consensus.

#### **D.1.6 STRUCTURATION LE LONG DES GROUPES DE LIAISON DES VALEURS DE DIFFERENCIATION INTERSPECIFIQUE**

Le but de notre étude étant de tester la structuration "spatiale" dans le génome des loci impliqués dans la différenciation génétique interspécifique, nous avons classé tous les marqueurs pour lesquels une valeur de  $G_{ST}$  était disponible en 2 catégories :

- ceux pour lesquels la valeur de  $G_{ST}$  était significative au seuil de  $p=0.05$
- ceux pour lesquels elle ne l'était pas.

Nous avons donc privilégié une analyse qualitative (différenciation significative versus non significative) à une analyse quantitative (niveau de signification). Cette manière de procéder est basée sur l'argumentaire suivant :

- les valeurs de  $G_{ST}$  ont été calculées sur la base de marqueurs aux propriétés différentes (dominants et codominants) dont les estimations sont affectées par ces propriétés.
- les tailles d'échantillon utilisées pour estimer  $G_{ST}$  ayant varié historiquement selon le type de populations étudiées (paire ou ISP), les intervalles de confiance correspondants aux différentes valeurs de  $G_{ST}$  sont donc également variables.

### D.1.6.1 Comparaison à la distribution aléatoire

De la même manière que pour les QTL, la distribution des marqueurs (à  $G_{ST}$  significatifs et non significatifs) a été comparée à une distribution de Poisson ayant comme moyenne la densité de marqueurs de chaque catégorie par intervalle. Pour cela, nous reprenons ici la subdivision du génome en  $N$  intervalles en modifiant cependant la taille de l'intervalle. Par opposition aux QTL, la position d'un marqueur est nettement plus précise (inférieur au cM, pour des marqueurs séparés par 1cM). Par ailleurs, nous nous doutons que s'il y a regroupement des marqueurs à  $G_{ST}$  élevés, ce regroupement est dû à la liaison génétique et à l'effet d'entraînement. Sachant que les chênes sont des espèces à taille de population élevée, l'effet d'entraînement doit sans doute se limiter à une faible distance. Nous considérerons donc tout particulièrement des intervalles de faible taille. Dans ce qui suit, nous considérerons 2 intervalles (0 à 1.5cM et 0 à 3cM).

En plus de la comparaison de la distribution des marqueurs à une Loi de Poisson faite séparément pour chaque classe de marqueurs (significatifs et non significatifs), nous avons également utilisé un test utilisant les deux classes simultanément. En effet la comparaison à la distribution de Poisson est faite sur l'ensemble du génome, alors que de nombreuses régions sont "non prospectées", violant en cela l'hypothèse d'homogénéité requise par les processus poissonien. Nous avons donc réalisé un second test qui lui n'est basé que sur les régions effectivement prospectées et dans lesquels il y avait au moins deux marqueurs disponibles.

Nous pouvons dénombrer les intervalles correspondant aux différentes cas de figures suivants :

	Nombre d'intervalles où les $G_{ST}$ sont significatifs	Nombre d'intervalles où les $G_{ST}$ ne sont pas significatifs
Nombre d'intervalles où les $G_{ST}$ sont significatifs	N11	N12
Nombre d'intervalles où les $G_{ST}$ ne sont pas significatifs	N21	N22

Un test de Fisher exact permettra de voir s'il y a co-occurrence des  $G_{ST}$  significatifs dans les mêmes régions.

**D.1.6.2 Analyse de la distribution spatiale par autocorrélation.**

La deuxième manière d'étudier la distribution des valeurs de  $G_{ST}$  le long des groupes de liaison est également inspirée des analyses spatiales faites en génétique des populations. Elle consiste à calculer la corrélation entre les paires de marqueurs situés à une distance  $l$  sur les groupes de liaison, chaque marqueur pouvant avoir deux valeurs (0, s'il ne manifeste pas de valeur de  $G_{ST}$  significatif, 1 dans le cas contraire). Cette corrélation est également appelée indice  $I$  de Moran (Sokal, 1978). Pour chaque classe de distance  $q$ , l'indice  $I_q$  de Moran est calculé :

$$I_q = \frac{n \sum_{i=1}^n \sum_{j \neq 1}^n w_{ij} (a_i - \bar{a})(a_j - \bar{a})}{W \sum_{i=1}^n (a_i - \bar{a})^2} \text{ et } W = \sum_{i=1}^n \sum_{j \neq 1}^n W_{ij}$$

$n$  est le nombre total de marqueurs,  $W_{ij}$  est égal à 1 si les marqueurs  $i$  et  $j$  appartiennent à la classe de distance  $q$ . Dans cette étude nous avons considéré deux classes de distances (pour  $q=1$ ,  $l$  varie de 0 à 1.5 cM et pour  $q=2$ ,  $l$  varie de 1.5 à 3 cM.).

$a_i$  est égal à 1 si le marqueur  $i$  présente un  $G_{ST}$  significatif et 0 dans le cas contraire,  $\bar{a}$  est la moyenne des  $a_i$  sur les  $n$  marqueurs.

Les valeurs de  $I$  sont testées par un test empirique de permutation. Chaque permutation consiste en une redistribution aléatoire des valeurs de  $G_{ST}$  sur l'ensemble des positions des marqueurs. Pour chaque classe de distance, les valeurs observées sont comparées à la distribution obtenue après 1000 permutations. Le logiciel SGS (Version 1.0 c, Degen B., 2000) a été utilisé pour réaliser les analyses (<http://kourou.cirad.fr/génétique/software.html>).

### D.1.7 ASSOCIATIONS ENTRE MARQUEURS ET CARACTÈRES QUANTITATIFS

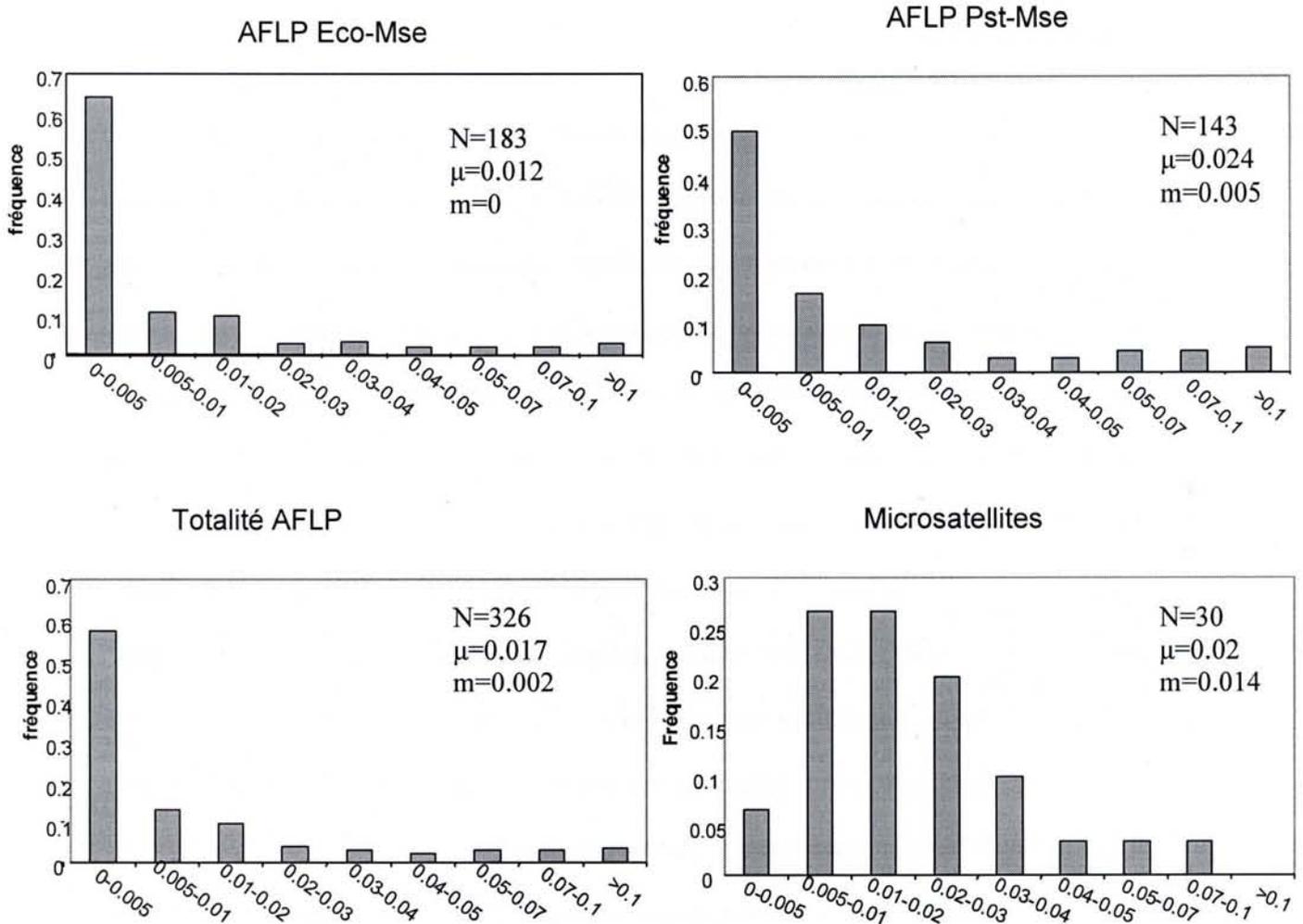
Dans la première partie de cette thèse, 14 caractères morphologiques foliaires qui discriminent le chêne sessile et le chêne pédonculé ont été utilisés pour détecter des QTL dans une famille  $F_1$  intra spécifique de chêne pédonculé. Sur tous les arbres constituant les peuplements de référence, 5 feuilles ont été mesurées en moyenne au niveau de ces caractères afin de leur assigner une espèce sur la base d'analyse en composantes principales (Kremer *et al.*, 2002a). Parmi les 14 caractères, 9 ont été retenus (variables dérivées non incluses) pour tester l'association statistique entre les 98 marqueurs AFLP cartographiés utilisés pour l'estimation des indices de différenciation interspécifique et des caractères phénotypiques discriminants les deux espèces. Malheureusement, les mesures des autres caractères phénotypiques étudiés dans les chapitres précédents sur la population de cartographie, n'étaient pas disponibles en populations naturelles. La recherche d'association statistique a été réalisée au niveau interspécifique et intraspécifique (intra sessile et intra pédonculé) en utilisant pour chaque individu la moyenne du caractère mesuré (sur les 5 feuilles).

Il est important de faire ici la distinction entre liaison statistique (ce que nous testons) et la liaison physique, sans préjuger des causes de cette association statistique. La liaison statistique entre marqueur et caractère a été testé par ANOVA avec le logiciel Splus (1996), en appliquant le modèle suivant :

$$Y_{ijk} = \mu + M_i + P_j + M_i * P_j + E_{ijk}$$

où  $Y_{ijk}$  est la valeur du caractère mesuré sur l'individu  $k$  au niveau du marqueur  $i$  dans le peuplement  $j$ ,  $\mu$  est la moyenne générale du caractère phénotypique,  $M_i$  est l'effet du marqueur,  $P_j$  est l'effet peuplement, et  $M_i * P_j$  est l'interaction entre marqueur et peuplement. Les deux analyses réalisées au niveau intraspécifique ont porté sur 100 arbres (10 arbres par peuplement) alors que l'analyse au niveau interspécifique a été réalisée en cumulant les données, soit sur 200 arbres.

**Figure 24** : Histogramme de distribution des indices de différenciation génétique interspécifique pour les marqueurs AFLP en fonction des combinaisons d'enzymes utilisées et pour les microsatellites.



Avec  $\mu$  la moyenne des valeurs de  $G_{ST}$ ,  $m$  la médiane.

**Tableau 35** : Bilan des marqueurs cartographiés

Marqueur	Marqueurs analysés en population naturelle		Marqueurs NON analysés en population naturelle
	Total	Cartographié	Cartographié
AFLP Eco-Mse	183 <sup>1</sup>	80	66
AFLP Pst-Mse	190 <sup>2</sup>	30	4
Microsatellites	30 <sup>3</sup>	30	18
SNP	8 <sup>4</sup>	6	0
Scars	9 <sup>5</sup>	5	0
Isozymes	11 <sup>6</sup>	4	0
RAPD	-	-	12
<b>Somme</b>	<b>423</b>	<b>155</b>	<b>100</b>

<sup>1</sup> Saintagne (travail de thèse)

<sup>2</sup> Mariette *et al.*, 2002 et Saintagne (travail de thèse)

<sup>3</sup> Mariette *et al.*, 2002 ; Goicoechea (non publié) ; Muir *et al.*, (2000)

<sup>4</sup> Porth *et al.*, non publié

<sup>5</sup> Bodénès *et al.*, 1997

<sup>6</sup> Zanetto *et al.*, 1994

L'étude d'association entre marqueurs et caractères a aussi été menée par Mariette au cours de sa thèse (résultats non publiés). Parmi les 155 marqueurs étudiés, seuls les résultats obtenus pour les marqueurs cartographiés (12) seront présentés.

## D.2 RESULTATS

### D.2.1 DISTRIBUTION DES VALEURS DE $G_{ST}$ SELON LES MARQUEURS

La distribution des classes de  $G_{ST}$  pour les combinaisons AFLP Eco-Mse et Pst-Mse suit une courbe en L (Figure 24) avec beaucoup de marqueurs qui présentent des faibles  $G_{ST}$  et peu qui présentent des forts  $G_{ST}$ , indiquant ainsi une forte homogénéité entre le génome des deux espèces de chênes avec quelques régions impliquées dans la différenciation interspécifique. On peut cependant noter plus de marqueurs à fort  $G_{ST}$  pour les Pst-Mse : 34% ont un  $G_{ST}$  supérieur à 0.01 contre 25% pour les Eco-Mse. Le test exact de Fisher réalisé pour tester s'il existe des différences entre la distribution des classes de  $G_{ST}$  pour les combinaisons Eco-Mse et Pst-Mse n'est pas significatif. Cependant, la moyenne des  $G_{ST}$  des marqueurs AFLP Pst-Mse est deux fois supérieure à celles des Eco-Mse (0.024 contre 0.012). La non normalité des distributions n'a pas permis de comparer ces deux moyennes avec un test t de Student cependant le test de Wilcoxon indique des différences significatives entre les médianes ( $p=7.610 \cdot 10^{-4}$ ). Ces différences pourraient s'expliquer par une sensibilité différente des deux enzymes vis à vis de la méthylation des nucléotides cytosines. En effet, différentes études montrent que les enzymes Pst sont sensibles à la méthylation de la cytosine fournissant ainsi des marqueurs localisés dans des régions hypo-méthylées (Young *et al.*, 1999, Castiglioni *et al.*, 1999, Vuylsteke *et al.*, 1999). Or, l'activité transcriptionnelle de nombreux gènes est réprimée par la méthylation des cytosines au niveau des régions régulatrices (Kass, 1997). En principe donc, Pst couperait préférentiellement dans les zones codantes qui sont exprimées, alors que l'enzyme de restriction Eco coupe de manière aléatoire dans le génome. Les marqueurs AFLP résultant des digestion par Pst-Mse

ne seraient donc pas totalement neutres, leur polymorphisme pourrait être en partie dû à des gènes différenciellement exprimés entre le chêne sessile et le chêne pédonculé.

La distribution des classes de  $G_{ST}$  pour les marqueurs microsatellites est significativement différente de la distribution des AFLP (Test exact de Fisher significatif). Comparés aux marqueurs AFLP, les microsatellites présentent moins de faibles valeurs de  $G_{ST}$ , avec 65% des valeurs supérieures à 0.01. La moyenne des  $G_{ST}$  sur les 30 microsatellites est de 0.02 ce qui est proche des marqueurs AFLP Pst-Mse (0.024) avec cependant des médianes significativement différentes ( $p=0.0014$ ). Si l'on cumule les marqueurs des deux combinaisons AFLP, le  $G_{ST}$  moyen est de 0.018, ce qui est inférieur au  $G_{ST}$  moyen des microsatellites. Ce résultat est surprenant puisqu'on s'attendrait plutôt à observer des  $G_{ST}$  en moyenne inférieurs pour les microsatellites (cf D.1.3.3). Il peut en partie être lié à la manière dont ces marqueurs ont été développés. Rappelons que la mise au point de marqueurs microsatellites suit un certain nombre de filtres de sélection comprenant une étape qualifiée de "Polymorphisme" où seuls les plus polymorphes sont retenus. Les microsatellites utilisés aujourd'hui en routine pour les analyses populationnelles ne représentent qu'un sous-ensemble de tous les microsatellites possibles, le sous-ensemble le plus reproductible et le plus polymorphe. Il n'y a pas de procédure semblable pour les AFLP qui sont de choix plus aléatoires que les microsatellites. Seuls ont été exclus de cette analyse les AFLP totalement monomorphes (présence d'une bande chez tous les individus sessile et pédonculé) qui ne pouvaient donc, de ce fait être cartographiés, et dont la valeur de  $G_{ST}$  était nulle.

### D.2.2 CONSTRUCTION DE LA CARTE GENETIQUE CONSENSUS

En plus des marqueurs sélectionnés au départ, nous avons rajouté quelques marqueurs RAPD de la précédente carte (Barreneche *et al.*, 1998) qui ségrégeaient en proportion 3 :1 de façon à réunir des groupes de liaisons mâle et femelle qui ne présentaient pas assez de marqueurs ponts.

La carte consensus est composée de 255 marqueurs (Tableau 35) répartis sur 902 cM avec en moyenne 1 marqueur tous les 4 cM, ce qui représente 75% de la taille du génome si l'on se base sur l'estimation réalisée par Barreneche *et al.* (1998). La couverture du génome est de même ordre que

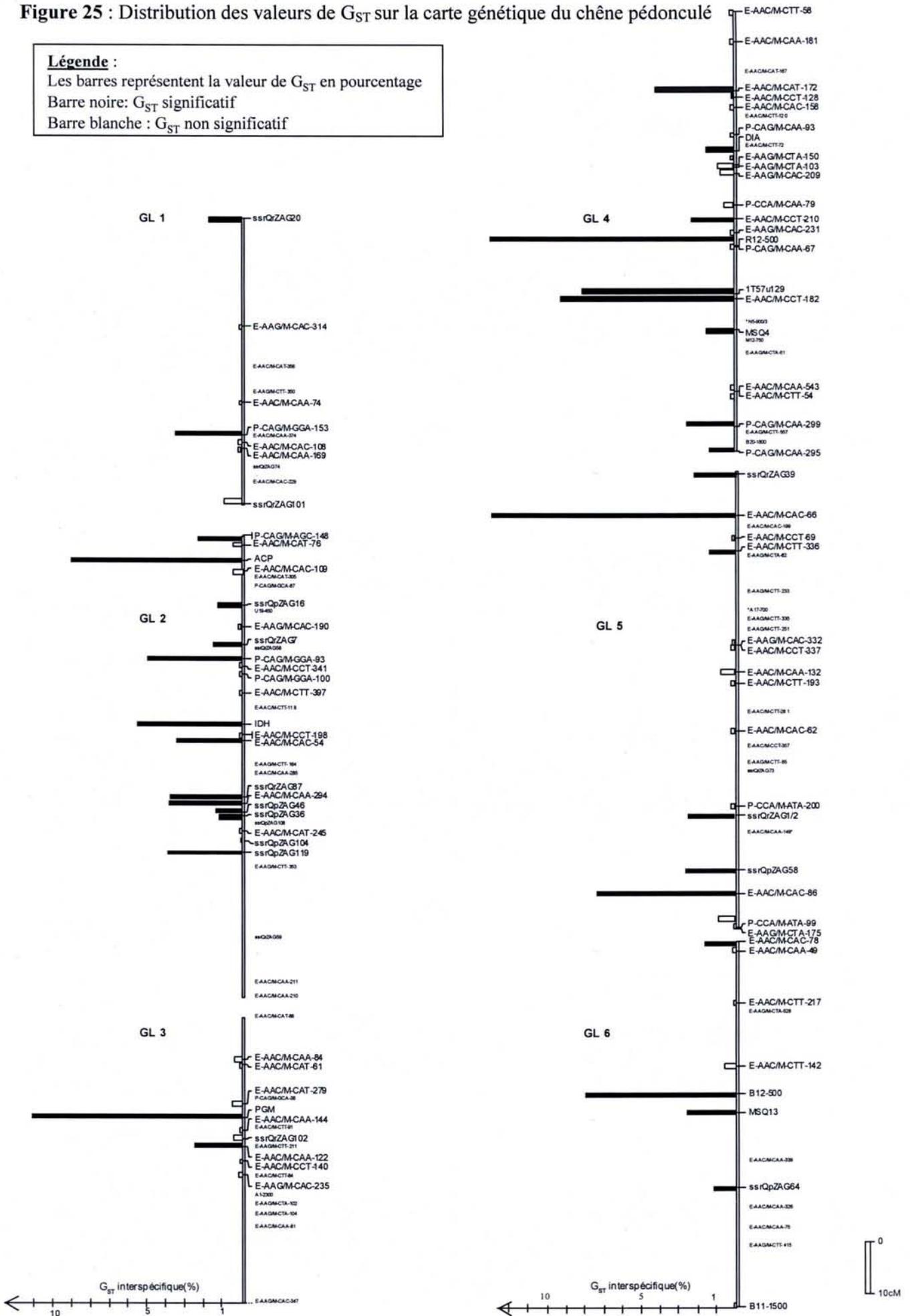
Figure 25 : Distribution des valeurs de  $G_{ST}$  sur la carte génétique du chêne pédonculé

**Légende :**

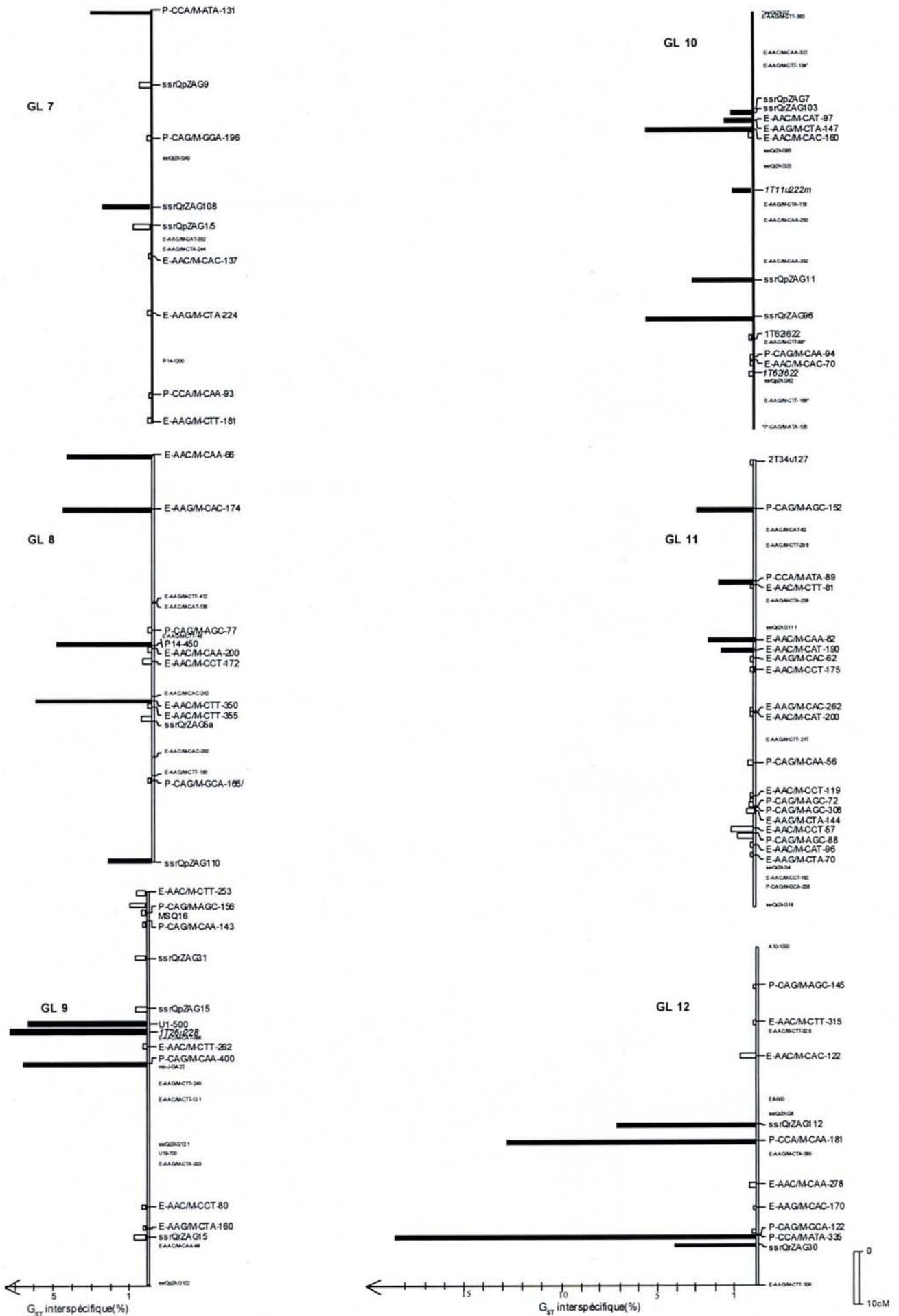
Les barres représentent la valeur de  $G_{ST}$  en pourcentage

Barre noire:  $G_{ST}$  significatif

Barre blanche :  $G_{ST}$  non significatif



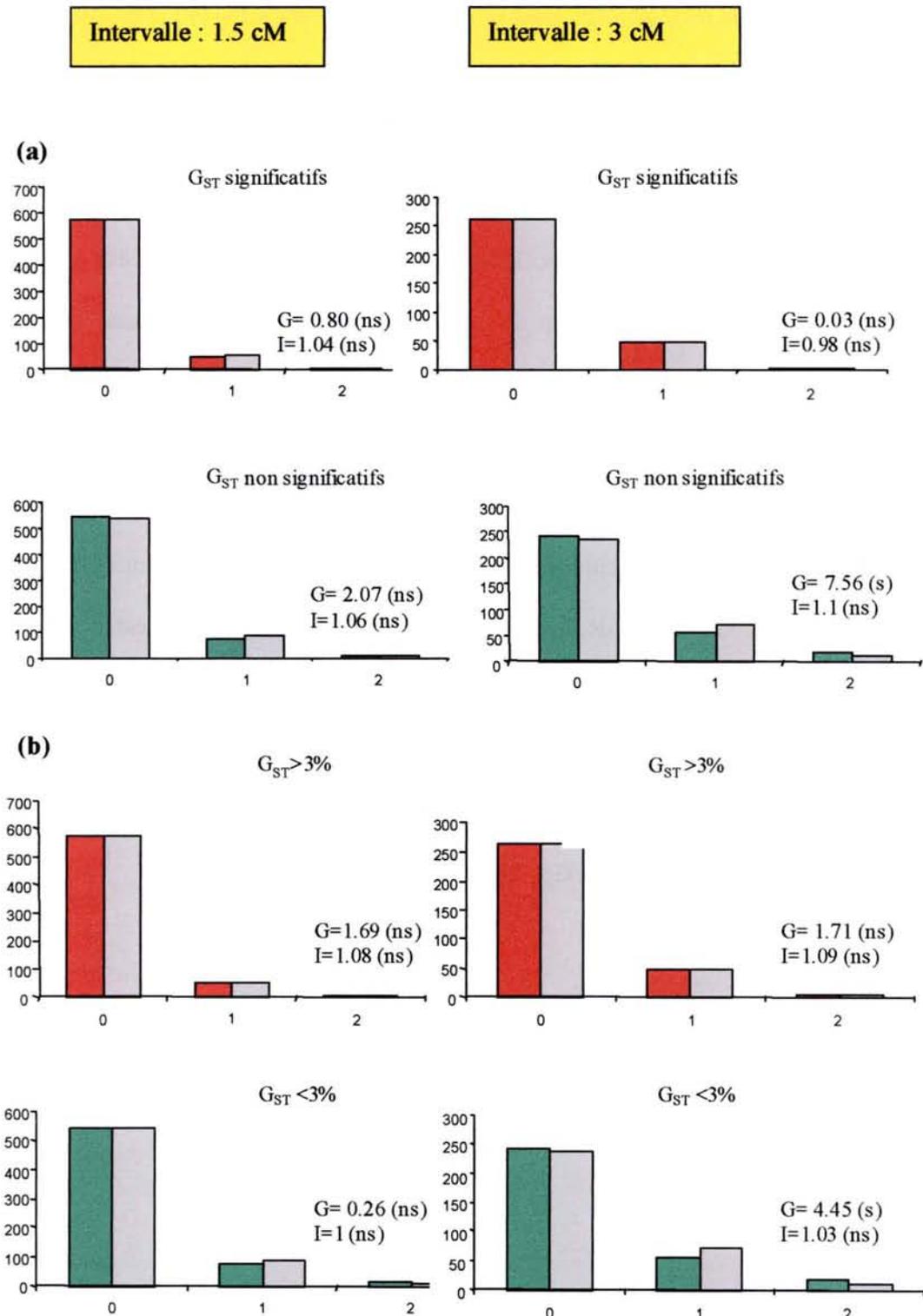
Suite figure 25 :



celle obtenue avec la carte QTL (902cM pour la femelle et 933 cM pour le mâle). De plus, l'ordre des marqueurs entre les cartes est conservé, excepté au niveau de 3 chromosomes où des inversions de position entre paires de marqueurs proches sont observées : entre U19-450 et E-AAC/M-CAT-305 et entre *ssrQpZAG119* et E-AAG/M-CTT-353 sur le chromosome 2, entre E-AAC/M-CCT-357 et *ssrQrZAG73* et entre *ssrQpZAG58* et E-AAC/M-CAA-149 sur le chromosome 5, et entre *ssrQpZAG1/5* et E-AAC/M-CAC-137 sur le chromosome 7.

Parmi les 255 marqueurs cartographiés, 155 possèdent une valeur de  $G_{ST}$ . Leur distribution sur la carte consensus est présentée figure 25. Les marqueurs impliqués dans la différenciation interspécifique sont repartis sur l'ensemble des 12 chromosomes avec cependant des zones où l'on constate des regroupements de  $G_{ST}$  significatifs. Sur le chromosome 2 par exemple, parmi les 21 marqueurs qui présentent une valeur de  $G_{ST}$ , 13 possèdent des  $G_{ST}$  significatifs qui varient entre 1.2% et 9%. On peut aussi noter sur le chromosome 12, la présence de 2 zones fortement impliquées dans la différenciation avec des valeurs de  $G_{ST}$  supérieures à 4%. Un résultat intéressant est la présence à courte distance de  $G_{ST}$  significatifs et élevés pour des types de marqueurs différents. Sur le groupe de liaison 4 par exemple, dans un intervalle de 0.5 cM un marqueur AFLP et un marqueur SNP présentent des  $G_{ST}$  de 8 %. Pareillement sur le groupe de liaison 9 un marqueur SCAR et un marqueur SNP distant de 1.5cM présentent des  $G_{ST}$  de 6 et 7 % respectivement.

**Figure 26** : Test de la structuration le long du génome de la distribution des  $G_{ST}$  significatifs et non significatifs (a) et des  $G_{ST}$  inférieurs et supérieurs à 3% (b) sous l'hypothèse de distribution totalement aléatoire (processus poissonien) (gris).



**Légende :**

G : valeur de la statistique G, I : indice de dispersion. Entre parenthèses est indiqué si la valeur est significative (s) ou pas (ns).  
**rouge** : Nombre d'intervalles qui présente 0, 1 ou au moins 2  $G_{ST}$  significatifs d'une part (a), et supérieurs à 3% d'autre part (b)  
**vert** : Nombre d'intervalles qui présente 0, 1 ou au moins 2  $G_{ST}$  non significatifs d'une part (a), et inférieurs à 3% d'autre part (b),  
**gris** : Distribution attendue sous l'hypothèse nulle de distribution aléatoire des  $G_{ST}$  le long du génome

### D.2.3 DISTRIBUTION DES $G_{ST}$ LE LONG DES GROUPES DE LIAISON

Afin de tester la distribution des valeurs de  $G_{ST}$  le long du génome trois approches complémentaires ont été suivies (voir paragraphe D 1.6) : comparaison à la distribution de Poisson, test de Fisher et autocorrélation.

La comparaison à la distribution de Poisson montre que la distribution n'est pas différente d'une distribution aléatoire (Figure 26). Sous l'hypothèse de répartition au hasard, les 59 marqueurs à  $G_{ST}$  significatif devaient se répartir en 53.7 intervalles célibataires de 1.5 cM (ne comprenant qu'un seul marqueur), 2.52 intervalles à deux marqueurs et 0.08 intervalles à 3 marqueurs. Ces chiffres observés sont respectivement de 51, 4 et 0. Cependant, la comparaison à un processus poissonien nécessite que la distribution des marqueurs soit homogène, notamment sur tous les groupes de liaison. Ceci n'est évidemment pas le cas comme le montre la figure 25 où certains groupes de liaison sont beaucoup plus prospectés que d'autres. Nous avons donc complété cette comparaison par une analyse plus détaillée dans des zones également prospectées (intervalles ayant au moins 2 marqueurs). La deuxième approche a consisté à comparer par un test de Fisher exact la distribution de co-occurrence des  $G_{ST}$  observée par rapport à la distribution attendue sous l'hypothèse nulle de distribution aléatoire des  $G_{ST}$  dans les zones du génome couvertes par les marqueurs. Pour ce faire, seuls les intervalles qui présentaient au moins deux marqueurs avec une valeur de  $G_{ST}$  ont été considérés. A courte distance, la distribution observée de la co-occurrence des  $G_{ST}$  s'écarte de celle attendue sous l'hypothèse de distribution aléatoire, alors qu'à plus grande distance (Tableau 36) le test n'est plus significatif.

La troisième méthode basée sur l'indice de Moran a consisté à calculer dans deux classes d'intervalles (0 à 1.5 cM et 1.5 à 3cM) les corrélations entre marqueurs séparés par une distance inférieure à ces seuils. Les résultats fournis dans le tableau 37 indiquent qu'à courte distance (inférieure à 1.5 cM), les valeurs de  $G_{ST}$  significatives ont tendance à se regrouper ce qui n'est plus vrai au-delà de 1.5cM. En d'autres termes, deux marqueurs situés à moins de 1.5cM ont tendance à avoir des valeurs de  $G_{ST}$

**Tableau 36** : Résultat du test de exact de Fisher réalisé pour tester la distribution non aléatoire des  $G_{ST}$  significatifs à courte distance (1.5cM) et moyenne distance (3cM).

a) Division du génome en intervalles de 1.5 cM

	$G_{ST}$ Non Significatif	$G_{ST}$ Significatif	Somme
$G_{ST}$ Non Significatif	11	2	13
$G_{ST}$ Significatif	2	4 <sup>a</sup>	6
Somme	13	6	19 <sup>b</sup>

Pvalue=0.046

b) Division du génome en intervalles de 3 cM

	$G_{ST}$ Non Significatif	$G_{ST}$ Significatif	Somme
$G_{ST}$ Non Significatif	17	7	24
$G_{ST}$ Significatif	6	5	11
Somme	23	12	35

Pvalue=0.29

Lecture du tableau :

<sup>a</sup> 4 intervalles de 1.5cM contiennent au moins deux  $G_{ST}$  significatifs

<sup>b</sup> 19 intervalles de 1.5cM présentent au moins deux marqueurs avec une valeur de  $G_{ST}$

**Tableau 37** : Résultat des autocorrélations spatiales testées par l'indice de Moran.

	$G_{ST}$ signif	$G_{ST} \geq 3\%$
0 - 1.5cM	<b>0.045</b>	<b>0.028</b>
1.5 - 3cM	0.493	0.860

Les valeurs en gras correspondent à des Pvalue significatives

appartenant à la même catégorie (significative ou non significative) en proportion plus importante que par une association au hasard.

Nous avons aussi effectué le calcul d'autocorrélation en considérant deux autres catégories de marqueurs : ceux pour lesquels les  $G_{ST}$  étaient supérieurs à 3% et ceux pour lesquels ils étaient inférieurs. La valeur de l'Indice de Moran est supérieure à celle attendue sous l'hypothèse d'association au hasard mais uniquement pour les intervalles de petites distances (entre 0 et 1.5 cM) ce qui confirme les précédents résultats.

La structuration des marqueurs impliqués dans la différenciation interspécifique laisse donc supposer des effets d'entraînements entre les marqueurs (effet d'hitch-hiking) sur des distances inférieures à 1.5 cM.

#### **D.2.4 ETUDE D'ASSOCIATION ENTRE MARQUEURS ET CARACTERES**

L'association statistique entre les marqueurs AFLP cartographiés qui portent une valeur de  $G_{ST}$  et les caractères qui discriminent les deux espèces a été testée par ANOVA. Trois types d'analyses ont été effectuées, une au niveau interspécifique, et deux au niveau intraspécifique (population sessile et pédonculé). En considérant que 110 marqueurs ont été testés au niveau de 9 caractères, soit 990 tests par analyse, nous avons fixé le risque de première espèce à un pourcent. Globalement sur les 3 analyses (2970 tests) 107 tests ont présenté des associations significatives entre des caractères et des marqueurs (Tableau 38). L'effet peuplement était significatif pour 105 d'entre-eux ; par contre, l'interaction marqueur peuplement n'est significative que pour 10. Indiquant ainsi que d'un peuplement à l'autre, l'effet des allèles sur la morphologie foliaire des arbres varie en amplitude mais sans changer leur classement. Par exemple, l'allèle qui engendre en moyenne des longs pétioles dans une population n'engendrera pas des petits pétioles dans une autre population.

Au niveau interspécifique, 86 tests sont significatifs qui impliquent seulement 30 marqueurs. Par exemple les marqueurs P-CCA/M-ATA-335, et P-CCA/M-CAA-181 sont associés à 6 caractères. Parmi les 30 marqueurs, 22 ont une valeur de  $G_{ST}$  significative supérieure à 0.023. Et parmi les 9

**Tableau 38** : Association statistique en population naturelle entre les marqueurs AFLP cartographiés et les caractères morphologiques foliaires impliqués dans la différenciation phénotypique entre le chêne sessile et le pédonculé (P-value=0.01).

MARQUEUR	GL <sup>1</sup>	G <sub>ST</sub> <sup>2</sup>	Analyse Intraspécifique		Analyse Interspécifique
			P (N=100) <sup>3</sup>	S(N=100) <sup>4</sup>	PxS (N=200) <sup>5</sup>
E-AAG/M-CAC-314	1	0			SW
P-CAG/M-GGA-153**	1	0.035			PL SW NL NV PU
E-AAC/M-CAA-74	1	0		LL PL	
E-AAC/M-CAC-109	2	0.005	SW		LL SW
E-AAC/M-CAC-54	2	0.035			BS PL PU SW NV
E-AAC/M-CAA-294	2	0.032			BS NL
P-CAG/M-GGA-100**	2	0			NV
P-CAG/M-GGA-93**	2	0.050			NV PU
E-AAG/M-CAC-190	2	0		NV	
E-AAC/M-CAT-61	3	0		LL LW WP	LL LW WP
E-AAC/M-CAA-122	3	0.025			NL
E-AAG/M-CAC-235	3	0.002	NL		
E-AAC/M-CCT-210	4	0.023			NL PL PU NV
P-CAG/M-CAA-299	4	0.025			PL
P-CCA/M-CAA-79**	4	0.004			SW BS
E-AAC/M-CAC-158	4	0	LW WP		
E-AAC/M-CTT-57	4	0	BS		
E-AAC/M-CCT-182	4	0.087			PL PU NV
E-AAC/M-CAT-171	4	0.041			PL PU NV
E-AAC/M-CAC-86	5	0.073			BS PL PU NV
E-AAC/M-CAC-66	5	0.128			BS NL PL PU SW NV
E-AAC/M-CTT-193	5	0.007	NL		NL PL
E-AAC/M-CCT-69	5	0		BS	
E-AAG/M-CAC-332	5	0		SW	
P-CCA/M-ATA-131	7	0.031			PU
E-AAG/M-CAC-174	8	0.046			BS PL SW NV
E-AAC/M-CAA-66	8	0.043			NV
E-AAC/M-CTT-350	8	0.060		WP	PL PU NV
P-CAG/M-CAA-400	9	0.063			BS NL PL PU NV
E-AAC/M-CTT-253	9	0.005			PU
E-AAC/M-CTT-262	9	0		PU	
E-AAC/M-CAC-160	10	0		PL	LW SW
E-AAG/M-CTA-147	10	0.055			LW PL PU SW NV
E-AAC/M-CAT-190	11	0.017			PL
P-CAG/M-AGC-152	11	0.029			PL PU NV
E-AAC/M-CAA-82	11	0.024		SW	PU
E-AAC/M-CAT-200	11	0		NV	
P-CAG/M-AGC-308	11	0.004	WP		
P-CCA/M-ATA-335	12	0.186			BS NL PL PU SW NV
P-CCA/M-CAA-181**	12	0.128		PU	PL SW NL NV BS PU
Test significatifs (p<0.01)			7	14	86

Modèle d'ANOVA :  $Y_{ijk} = \mu + M_i + P_j + M_j * P_j + E_{ijk}$

où  $Y_{ijk}$  est la valeur du caractère mesuré sur l'individu  $k$  au niveau du marqueur  $i$  dans le peuplement  $j$ ,  $\mu$  est la moyenne générale du caractère phénotypique,  $M_i$  est l'effet du marqueur,  $P_j$  est l'effet peuplement,  $M_i * P_j$  est l'interaction marqueur peuplement et  $E_{ijk}$  l'erreur résiduelle.

\*\* marqueurs analysés par S. Mariette

**En gras** : interaction significative entre marqueur et peuplement ( $M_i * P_j$ )

*En italique* : Effet non significatif du peuplement ( $P_j$ )

<sup>1</sup> groupe de liaison

<sup>2</sup> valeurs des indices de différenciation interspécifiques ( $G_{ST}$ )

<sup>3</sup> Caractères en association avec un marqueur dans la population sessile

<sup>4</sup> Caractères en association avec un marqueur dans la population pédonculé

<sup>5</sup> Caractères en association avec un marqueur dans la population globale (sessile et pédonculé)

longueur du limbe (LL), longueur de pétiole (PL), largeur maximale du lobe (LW), largeur du sinus (SW), longueur de la base du limbe au niveau de sa largeur maximale (WP), nombre de lobes (NL), nombre de veines intercalaires (NV), forme de l'oreillette (BS), pubescence dans la partie centrale du limbe (PU)

caractères, ce sont surtout les plus discriminants (PL, NV, SW, BS, PU) qui sont en association avec les marqueurs les plus discriminants. Par opposition, les caractères moins impliqués dans la différenciation présentent seulement 6 associations significatives avec les marqueurs (WP : 1, LL : 2, LW : 3) dont 5 impliquent des marqueurs à faibles  $G_{ST}$ . Ainsi le pouvoir discriminant des marqueurs AFLP à forts  $G_{ST}$  est confirmé par l'association avec des caractères morphologiques discriminants. Mais ces liaisons n'informent en aucun cas sur la liaison physique. Au niveau intraspécifique, le nombre de tests significatifs est plus faible, une des raisons pourrait être l'effectif plus faible. 14 associations significatives ont été obtenues pour la population de chêne sessile impliquant 11 marqueurs et 7 pour la population de chêne pédonculé impliquant 6 marqueurs. Les associations impliquent tous les caractères, qu'ils soient fortement ou peu discriminants. Par contre, ce sont des marqueurs à faibles  $G_{ST}$  qui sont principalement en association : sur les 17 marqueurs impliqués, 14 présentent des valeurs de  $G_{ST}$  inférieures à 1%.

Aucune association n'est significative simultanément dans la population sessile et la population pédonculé. Par contre 6 associations sont conjointement significatives dans l'analyse interspécifique et l'une ou l'autre des populations intraspécifiques. On peut par exemple noter l'association entre le marqueur E-AAC/M-CAT-61 et les caractères LL, WP, et LW qui est significative dans la population interspécifique et la population sessile. Dans l'hypothèse d'absence de structuration dans la population sessile, ces caractères pourraient être en déséquilibre de liaison avec ce marqueur. La liaison physique pourra être confirmée en étudiant les colocalisations avec les QTL.

S.C.D. - U.H.P. NANCY 1  
BIBLIOTHÈQUE DES SCIENCES  
Rue du Jardin Botanique - BP 11  
54601 VILLERS-LES-NANCY Cédex

## Chapitre E

# COLOCALISATIONS ENTRE $G_{ST}$ ET QTL

## CHAPITRE E COLOCALISATIONS ENTRE $G_{ST}$ ET QTL

La colocalisation entre les QTL impliqués dans des caractères discriminants et les  $G_{ST}$  significatifs a été testée en considérant l'ensemble des QTL détectés au seuil de 5% au niveau du chromosome. Cependant, lorsque plusieurs QTL impliqués dans le même caractère se situaient dans le même intervalle un seul QTL a été considéré. Pour repositionner les QTL nous avons considéré l'intervalle de la position du QTL qui correspond à la distance entre P1, la position au niveau du LOD maximal et P2 la position obtenue après 1000 ré-échantillonnages par bootstrap. Nous avons ensuite considéré sur les cartes jalons mâle et femelle la distance qui sépare chaque extrémité des intervalles avec les marqueurs les plus proches afin de la reporter sur la carte consensus.

Les QTL ont été classés en deux catégories. D'une part les QTL impliqués dans les caractères discriminants et d'autre part les QTL impliqués dans des caractères non discriminants (Voir paragraphe C.2.4).

Pour tester les colocalisations entre les QTL impliqués dans des caractères discriminants et  $G_{ST}$  significatifs le génome a de nouveau été divisé en N intervalles de 5 cM d'une part puis de 10 cM d'autre part. A la différence des tests réalisés pour les colocalisations entre QTL, ont été exclus tous les intervalles qui ne contiennent pas de  $G_{ST}$ . Ainsi certains QTL ont du être écartés de l'analyse. Deux types de test ont été réalisés. Le premier test basé sur la méthode de Feltus *et al.* (2003) décrite dans le paragraphe C.1.3.3 et déjà utilisé pour les colocalisations entre QTL consiste à calculer la probabilité d'observer dans l'intervalle d'un QTL impliqué dans la différenciation interspécifique un  $G_{ST}$  significatif.

Le deuxième test a consisté à comparer la distribution des QTL et des  $G_{ST}$  avec celle attendue sous l'hypothèse de distribution aléatoire des données. Pour cela un tableau de contingence a été réalisé en considérant les 4 classes suivantes :

**Tableau 39** : Test de colocalisation entre les  $G_{ST}$  significatifs et les QTL selon la méthode de Feltus (2003).

<i>Caractères</i>	<i>N</i>	<i>P</i>	<i>G</i>	<i>C</i>	<i>Pvalue</i>
Morpho1	100	24	47	14	0.084
$\delta^{13}C$	100	7	47	3	0.297
Crois	100	6	47	3	0.319
Global D	100	35	47	20	0.055
Morpho2	100	7	47	0	-
deb	100	12	47	7	0.172
bout	100	8	47	4	0.281
Global ND	100	24	47	10	0.156

*N* : le nombre d'intervalle qui contient une valeur de  $G_{ST}$

*P* : le nombre d'intervalle qui contient un QTL

*G* : le nombre d'intervalle qui contient au moins un  $G_{ST}$  significatif

*C* : le nombre d'intervalle où l'on observe la présence d'un QTL et d'au moins un  $G_{ST}$  significatif

Morpho1 : Caractères morphologiques discriminants (PL, SW, NL, NV, BS, Puv, Pul)

$\delta^{13}C$  : Discrimination isotopique du carbone

Crois : Caractères impliqués dans la croissance (Luc, Nuc, Ht)

Global D : Ensemble des caractères discriminants (Morpho1,  $\delta^{13}C$ , Crois)

Morpho2 : Caractères morphologiques non discriminants (LL, LW, WP)

deb : Débourrement

bout : Aptitude au bouturage

Global ND : Ensemble des caractères non discriminants (Morpho2, Deb, Bout).

	$G_{ST}$ Significatif	$G_{ST}$ non Significatif
QTL présent		
QTL absent		

Un test de Fisher exact a ensuite permis de voir s'il y avait co-occurrence ou non des QTL avec les  $G_{ST}$  significatifs dans les mêmes régions du génome

### E.1.1 COLOCALISATION ENTRE QTL ET MARQUEURS IMPLIQUES DANS LA DIFFERENCIATION GENETIQUE INTERSPECIFIQUE

Ces études de colocalisations sont faites sur la carte consensus, dont les principes de construction ont été évoqués dans le paragraphe D.1.5. La distribution des QTL le long de la carte de distribution des valeurs de  $G_{ST}$  est donnée Figure 27. Ont aussi été mentionnés sur la figure, les marqueurs qui présentent une association significative avec un caractère de morphologie foliaire en population naturelle (tableau 38) et qui sont à proximité d'un QTL.

Pour tester si la présence des  $G_{ST}$  significatifs à l'intérieur des intervalles des QTL est due au hasard ou pas, nous avons calculé la probabilité d'observer  $C$   $G_{ST}$  significatifs dans l'intervalle des QTL sous l'hypothèse nulle de distribution aléatoire selon la méthode de Feltus *et al.*(2003) (paragraphe C.1.3.3). La méthode consiste à calculer la probabilité de co-occurrence de QTL (impliqués dans la différenciation phénotypique entre les deux espèces) et les marqueurs (à  $G_{ST}$  significatif) dans des intervalles de 5cM d'une part et 10 cM d'autre part.. L'analyse a tout d'abord été réalisée pour chaque type de caractère puis globalement en groupant les caractères impliqués dans la différenciation d'une part, et ceux impliqués dans des caractères non discriminants, d'autre part. Les résultats sont fournis dans le tableau 39. Aucun test n'est significatif au seuil de 5%, toutefois on peut noter une tendance au regroupement entre les QTL impliqués dans des caractères discriminants et les  $G_{ST}$  significatifs puisque le test est à la limite du risque de première espèce (pvalue=0.055).

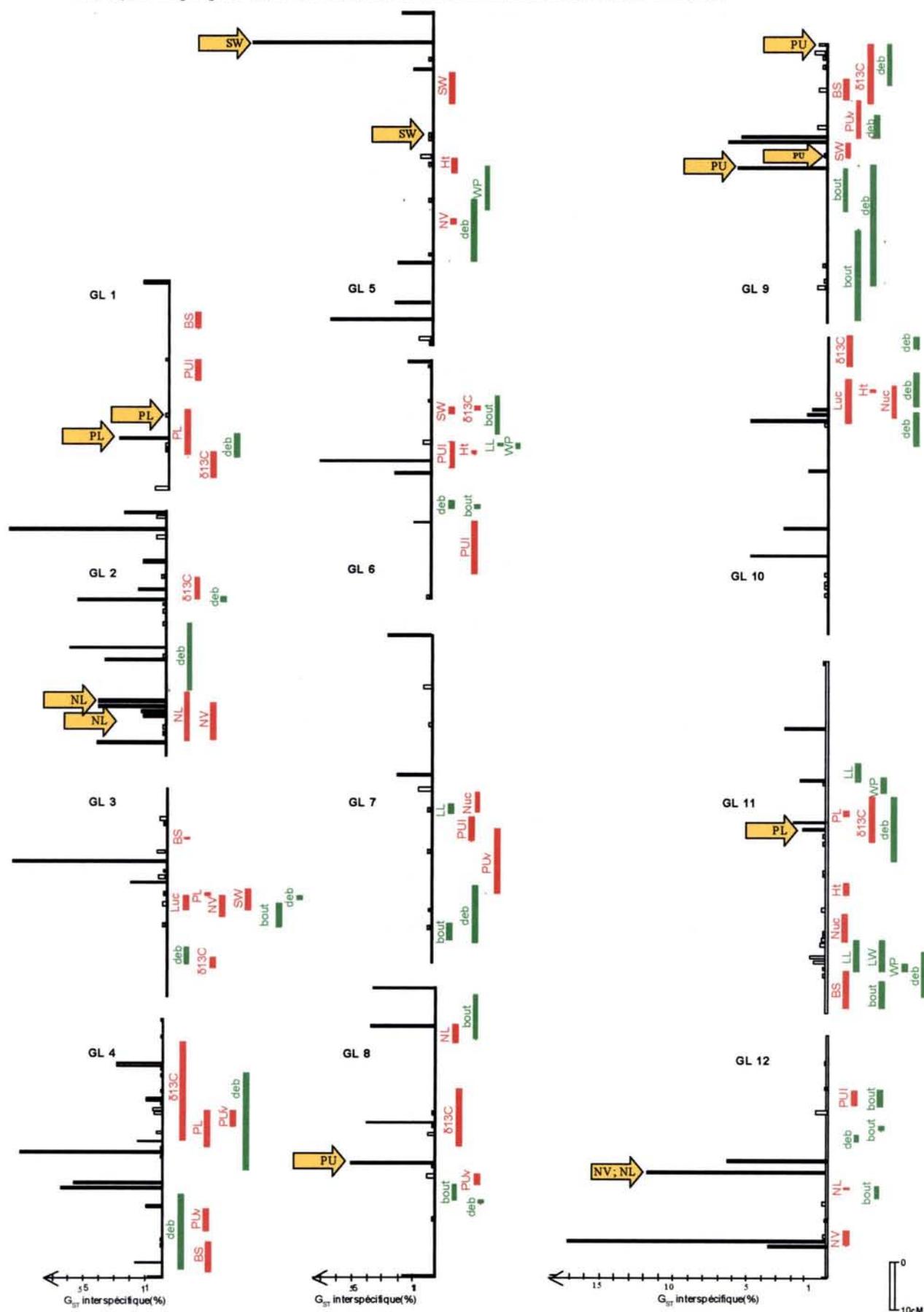
**Tableau 40** : Test de colocalisation entre les  $G_{ST}$  significatifs et les QTL impliqués dans les caractères discriminants par un test de Fisher exact.

	$G_{ST}$ Significatif	$G_{ST}$ Non Significatif	Somme
QTL présent	21	13	34
QTL absent	29	37	66
Somme	50	50	100

Pvalue=0.07

Une deuxième façon de tester si le regroupement des  $G_{ST}$  significatifs dans l'intervalle des QTL est dû au hasard ou pas, a été de comparer par un test exact de Fisher la distribution observée des  $G_{ST}$  dans les intervalles des QTL avec la distribution attendue sous l'hypothèse nulle de distribution aléatoire. Pour cela nous avons subdivisé le génome en N intervalles de 5 cM en ne considérant que les intervalles où au moins une valeur de  $G_{ST}$  était présente. Les résultats ne sont pas significatifs (tableau 40) indiquant ainsi l'absence de regroupement des  $G_{ST}$  significatifs dans l'intervalle des QTL impliqués dans la différenciation. Une des explications pourrait être liée au type de famille utilisé pour la recherche de QTL. En effet, il est possible que les QTL les plus discriminants ne soient pas polymorphes au marqueur dans la descendance intraspécifique en raison de la fixation d'allèles différents dans les deux espèces. Pour voir ségréger ces QTL il faudrait utiliser une descendance interspécifique de deuxième génération. Une étude similaire a été menée chez le maïs, dans une famille  $F_2$  interspécifique, dont le but était d'étudier les colocalisations entre des coefficients de différenciation interspécifique ( $F_{ST}$ ) et des QTL impliqués dans des caractères qui différencient le maïs et son espèce sauvage le téosinte (Matsuoka et Vigouroux, résultats non publiés, voir [http://plantgenome.sdsc.edu/AwardeesMeeting/poster\\_Doebley.pdf](http://plantgenome.sdsc.edu/AwardeesMeeting/poster_Doebley.pdf)). Les résultats montrent l'absence de fortes relations entre la localisation des  $F_{ST}$  élevés et les QTL impliqués dans la domestication. Pour l'ensemble du génome, l'effet des QTL de domestication sur la valeur des  $F_{ST}$  est significatif mais explique moins de 5 % de leur variabilité. La sélection n'aurait donc pas entraîné une augmentation de la différenciation interspécifique au niveau des régions qui présentent des gènes de domestication.

**Figure 27** : Colocalisations entre les marqueurs impliqués dans la différenciation interspécifique et les QTL impliqués dans les caractères discriminants et non discriminants.



**Légende :**

Les barres représentent la valeur de  $G_{ST}$  en pourcentage

Barre noire:  $G_{ST}$  significatif

Barre blanche :  $G_{ST}$  non significatif



QTL impliqué dans un caractère discriminant (PL, NL, NV, BS, SW, PUI, PUv,  $\delta 13c$ , Ht,  $L_{UC}$ ,  $N_{UC}$ )

QTL impliqué dans un caractère non discriminant (LL, LW, WP, deb, bout)



Le caractère indiqué dans la flèche est en association en population naturelle avec le marqueur pointé

## Chapitre F

# CONCLUSION GENERALE

## CHAPITRE F CONCLUSION GENERALE

### F.1 RAPPELS DES PRINCIPAUX RESULTATS

L'étude de l'organisation moléculaire de la différenciation génétique entre le chêne sessile et le chêne pédonculé a été menée en conduisant deux types d'approches complémentaires. La première a consisté à rechercher des QTL impliqués dans des caractères phénotypiques qui différencient les deux espèces dans une population  $F_1$  intraspécifique de chêne pédonculé. Parallèlement des QTL ont été détectés pour des caractères non impliqués dans la différenciation afin de pouvoir comparer le déterminisme génétique ainsi que la distribution sur le génome des deux types de QTL. En moyenne l'effet des QTL impliqués dans les caractères discriminants est supérieur à l'effet des QTL impliqués dans les caractères non discriminants. Il faut cependant rester vigilant sur la comparaison de ces résultats qui sont très dépendants de l'effectif pris en compte dans les analyses (Beavis, 1995). La multiplication végétative des clones a cependant permis de préciser la proportion de la variance phénotypique et génétique (additive) expliquée par l'effet clone, en faisant des hypothèses sur la variance de dominance. La multiplication végétative des clones a enfin permis de comparer les QTL détectés dans différents dispositifs et de mettre en évidence une bonne stabilité de certains d'entre eux dans l'espace et le temps. Enfin dans une dernière analyse, nous avons montré que les QTL impliqués dans les caractères discriminants tendent à se regrouper sur les carte mâle et femelle alors que pour les caractères non discriminants, les QTL tendent à se regrouper seulement sur la carte femelle.

La deuxième approche a consisté à prospecter le génome de manière systématique à l'aide des différentes techniques de marquage disponible et d'estimer pour chaque marqueur un coefficient de différenciation interspécifique ( $G_{ST}$ ). La distribution des valeurs de  $G_{ST}$  pour les marqueurs AFLP suit une courbe en L indiquant ainsi la présence sur le génome d'un grand nombre de marqueurs non impliqués dans la différenciation, associés à quelques marqueurs discriminants mais non diagnostiques. Les combinaisons Pst-Mse fournissent en moyenne des  $G_{ST}$  plus élevés que les

marqueurs issus des combinaisons Eco-Mse, ce qui pourrait s'expliquer par une tendance des enzymes Pst à couper préférentiellement dans les parties codantes du génome. Les régions impliquées dans la différenciation sont réparties sur l'ensemble du génome avec néanmoins des groupes de liaison qui présentent des hot spot de différenciation (GL2). Un résultat original est la mise en évidence d'un regroupement des  $G_{ST}$  significatifs dans des intervalles de courte distance (entre 0 et 1.5 cM) qui laisse supposer la présence d'un effet d'entraînement des gènes sélectionnés sur les fréquences alléliques des marqueurs voisins.

Enfin, la comparaison de la distribution des  $G_{ST}$  avec celle des QTL tend à indiquer la présence d'un regroupement entre les  $G_{ST}$  significatifs et les QTL discriminants mais les tests ne sont pas significatifs ( $p=0.055$ ).

De cette comparaison de cartographie de QTL et de  $G_{ST}$  on peut donc conclure :

- (1) que les régions du génome qui sont impliquées dans la différenciation interspécifique sont distribuées sur l'ensemble des groupes de liaison, et que cette distribution au vu de nos résultats est proche d'une répartition aléatoire.
- (2) que localement au niveau d'une région impliquée dans la différenciation entre *Q. petraea* et *Q. robur*, un effet d'entraînement est perceptible autour d'une distance de 1.5 cM. Il s'agit là d'un effet d'entraînement moyen sur l'ensemble des régions. Nos résultats suggèrent cependant une variation de la distance d'entraînement en fonction sans doute de variations des taux de recombinaisons.
- (3) que certains hot spots de différenciation ont clairement pu être identifiés par des approches indépendantes associant QTL,  $G_{ST}$  (et souvent  $G_{ST}$  mesurés à partir de marqueurs différents, sur des populations différentes) et déséquilibre en populations naturelles. On peut par exemple noter les groupes de liaisons 1, 2, 8, 9, 11, 12 (Figure 27).

## **F.2 PROBLEMES RENCONTRES ET REPONSES PROPOSEES**

### **F.2.1 PUISSANCE DE DETECTION DES QTL**

Malgré la multiplication végétative des descendants, la puissance de détection de QTL est limitée par l'effectif de la population. Pour améliorer la recherche de régions discriminantes par l'approche QTL il serait nécessaire d'augmenter l'effectif de la descendance en vue de préciser l'effet des QTL et de pouvoir détecter ceux ayant des effets plus faibles. A l'heure actuelle la population de cartographie intraspécifique est composée de 278 individus mais sera doublée dans les années à venir. La position et l'effet des QTL pourront ainsi être précisés et la détection de QTL à effets plus faibles pourra être possible.

### **F.2.2 RECHERCHE DE NOUVEAUX QTL ET VALIDATION DES QTL DEJA DETECTES**

Comme nous l'avons évoqué en conclusion du chapitre D la détection de QTL n'est pas une recherche exhaustive des régions impliquées dans la différenciation interspécifique. La principale limite de notre étude réside dans le dispositif expérimental composé d'une famille intraspécifique. En effet, comme nous l'avons déjà mentionné, les QTL les plus discriminants ne sont pas forcément polymorphes entre les deux parents. On ne peut pas exclure le risque que certains d'entre eux soient chez les deux espèces et dans ce cas il seraient passés totalement inaperçus dans notre analyse. Pour voir ségréger ces allèles la population idéale serait une famille interspécifique de deuxième génération. Or à l'heure actuelle cette famille n'est pas disponible.

Il est aussi possible que certains QTL n'aient pas été détectés car non polymorphes dans les deux parents choisis pour construire la population de cartographie. Pour remédier à ce problème il sera possible d'envisager la détection de QTL dans d'autres populations de cartographie qui sont en cours de construction. Par ailleurs cela permettra de valider les QTL déjà détectés.

Cinq populations de cartographie de première génération supplémentaires seront disponibles : une deuxième population intraspécifique de chêne pédonculé qui correspond à la réciproque de celle utilisée dans notre étude (l'arbre mâle est utilisé comme femelle et la femelle comme mâle). Deux populations intraspécifiques de chênes sessiles (les deux réciproques), et deux populations interspécifiques (les deux réciproques).

A l'heure actuelle seule la carte interspécifique est en cours de construction basée sur le typage de 83 individus, ce qui n'est pas encore suffisant pour détecter des QTL.

### **F.2.3 L'APPROCHE GENES CANDIDATS : UNE NOUVELLE STRATEGIE POUR DETECTER DES REGIONS DISCRIMINANTES**

Deux types d'approches peuvent être suivies pour rechercher des gènes impliqués dans des caractères qui différencient les deux espèces. La première consiste à rechercher dans les bases de données des séquences d'ADN de gènes d'intérêt déjà séquencés chez d'autres espèces.

C'est par exemple le cas du gène *GLABROUS1* d'*Arabidopsis*, impliqué dans la formation des trichomes (Oppenheimer, 1991). Compte tenu de la forte différenciation entre le chêne sessile et le chêne pédonculé au niveau de la pilosité des feuilles, ce gène a été amplifié chez le chêne. L'amplification a permis d'obtenir un fragment, qui après séquençage ne correspondait pas à la séquence de *GLABROUS 1* mais à la séquence d'un gène *Myb* (*Myb 36*) impliqué dans la formation des trichomes chez le coton. Il serait maintenant intéressant de rechercher du polymorphisme dans ces séquences pour pouvoir cartographier ce gène sur la carte du chêne et regarder s'il colocalise avec des QTL.

La deuxième approche pour rechercher des gènes candidats est l'analyse d'expression différentielle de gènes entre le chêne sessile et le chêne pédonculé. De telles expérimentations sont en cours. Elles consistent à étudier les transcrits chez les deux espèces soumises à des stress, auxquels les espèces répondent de manière différentielle. C'est notamment le cas pour le facteur hydrique (en excès : stress

anoxique, ou en manque : sécheresse). Une étude, en cours au laboratoire, a déjà permis d'identifier 130 transcrits différenciellement exprimés entre les deux espèces en condition de stress anoxique.

Une étude similaire est menée par une équipe autrichienne (thèse d'Ilga Porth, dans le cadre du projet OAKFLOW) dans le but d'amplifier des gènes différenciellement exprimés en condition de stress osmotique. Certains de ces gènes (marqueur SNP, voir tableau 32) ont d'ailleurs été cartographiés et utilisés dans cette thèse pour étudier la distribution des  $G_{ST}$  le long du génome (Porth *et al.*, 2002).

Enfin des gènes candidats à l'efficacité d'utilisation de l'eau vont être recherchés en adoptant la méthode de "Bulked segregant analysis" (Michelmore *et al.*, 1991). La détection de QTL a permis de mettre en évidence une région fortement impliquée dans ce caractère sur le groupe de liaison 11 ce qui suggère la présence d'un gène à effet majeur. Pour rechercher ce gène deux groupes d'individus extrêmes vont être considérés : un groupe constitué des individus qui présentent les plus fortes efficacités d'utilisation de l'eau et un deuxième groupe constitué des individus qui présentent les efficacités les plus faibles.

Nous allons donc avoir dans les années qui viennent toute une batterie de gènes candidats impliqués dans des caractères écophysiologiques qui différencient les deux espèces qui permettront de compléter l'étude de QTL. Parallèlement ces gènes pourront être analysés en population naturelle en vue de saturer la carte génétique en valeur de  $G_{ST}$ .

#### **F.2.4 AUGMENTER LA DENSITE DES $G_{ST}$ SUR LA CARTE GENETIQUE**

Les résultats obtenus au cours de cette thèse indiquent qu'à faible distance les  $G_{ST}$  significatifs ont tendance à se regrouper.

Dans les études à venir il sera important de densifier ces régions en marqueurs pour préciser l'effet d'entraînement. Cependant, la tâche ne sera pas facile compte tenu de l'effectif de la population de cartographie. En effet, avec un effectif de 300 individus, l'erreur standard de la position entre deux marqueurs distants de 1cM est de 0.5 cM. Néanmoins, tous les gènes candidats qui présentent du

polymorphisme dans la population de cartographie pourront être étudiés en population naturelle pour augmenter la densité des  $G_{ST}$  sur la carte. Si ces gènes ne permettent pas d'obtenir un nombre de marqueurs suffisant, il sera alors possible de cartographier d'autres AFLP qui procurent rapidement un grand nombre de marqueurs. Il serait pour cela préférable de choisir des combinaisons de type Pst-Mse qui fournissent des  $G_{ST}$  en moyenne supérieurs aux combinaisons Eco-Mse.

### **F.2.5 RECHERCHE DE GENES LETAUX**

L'incompatibilité entre certains allèles a pu être mise en évidence lors de la réalisation de la carte interspécifique par un fort taux de marqueurs présentant une distorsion de ségrégation (26% des marqueurs)(données non publiées). La mortalité de certains hybrides a par ailleurs été observée lors de la réalisation des croisements contrôlés dont le principal symptôme de dégénérescence est une déficience en chlorophylle. A chaque fois qu'un hybride déficient chlorophyllien se présente son ADN est extrait avant qu'il meure. Ainsi lorsque le nombre d'individus sera assez élevé, il sera possible d'entreprendre une détection de QTL pour la déficience en chlorophylle et vérifier si ces QTL tombent dans des régions distordues du génome.

### **F.2.6 LE PARADOXE DIFFERENCIATION MOLECULAIRE/ DIFFERENCIATION PHENOTYPIQUE DES DEUX ESPECES**

Les précédentes investigations sur la différenciation moléculaire entre les deux espèces de chênes ont mis en évidence le fait, qu'en proportion, le nombre de loci impliqués dans la différenciation reste faible (courbe en L des valeurs de  $G_{ST}$  interspécifique) et que les loci ne sont jamais fixés à des allèles différents chez les deux espèces. Ces observations sont clairement confirmées ici. La différenciation interspécifique ne serait donc pas due à la fixation d'allèles différents entre les deux chênes. Nos résultats montrent que les différences interspécifiques résultent plutôt de la somme de petites différences sur plusieurs loci. En outre, nos observations suggèrent à l'évidence, que même en proportion faible, ces loci sont distribués sur l'ensemble du génome, sur tous les groupes de liaison.

Nous sommes ici aux antipodes des notions de "linkats" ou groupes de gènes coadaptés caractérisés par de faibles taux de recombinaison (Demarly, 1977), à moins de considérer que les linkats sont de taille inférieure à 1.5 cM. Nos résultats suggèrent au contraire que ce sont des associations alléliques sur des loci *à priori* non liés qui sont à la base des différences interspécifiques. Cette observation conduit alors à une double conclusion : tout d'abord ce sont des effets multilocus qui sont à la base des différences interspécifiques et ces associations multilocus doivent être induites par des forces évolutives suffisamment fortes pour ne pas être gommées par les flux de gènes. En ce qui concerne les effets multilocus, on peut invoquer soit les effets épistatiques, soit les déséquilibres de liaison. Des études théoriques récentes portant sur l'évolution comparée de la variabilité de caractères phénotypiques et de la diversité aux QTL contrôlant ces caractères phénotypiques, ont montré que de très importantes différences phénotypiques inter-populations pouvaient coexister avec de petites différences de fréquences alléliques à ces loci (Latta, 1998 ; Le Corre et Kremer, 2003). L'apparent paradoxe de cette situation réside dans la nature multilocus du caractère phénotypique : de la même manière qu'une même valeur phénotypique peut être atteinte par de nombreuses combinaisons alléliques, de très faibles différences de fréquences alléliques peuvent conduire à de fortes différences phénotypiques. Le Corre et Kremer (2003) montrent que cette discordance diversité moléculaire - diversité phénotypique s'explique par le rapport des déséquilibres inter et intra-population entre les loci impliqués dans le contrôle du caractère phénotypique, et que ces déséquilibres sont très rapidement (en nombre de générations) induits par la sélection naturelle divergente entre les populations. Nous pouvons très aisément transposer ce raisonnement au cas interspécifique qui nous concerne.

Les deux espèces sont connues pour leur préférence écologique à certains milieux. Sous l'effet de la sélection divergente (milieux plus secs pour le chêne sessile, milieux plus humides pour le chêne pédonculé) portant sur des caractères d'adaptation, des phénotypes particuliers vont être favorisés pour les deux pôles constitués du chêne sessile et du chêne pédonculé. Cette sélection va entraîner des associations particulières d'allèles aux différents QTL, sans pour autant aboutir à de fortes différences

alléliques à ces QTL. Rappelons à ce sujet que plus le nombre de QTL est grand, moins il faut de différences entre fréquences alléliques pour rendre compte de différences phénotypiques importantes (Le Corre et Kremer, 2003). Les résultats expérimentaux de cette thèse suggèrent en plus qu'il n'y a pas besoin de liaison physique (par déficit de recombinaison) pour renforcer ces associations. En résumé, ce scénario s'adapte assez facilement à la vision de Wu (2001) pour qui la différenciation adaptative est à la base de la différenciation spécifique et qui "subordonne" l'isolement reproducteur à la différenciation adaptative. Il corrobore aussi les données en biologie évolutive (Petit et al, 1997; Kremer et al, 2002) qui suggèrent que le chêne sessile résulte en fait de backcross successifs et unidirectionnels avec le chêne pédonculé, et ceci de manière indépendante dans différentes forêts. Dans le cas de différences interspécifiques distribuées dans de nombreuses régions génomiques, il faudrait de nombreuses générations pour "régénérer" le chêne sessile dans le chêne pédonculé. Seule la présence d'une forte pression de sélection favorisant les associations particulières peut accélérer l'introgression et aboutir en un faible nombre de générations au maintien des deux espèces.

Au delà du couple chêne sessile-chêne pédonculé, il existe d'autres espèces de chênes qui peuvent s'hybrider et qui maintiennent cependant des différences morphologiques et écologiques nettes. C'est par exemple le cas couple chêne sessile/chêne pubescent ou du couple chêne tauzin /chêne pédonculé. Pour chaque couple, les deux espèces se caractérisent par des exigences écologiques particulières. Il sera important dans des études futures de mener de front l'étude de la différenciation interspécifique entre ces différents chênes pour vérifier si les régions impliquées dans la différenciation génétique interspécifique sont toujours impliquées dans la différenciation générale ou plutôt spécifique du couple considéré.

REFERENCES  
BIBLIOGRAPHIQUES

- Aas G., 1993. Taxonomical impact of morphological variation in *Quercus robur* and *Q. petraea* : a contribution to the hybrid controversy. *Ann. Sci. For.* **50**, Suppl. 1 : 107s-113s.
- Allard R., 1956. Formulas and tables to facilitate the calculation of recombination values in heredity. *Hilgardia* **24** : 235-278
- Axelrod D.I., 1983. Biogeography of oaks in the Arcto-Tertiary province. *Ann. Missouri Bot. Gard.* **70** : 629-657.
- 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 in the northwest of France. *Silvae Genetica* **44** : 1-10.
- Bacilieri R., Ducouso A., Petit R.J., Kremer A., 1996. Mating system and asymmetric hybridization in a mixed stand of European oaks. *Evolution* **50** : 900-908.
- Bakker E., Van Dam B., van Heck H.J., Jacobsen E., 2000. A discrimination between *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. based on species-indicative AFLP markers. In : Towards molecular tools for management of oak forests. PhD. Pp 7-114
- Barreneche T., Bahrman N., Kremer A., 1996. Two dimensional gel electrophoresis confirms the low level of genetic differentiation between *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. *Forest Genetics* **3** : 89-92.
- Barreneche T., Bodénès C., Lexer C., Trontin J.F., Fluch S., Streiff R., Plomion C., Roussel G., Steinkellner H., Burg K., Favre J.M., Glössl 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. *Theor Appl Genet.* **97** : 1090-1103.
- Beavis W.D., 1995. The power and deceit of QTL experiments : lessons from comparative studies. In : Proceedings of the 49th annual corn and sorghum industry research conference, Chicago, IL. pp. 250-266
- Becker M., Lévy G., 1982. Le dépérissement du chêne en forêt de Tronçais. Les causes écologiques. *Ann. Sci. For.* **39** : 439-444.
- Becker M., Lévy G., 1990. Le point sur l'écologie comparée du chêne sessile et du chêne pédonculé. *Revue Forestière Française* **42** : 148-154.
- Billington H.L., Pelham J., 1991. Genetic variation in the date of budburst in Scottish birch populations : implications for climate change. *Functional Ecology* **5** : 403-409.
- Blears M.J., De Grandis S.A., Lee H., Trevors J.T., 1998. Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. *J. Industrial Microbiol. Biotech.* **21** : 99-114

- Bodénès C., 1996c. Différenciation moléculaire entre chêne sessile *Quercus petraea* (Matt) Liebl. et chêne pédonculé *Quercus robur*. Thèse de Doctorat. Université Bordeaux I. 131p.
- Bodénès C., Joandet S., Laigret F., Kremer A., 1996a. Detection of genomic regions differentiating two closely related oak species *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. *Heredity* **78** : 433-444.
- Bodénès C., Labbé T., Pradère S., Kremer A., 1997. General vs. local differentiation between two closely related white oak species. *Molecular Ecology* **6** :713-724.
- Bodénès C., Laigret F., Kremer A., 1996b. Inheritance and molecular variations of PCR-SSCP fragments in pedunculate oak (*Quercus robur* L.). *Theor Appl Genet* **93** : 348-354.
- Box G.E.P., Cox D.R., 1964. An analysis of transformation. *J Amer Stat Assn.* **39** : 357-365.
- Bradshaw H.D, Foster G.S., 1992. Marker-aided selection and propagation systems in trees: advantages of cloning for studying quantitative inheritance. *Can. J. For. Res.* **22** : 1044-1049
- Bradshaw H.D., Otto K., Frewen B., McKay J., Schemske D., 1998. Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (*Mimulus*). *Genetics* **149** : 367-382.
- Bradshaw H.D., Wilbert S., Otto K., Schemske D., 1995. Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). *Nature* **376** : 762-765.
- Brewer S., Cheddadi, R., de Beaulieu, J.L., Reille, M., 2002. The spread of deciduous *Quercus* throughout Europe since the last glacial period. *For. Ecol. Manage.***156** : 27-48.
- Britvin E., Minkov D., Glikson L., Ronin Y., and Korol A., 2001. MultiQtl, an interactive package for genetic mapping of correlated quantitative trait complexes in multiple environments, version 2.0 (Demo). Plant & Animal Genome IX, San Diego, CA, Abstract C01\_01.html.
- Burger W.C., 1975. The species concept in *Quercus*. *Taxon*, **24** : 45-50.
- Burke J.M., Voss T.J., Arnold M.L., 1998. Genetic interactions and natural selection in Louisiana iris hybrids. *Evolution* **52** : 1304-1310.
- Camargo L.E.A., Osborn T.C., 1996. Mapping loci controlling flowering time in *Brassica oleracea*. *Theor. Appl. Genet.* **92**: 610-616.
- Castiglioni P., Ajmone-Marsan P., Van Wijk R., Motto M., 1999. AFLP markers in a molecular linkage map of maize: codominant scoring and linkage group distribution. *Theor Appl Genet* **99** : 425-431
- Churchill G.A., Doerge R.W., 1994 . Empirical threshold values for quantitative trait mapping. *Genetics* **138** : 963-971.

- Coart E., Lamote V., De Loose M., Van Bockstaele E., Lootens P., Roldán-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. *Theor Appl Genet* **105** : 431-439.
- Cornelius J., 1994. Heritabilities and additive genetic coefficients of variation in forest trees. *Can. J. For. Res.* **24** : 372-379.
- Coyne J.A., Orr H.A., 1998. The evolutionary genetics of speciation, *Phil. Trans. R. Soc. Lond.* **353** : 287-305.
- Coyne J.A., 1992. Genetics and speciation, *Nature* **355** : 511-515.
- Cracraft J., 1983. Species concepts and speciation analysis. *Curr. Ornithol.* **1** : 159-187.
- Degen, B. (2000). SGS: Spatial Genetic Software. Computer program and user's manual.
- Demarly Y., 1977. Génétique et amélioration des plantes. Ed Masson . 287 pp.
- de Vicente M. C., Tanksley S. D., 1993. QTL analysis of transgressive segregation in an interspecific tomato cross. *Genetics* **134** : 585-596.
- de Vienne D., 1997. Etablissement de cartes de liaisons génétiques. In: de Vienne D (eds). Les marqueurs moléculaires en génétique et biotechnologie végétales. INRA, Paris pp 49-79.
- de Vienne D., Causse M., 1997. La cartographie et la localisation des locus contrôlant la variation des caractères quantitatifs. In: de Vienne D (eds). Les marqueurs moléculaires en génétique et biotechnologie végétales. INRA, Paris pp 99-118
- Dieckmann U., Doebeli M., 1999. On the origin of species by sympatric speciation, *Nature* **400** : 354-357.
- Dobb D.M.B., 1989. Reproductive isolation as a consequence of adaptive divergence in *Drosophila pseudoobscura*. *Evolution* **43** : 1308-1311.
- Dobzhansky T., 1937. Genetic of the Evolutionary process. Columbia University press, New York.
- Dobzhansky T., Pavlovsky O., 1957. An experimental study of interaction between genetic drift and natural selection. *Evolution* **11**: 311-319.
- Doebley J., Stec A., 1993. Inheritance of the morphological differences between maize and teosinte :Comparison of results for two F2 populations , *Genetics* **134** : 559-570.
- Ducousso A. Guyon J.P., Kremer A., 1996. Latitudinal and altitudinal variation of bud burst in western populations of sessile oak (*Quercus petraea* (Matt) Liebl). *Ann. Sci. For.* **53** : 775-782.
- Ducousso A., Bennehart H., Kremer A., 1997. Geographic variation of *Quercus rubra* L., in European tests. Diversity and Adaptation in oak species. In Proceedings of the second meeting of working party 2.08.05, Genetics of *Quercus*, of the International Union of Forest Research Organisations. pp 58-67.

- Dumolin-Lapègue S., Demesure B., Le Corre V., Fineschi S., Petit R.J., 1997. Phylogeographic structure of white oaks throughout the European continent. *Genetics* **146** : 1475–1487.
- Dumolin-Lapègue S., Kremer A., Petit R.J., 1999. Are chloroplast and mitochondrial DNA variation species-independent in oaks? *Evolution* **53** : 1406–1413.
- Dupouey J.L., 1983. Analyse multivariable de quelques caractères morphologiques de populations de chênes (*Quercus petraea* (Matt.) Liebl. et *Quercus robur* L.) du Hurepoix. *Ann. Sci. For.* **40** : 187-194.
- Dupouey J.L., Badeau V., 1993. Morphological variability of oaks (*Quercus robur* L, *Quercus petraea* (Matt) Liebl, *Quercus pubescens* Willd.) in northeastern France :preliminary results. *Ann. Sci. For.* **50**, Suppl 1 : 35s-40s.
- Endler J.A., 1977. Geographic variation, speciation, and clines. Princeton University Press, Princeton, New Jersey.
- Farquhar G.D., Ehleringer J.R., Hubick K.T., 1989. Carbon isotope discrimination and photosynthesis. *Annual Review of Plant Physiology and Plant Molecular Biology* **40** : 503–537.
- Feltus F.A., Hart G.E., Paterson A.H., 2003. Comparative Analysis of Two Sorghum Mapping Populations: Statistically Significant Overlap of Tiller Height, Flowering Time, Seed Weight, and Panicle Length QTLs. Plant and Animal Genome XI Conference. San Diego, California, January 2003.
- Fisher R.A., 1930. The genetical theory of natural selection. Clarendon Press Oxford, UK.
- Fisher, R.A., 1921. On the 'probable error' of a coefficient of correlation deduced from a small sample. *Metron* **1** pt 4, 3-32.
- Frewen B.E., Chen T., Howe G.T., Davis J., Rohde A., Boerjan W., Bradshaw H.D., 2000. Quantitative trait loci and candidate gene mapping of bud set and bud flush in *Populus*. *Genetics* **154** : 837–845.
- Futuyma D.J., and Mayer G.C., 1980. Non-allopatric speciation in animals. *Syst. Zool.* **29** : 254-271.
- Gadau J., Page R.E., Werren J.H., 1999. Mapping of hybrid incompatibility loci in *Nasonia*, *Genetics* **153** : 1731-1741.
- Gadau J., Page R.E., Werren J.H., 2002. The genetic basis of the interspecific differences in wing size in *Nasonia* (Hymenoptera; Pteromalidae): Major quantitative trait loci and epistasis. *Genetics* **161** : 673-684.
- Georgiady M.S., Whitkus R.W., Lord E.M., 2002. Genetic analysis of traits distinguishing outcrossing and self-pollinating forms of currant tomato, *Lycopersicon pimpinellifolium* (Jusl.) Mill. *Genetics* **161** : 333-344.

- Gömöry D., Yakovlev I., Zhelev P., Jedináková 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
- Gömöry D.A., 2000. Gene coding for a non-specific NAD-dependent dehydrogenase shows a strong differentiation between *Quercus robur* and *Quercus petraea*. *Forest Genetics* **7** : 167-170.
- Gottlieb L.D., 1984. Genetics and morphological evolution in plants, *Am. Nat.* **123** : 681-709.
- Gottlieb, L.D., 1973. Genetic differentiation, sympatric speciation, and the origin of a diploid species of *Stephanomeira*. *Am. J. Bot.* **60** : 545-553.
- Goudet J., Raymond M., Demeus T., Rousset F., 1996. Testing differentiation in diploid populations. *Genetics* **144** : 1933-1940.
- Grandillo S., Tanksley S.D., 1996. QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species *Lycopersicon pimpinellifolium*. *Theor. Appl. Genet.* **92** : 935-951.
- Grandjean G., Sigaud P., 1987. Contribution à la taxonomie et à l'écologie des chênes du Berry. *Ann. Sci. For.* **44** : 35-66.
- Grant V., 1958. The regulation of recombination in plants. *Cold Spring Harbor Symp. Quant. Biol.* **23** : 337-363.
- Grattapaglia D., Bertolucci F.L., Sederoff R.R., 1995. Genetic mapping of quantitative trait loci (QTLs) controlling vegetative propagation in *Eucalyptus grandis* and *E. urophylla*, using the pseudo-testcross mapping strategy and RAPD markers. *Theor Appl Genet.* **90** : 933-947.
- Grattapaglia D., Sederoff R., 1994. Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross mapping strategy and RAPD markers. *Genetics* **137** : 1121-1137.
- Grattapaglia, D., Bertolucci, F.L.G.; Penchel, R., Sederoff, R., 1996. Genetic mapping of QTLs controlling growth and wood quality traits in *Eucalyptus grandis* using maternal half-sib family and RAPD markers. *Genetics* **144** : 1205-1214.
- Haldane J., 1919. The combination of linkage values and the calculation of distance between the loci of linked factors. *J. Genet.* **8** : 299-309.
- Hamrick J.L., Godt M.J.W., Scherman-Brayles S.T., 1992. Factors influencing levels of genetic diversity in woody plant species. *New Forest* **6** : 95-124.
- Han K.-H., Bradshaw H.D., Gordon, M.P., 1995. Adventitious root and shoot regeneration in vitro is under major gene control in an F2 family of hybrid poplar (*Populus trichocarpa* x *P. deltoides*). *Forest Genetics* **1** : 139-146.
- Hardy O., Vekemans X., 2002. SPAGeDI 1.0. A program for Spatial Pattern Analysis of Genetic Diversity.

- Hartl D.L., Clark A.G., 1989. Principles of Populations Genetics. 2Ed. Sinauer Associates.
- Hatfield T., Schluter D., 1999. Ecological speciation in sticklebacks: environment-dependent hybrid fitness. *Evolution* **53** : 866-873.
- Huntley B., Birks H.J., 1983. An atlas of past and present pollen maps for Europe, 0-13 000 years ago. Cambridge University Press, Cambridge.
- Hurme P., Sillanpaa M.J., Arjas E., Repo T., Savolainen O., 2000. Genetic basis of climatic adaptation in Scots pine by bayesian quantitative trait locus analysis. *Genetics* **156** : 1309-1322.
- Huxley J.S., 1942. Evolution: The Modern Synthesis. London: Allen and Unwin.
- Ioerger T.R., Clark A.G., Kao T.H., 1990. Polymorphism at the self-incompatibility locus in *Solanaceae* predates speciation. *Proc. Natl. Acad. Sci. USA* **87** : 9732-9735.
- Jansen R.C., Stam P., 1994. High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136** : 1447-1455.
- Jansen, R.C., 1993. Interval mapping of multiple quantitative trait loci. *Genetics* **135** : 205-211.
- Jermstad K.D., Bassoni D.L., Jech K.S., Wheeler N.C., Neale D.B., 2001. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. I. Timing of vegetative bud flush. *Theor. Appl. Genet.* **102** : 1142-1151.
- Johannesson K., 2001. Parallel speciation: a key to sympatric divergence. *Trends Ecol. Evol.* **16** : 148-153.
- Jones C.D., 1998. The genic basis of *Drosophila sechellia*'s resistance to a host plant toxin. *Genetics* **149** : 1899-1908.
- Jones E.W., 1959. Biological flora of the British Isles. *Quercus* L.. *Journal of Ecology* **47** : 169-222.
- Kampfer S., Lexer C., Glössl J., Steinkellner H., 1998. Characterization of (GA)<sub>n</sub> microsatellite loci from *Quercus robur*. *Hereditas* **129** : 183-186.
- Kass S.U., Landsberger N., Wolffe A.P., 1997. DNA methylation directs a time-dependent repression of transcription initiation. *Curr Biol* **7** : 157-165.
- Kaya Z., Sewell M.M., Neale D.B., 1999. Identification of quantitative trait loci influencing annual height- and diameter-increment growth in loblolly pine (*Pinus taeda* L.). *Theor Appl Genet* **98** : 586-592.
- Killias, 1980. A multifactorial genetic investigation of speciation theory using *Drosophila melanogaster*. *Evolution* **34** : 730-737.
- Kim S.C., Rieseberg L.H., 1999. Genetic architecture of species differences in annual sunflowers : Implications for adaptive trait introgression. *Genetics* **153** : 965-977.

- Kingsolver J.G., Hoekstra H.E., Hoekstra J.M., Berrigan D., Vignieri S.N., Hill C.E., Hoang A., Gibert P., Beerli P., 2001. The strength of phenotypic selection in natural populations. *Am. Nat.* **157** : 245-261.
- Kissling P., 1980. Clé de détermination des chênes médioeuropéens (*Quercus* L.). *Ber.schweiz.bot.Ges.* **90** : 1-28.
- Kleinschmit J., Bacilieri R., Kremer A., Roloff A., 1995a. Comparison of morphological and genetic traits of pedunculate oak (*Q. robur* L.) and sessile oak (*Q. petraea* [Matt.] Liebl.). *Silvae Genetica.* **44** : 256-269.
- Kleinschmit J., Kleinschmit J.G.R., 2000. *Quercus robur* - *Quercus petraea*: a critical review of the species concept. *Glasnik Za Sumske Pokuse* **37** : 441-452.
- Kleinschmit, J. , Kremer A., Roloff A., 1995b. Sind stieleiche und traubeneiche zwei getrennte arten? *AFZ.* **26** : 1453-1456.
- Kondrashov A.S, Kondrashov F.A, 1999. Interactions among quantitative traits in the course of sympatric speciation. *Nature*: **400** : 351-354.
- Kosambi D.D., 1944. The estimation of map distance from recombination values. *Ann Eugen* **12** : 172-175.
- Kremer A., 1997. Les marqueurs moléculaires en génétique des populations. Dans Les marqueurs moléculaires en génétique et biotechnologie végétales. Ed INRA p 119-138.
- Kremer A., Dupouey J.L., Deans J.D., Cottrell J., Csaikl U., Finkeldey R., Espinel S., Jensen J., Kleinschmit J., Van Dam B., Ducouso A., Forrest I., de Heredia U.L., Lowe A., Tutkova M., Munro R.C., Steinhoff S., Badeau V., 2002a. Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands. *Ann. For. Sci.* : **59** : 777-787.
- Kremer A., Le Corre V., Mariette S., 2000. Population differentiation for adaptive traits and their underlying loci in forest trees : theoretical predictions and experimental results. *Forest Genetics and Sustainability*, (C. Mátyás ed) **63** : 59-74.
- Kremer A., Petit R.-J., Ducouso A., 2002b. Biologie évolutive et diversité génétique des chênes sessile et pédonculé. *Revue Forestière Française* **LIV** : 111-130.
- Lander E.S., Botstein D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121** : 185-199.
- Lander E.S., Green P., Abrahamson J., Barlow A., Daly M.J., Lincoln S.E., Newburg L. ,1987. MAPMAKER : an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1** : 174-181.

- Latta L.G., 1998. Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. *Am. Nat.* **151** : 283-292.
- Laurie C.C., True J.R., Liu J., Mercer J.M., 1997. An introgression analysis of quantitative trait loci that contribute to a morphological difference between *Drosophila simulans* and *D. mauritiana*. *Genetics* **145** : 339-348.
- Le Corre V., Dumolin-Lapègue S., Kremer A., 1997. Genetic variability at enzyme and RAPD loci in sessile oak *Quercus petraea* (Matt.) Liebl.: the role of history and geography. *Molecular Ecology* **6** : 519-529.
- Le Corre V., Kremer A., 2003. Genetic variability at neutral markers, Quantitative Trait Loci and trait in a subdivided population under selection. *Genetics* **164**:1205-1219.
- Lerceteau E., Plomion C., Andersson B., 2000. Quantitative Trait Loci (QTLs) for economically important traits in *Pinus sylvestris* L. *Mol. Breed.* **6** : 451-458.
- Li Z., Pinson S.R.M., Paterson A.H., Park W.D., Stansel J.W., 1997. Genetics of hybrid sterility and hybrid breakdown in an intersubspecific rice (*Oryza sativa* L.) population. *Genetics* **145** : 1139-1148.
- Li, Z., Pinson S.R.M., Stansel J.W., Park W.D., 1995. Identification of quantitative trait loci (QTLs) for heading date and plant height in cultivated rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **91** : 374-381.
- Lin J.Z., Ritland K., 1997. Quantitative trait loci differentiating the outbreeding *Mimulus guttatus* from the interbreeding *M. platycalyx*. *Genetics* **146** : 1115-1121.
- Lynch M., Milligan. B., 1994. Analysis of population-genetic structure using RAPD markers. *Molecular Ecology* **3** : 91-99.
- Macdonald S.J., Goldstein D.B., 1999. A quantitative genetic analysis of male sexual traits distinguishing the sibling species *Drosophila simulans* and *Drosophila sechellia*. *Genetics* **153** : 1683-1699.
- Malamy J. E., and Benfey, P. N., 1997. Organization and cell differentiation in lateral roots of *Arabidopsis thaliana*. *Development* **124** : 33-44.
- Manos P.S., Stanford, A.M., 2001. The historical biogeography of *Fagaceae* :tracking the Tertiary history of temperate and subtropical forests of the northern hemisphere. *Int. J. Plant Sci.* **162** (6 Suppl.) : s77-s93.
- Mariette S., 2001. Mesure de la diversité génétique intra et inter-populations : Echantillonnage intragénome et choix des marqueurs. Applications expérimentales sur *Quercus petraea*, *Quercus robur* et *Pinus pinaster*. Thèse de Doctorat. ENGREF. 90p.

- Mariette S., Cottrell J., Csaikl U.M., Goikoechea P., König A., Lowe A.J., Van Dam B.C., Barreneche T., Bodénès C., Streiff R., Burg K., Groppe K., Munro R.C., 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
- Marques C.M., Vasquez K.J., Carocha V.J., Ferreira J.G., O'Malley D.M., Liu B.H., Sederoff R.R., 1999. Genetic dissection of vegetative propagation traits in *Eucalyptus tereticornis* and *E. globulus*. *Theor. Appl. Genet.* **99** : 936-946.
- Martinez O., Curnow R.N., 1992. Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theor Appl Genet* **85** : 480-488.
- Mayden R. L., 1997. A Hierarchy of Species Concepts: The Denouement in the Saga of the Species Problem, p. 381-424. In: Species: The Units of Biodiversity. Claridge, M.F., Dawah, H.A. and Wilson, M.R. (eds). Chapman and Hall Ltd., London.
- Maynard Smith J., 1966. Sympatric speciation. *Am. Nat.* **100** : 637-650.
- Mayr E., 1942. Systematics and the Origin of Species, Columbia University press.
- Mayr E., 1963. Animal Species and Evolution. Belknap Press, Cambridge, Mass.
- Michelmore, R.W., Para, I., Kasser, R.V., 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations *Proc. Natl. Acad. Sci. USA* **88** : 9828-9832.
- Mitchell-Olds T., 1996. Genetic constraints on life-history evolution :Quantitative trait loci influencing growth and flowering in *Arabidopsis thaliana*. *Evolution* **50** : 140-145.
- Moreau F., Kleinschmit J., Kremer A., 1994. Molecular differentiation between *Q. petraea* and *Q. robur* assessed by random amplified DNA fragments. *Forest Genetics* **1** : 51-64.
- Moritz D.M.L., Kadereit J.W., 2001. The genetics of evolutionary change in *Senecio vulgaris* L.: a QTL mapping approach. *Plant Biology* **3** : 544-552.
- Morton N. E., 1955. Sequential tests for the detection of linkage *Amer. J. Hum. Gen* **7** : 277-318.
- Muir G., Fleming C.C., Schlötterer C., 2000. Species status of hybridizing oaks. *Nature* **405** : 1016.
- Muir G., Fleming C.C., Schlötterer C., 2001. Three divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. *Mol. Biol. Evol.***18** : 112-119.
- Muller H.J., 1940. Bearings of the *Drosophila* work on systematics. In: The new Systematics (Husley J.S., ed) pp.185-268, Clarendon Press.
- Muller H.J., 1942. Isolating mechanisms, evolution, and temperature. *Biol. Symp.* **6** : 71-125
- Nei M., 1973. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* **70** : 3321-3323.
- Nei M., 1987. Molecular Evolutionary Genetics. Columbia University Press, New York

- Oppenheimer D.G., Herman P.L., Sivakumaran S., Esch J., Marks M.D., 1991. A myb gene required for leaf trichome differentiation in *Arabidopsis* is expressed in stipules. *Cell* **67** (3) : 483-93.
- Orr A.H., 1998. Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait loci data. *Genetics* **149** : 2099-2104
- Orr A.H., 2001. The genetics of species differences. *Trends in Ecol. Evol.* **16** : 343-350.
- Orr H.A., Coyne J.A., 1992. The genetics of adaptation: a reassessment. *Am. Nat.* **140** : 725-742
- Otto S.P., Whutton J., 2000. Polyploid incidence and evolution. *Annu.Rev.Genet.* **34** : 401-437.
- Paterson, H. E. H. 1985. The recognition concept of species. pp. 21-29 in E. S. Vrba, ed. *Species and Speciation*. Transvaal Museum, Pretoria.
- Pernes J, Lourd M., 1984. Organisation des complexes d'espèces. In :Gestion des ressources génétiques des plantes. Paris : Agence de coopération culturelle et technique : 7-108.
- Petit R.J., Brewer S., Bordács S., Burg K., Cheddadi R., Coart E., Cottrell J., Csaikl U.M., van Dam B.C., Deans J.D., Fineschi S., Finkeldey R., Glaz I., Goicoechea P.G., Jensen J.S., König A.O., Lowe A.J., Madsen S.F., Mátyás G., Munro R.C., Popescu F., Slade D., Tabbener H., de Vries S.M.G., Ziegenhagen B., de Beaulieu J.-L., Kremer A., Espinel S., 2002a. Identification of refugia and postglacial colonization routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *For. Ecol. Manage.* **156** : 49–74.
- Petit R.J., Csaikl U.M., Bordács S., Burg K., Coart E., Cottrell J., van Dam B.C., Deans J.D., Dumolin-Lapègue S., Fineschi S., Finkelday R., Gillies A., Glaz I., Goicoechea P.G., Jensen J.S., König A., Lowe A.J., Madsen S.F., Mátyás G., Munro R.C., Pemonge M.-H., Popescu F., Slade D., Tabbener H., Turchini D., de Vries S.M.G., Ziegenhagen B., Kremer A., Olalde M., 2002b . Chloroplast DNA variation in European white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. *For. Ecol. Manage.* **156** : 5–26.
- Petit R.J., Kremer A., 1993. Ribosomal DNA and chloroplast DNA polymorphisms in a mixed stand of *Quercus robur* and *Q. petraea*. *Ann. Sci. For.* **50** Suppl. 1 : 41s-47s.
- Petit R.J., Pineau E., Demesure B., Bacilieri R., Ducouso A., Kremer A., 1997. Chloroplast DNA footprints of postglacial recolonization by oaks. *Proc. Natl. Acad. Sci. USA* **94** : 9996–10001.
- Plomion C., Costa P., Bahrman N., Frigerio J.M., 1997. Genetic analysis of needle proteins in maritime pine. 1. Mapping dominant and codominant protein markers assayed on diploid tissue, in a haploid-based genetic map. *Silvae genetica* **46** : 161-165.
- Poncet V., Lamy F. , Devos K.M. , Gale M.D. , Sarr A. , Robert T. , 2000. Genetic control of domestication traits in pearl millet (*Pennisetum glaucum* L., *Poaceae*). *Theor.Appl.Genet.* **100** : 147-159.

- Poncet V., Martel E., Allouis S., Devos K.M., Lamy F., Sarr A., Robert T., 2002. Comparative analysis of QTL affecting domestication traits between two domesticated x wild pearl millet (*Pennisetum glaucum* L., *Poaceae*) crosses. *Theor. Appl. Genet.* **104** : 965-975.
- Ponton S., Dupouey J.L., Breda N., Dreyer E., 2002. . Comparison of water-use efficiency of seedlings from two sympatric oak species: genotype x environment interactions. *Tree Physiology* **22** : 413-422.
- Ponton S., Dupouey J.L., Bréda N., Feuillat F., Bodénès C., Dreyer E., 2001. Carbon isotope discrimination and wood anatomy variations in mixed stands of *Quercus robur* and *Quercus petraea*. *Plant, cell and Environment* **24** : 861-868.
- Porth I, Koch M., Berenyi M., Bodénès C., Barreneche T., Saintagne C., Kremer A., Burg K. 2002. Isolation and mapping of putative adaptation specific osmotic induced expressed sequenced sequence tags of the oak genome. Poster presentation in Dyden Conference decembre 2002.
- Quillet M.C., Madjidlan N., Griveau T., Serieyss H., Tersac M., 1995. Mapping genetic factors controlling pollen viability in an interspecific cross in *Helianthus* section *Helianthus*. *Theor. Appl. Genet.* **91** : 1195-1202.
- Raymond M., Rousset F., 1995a. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86** :248-249.
- Raymond M., Rousset F., 1995b. An exact test for population differentiation. *Evolution* **49** : 1280-1283.
- Rice W.R., Hostert E.E., 1993. Laboratory experiments on speciation :What have we learned in 40 years? *Evolution* **47** : 1637-1653.
- Ridley M., 1993. *Evolution*. Blackwell Scientific Publications, Cambridge, UK.
- Rieseberg L.H, Baird S.J.E., Gardner K.A., 2000. Hybridization, introgression, and linkage evolution, *Plant Molecular Biology* **42** : 205-224.
- Rieseberg L.H, Whitton J., Gardner K., 1999b. Hybrid zones and genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* **152** : 713-727.
- Rieseberg L.H., Archer M.A., Wayne R.K., 1999a. Transgressive segregation, adaptation and speciation. *Heredity* **83** : 363-372.
- Rieseberg L.H., Burke J.M., 2001. A genic view of species integration. *Journal of Evolutionary Biology* **14** : 883-886.
- Rieseberg L.H., Widmer A., Arntz A.M., Burke J.M., 2002. Directional selection is the primary cause of phenotypic diversification. *Proc. Natl. Acad. Sci. USA* **99** : 12242-12245.
- Rieseberg, L.H., C.R. Linder, G. Seiler, 1995. Chromosomal and genic barriers to introgression in *Helianthus*. *Genetics* **141** : 1163-1171.

- Ritter E., Gebhardt C., Salamini F., 1990. Estimation of recombination frequencies and construction of RFLP linkage maps in plant from crosses between heterozygous parents. *Genetics* **125** : 645-654.
- Ritter E., Salamini F., 1996. The calculation of recombination frequencies in crosses of allogamous plant species with application to linkage mapping. *Genet. Res.* **67** : 55-65.
- Rundle D., Nagel L., Wenrick Boughman J., Schluter D., 2000. Natural selection and parallel speciation in sympatric Sticklebacks, *Science* **287** : 306-308.
- Sano Y., 1990. The genic nature of gamete eliminator in rice. *Genetics* **125** : 183-191.
- Schemske D.W., Bradshaw H.D., 1999. Pollinator preference and the evolution of floral traits in monkey flowers (*Mimulus*). *Proc. Natl. Acad. Sci. USA* **96**: 11910-11915.
- Schliewen U., Rassmann K., Markmann M., Markert J., Kochers T., Tautz D., 2001. Genetic and ecological divergence of a monophyletic cichlid species pair under fully sympatric conditions in lake Ejagham, Cameroon, *Molecular Ecology* **10** : 1471-1488.
- Schluter D., 2001. Ecology and the origin of species. *Trends Ecol. Evol.* **16** : 372-380.
- Schwarzbach A.E., Donovan L.A., Rieseberg L.H., 2001. Transgressive character expression in a hybrid sunflower species, *Am. J. Bot.* **88** : 270-277.
- Simpson G.G., 1951. The species concept. *Evolution* **5** : 285-298.
- Sokal R.R., Oden N.L., 1978. Spatial autocorrelation in biology. 1. Methodology. *Biol. J. Linnean Soc.* **10** : 199-228.
- Spethmann W., 1986. Stecklingsvermehrung von Stiel- und Traubeneiche (*Quercus robur* L. und *Quercus petraea* (Matt.) Liebl.). Schriften aus der Forstlichen Fakultät der Universität Göttingen. In : J.D. Sauerländer's Verlag ed , Frankfurt am Main. **86** : 99 pp.
- Stebbins G.L., 1957. The hybrid origin of microspecies in the *Elymus glaucus* complex. *Cytologia Suppl.* **36** : 336-340.
- Steinhoff S., 1993. Results of species hybridization with *Quercus robur* L. and *Q. petraea* (Matt.) Liebl.. *Ann.Sci.For* **50** suppl.1: 137-143.
- Steinhoff S., 1997. Results of *Quercus* hybridization work from 1989-1996 at Escherode (*Quercus petraea* (Matt.) Liebl.and *Quercus robur* L.). In : Diversity and adaptation in oak species. Proceedings Pennsylvania State University Oct.12-17, 1997.
- Steinkellner H., Fluch S., Turetschek E., Lexer C., Streiff R., Kremer A., Burg K., Glossl J., 1997. Identification and characterization of (GA/CT)<sub>n</sub> - microsatellite loci from *Quercus petraea*. *Plant Mol. Biol.* **33** :1093-1096.
- Stephan B.R., König A., Liepe K., Venne H., 1995. Klimakammer- und Freilandversuche zur Frosthärte und Phänologie von Herkünften der Traubeneiche (*Quercus petraea* [Mattuschka] Liebl.). *Mitteilungen aus der Forstlichen Versuchsanstalt Rheinland-Pfalz* **34** : 50-74.

- Streiff R., 1998. Structure génétique spatiale et flux polliniques dans une chênaie mixte : *Quercus petraea* (Matt.) Liebl., *Quercus robur* L.. Thèse de Doctorat. Université Montpellier II. 84p
- Sucena E., Stern, D.L., 2000. Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by cis-regulatory evolution of ovo/shaven-baby. *Proc. Natl. Acad. Sci. USA* **97** : 4530-4534.
- Takahashi A., Tsaur S.C., Coyne J.A., Wu C.I., 2001. The nucleotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **98** : 3920-3925.
- Tanksley S.D., 1993. Mapping polygenes. *Annu.Rev.Genet.* **27** : 205-233.
- Templeton A.R., 1981. Mechanisms of speciation – a population genetics approach. *Annu. Rev. Ecol. Syst.* **12** : 23-48.
- Templeton A.R., 1989. The meaning of species and speciation: a genetic perspective. In D. Otte and J. A. Endler (eds.), *Speciation and its consequences*, Sinauer Associates, Sunderland, Mass.
- Ting C.T., Tsaur S.C., Wu C.-I., 2000. The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. *Proc. Natl. Acad. Sci. USA* **97** : 5313-5316.
- True T.R., Liu J., Stam L.F., Zeng Z.-B., Laurie C.C., 1997. Quantitative genetic analysis of divergence in male secondary sexual traits between *Drosophila simulans* and *Drosophila mauritiana*, *Evolution* **51** : 816-832.
- Tsarouhas V., Gullberg U., Lagercrantz U., 2002. An AFLP and RFLP linkage map and quantitative trait locus (QTL) analysis of growth traits in *Salix*. *Theor. Appl. Genet.* **105** : 277-288.
- Turreli M., Orr A.H., 2000. Dominance, epistasis, and the genetics of postzygotic isolation. *Genetics* **154**: 1663-1679.
- Ungerer M.C., Halldorsdottir S.S, Modliszewski J.L, Mackay T.F.C., Purugganan M.D, 2002. Quantitative Trait Loci for inflorescence development in *Arabidopsis thaliana*, *Genetics* **160** : 1133-1151.
- Van Ooijen J.W., Voorrips R.E., 1993. Plant Research International, Biometris, Wageningen. See P Stam, *Plant J* **3**: 739-744.
- Van Valen L., 1976. Ecological species, multispecies and oaks. *Taxon* **18** : 233-239.
- Via S., 2001. Sympatric speciation in animals : the ugly duckling grows up. *Trends Ecol. Evol.* **16** : 381-390.
- Visscher, P.M., Thompson, R., Haley, C.S., 1996. Confidence intervals in QTL mapping by bootstrapping. *Genetics* **143** : 1013-1020.

- Vos P., Hogers R., Bleeker M., Rijans M., Van de Lee T., Hornes M., Frijters A., Pot J., Kuiper M., Zabeau M., 1995. AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res.* **23** : 4407-4414.
- Vuylsteke M., Mank R., Antonise R., Bastiaans E., Senior M.L., Stuber C.W., Melchinger A.E., Lübberstedt T., Xia X.C., Stam P., Zabeau M., Kuiper M. (1999). Two high-density AFLP linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. *Theor Appl Genet* **99**: 921-935.
- Weir B.S., Cockerham C.C., 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38** : 1358-1370.
- Westerbergh A., Doebley J., 2002. Morphological traits defining species differences in wild relatives of maize are controlled by multiple quantitative trait loci, *Evolution* **56** : 273-283.
- Wright S., 1969. Evolution and the genetics of populations. Vol. 2. The theory of gene frequencies. University of Chicago Press.
- Wu C.I., 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology* **14** : 851-865.
- Wu R., Stettler R.F.; 1996. The genetic resolution of juvenile canopy structure and function in a three-generation pedigree of *Populus*. *Trees* **11** : 99-108.
- Young W.P., Schupp J.M., Keim P., 1999. DNA methylation and AFLP marker distribution in the soybean genome. *Theor. Appl. Genet.* **99** : 785-792.
- Zanetto A., 1998. Structuration géographique de la diversité allozymique dans le complexe des chênes blancs européens : *Quercus petraea* (Matt.) Liebl. et *Quercus robur* L.. Thèse de Doctorat. Université Paris XI Orsay. 113p.
- Zanetto A., Roussel G., Kremer A., 1994. Geographic variation of inter-specific differentiation between *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. *Forest Genetics* **1** : 111-123.
- Zeng K.-B., Liu J., Stam L.F., Kao C.-H., Mercer J.M., Laurie C.C., 2000. Genetic architecture of a morphological shape difference between two drosophila species, *Genetics* **154** : 299-310.
- Zeng Z.B., 1994. Precise mapping of quantitative trait loci. *Genetics* **136** : 1457-1468.
- Zoldos V., Siljak-Yakovlev S., Papes D., Sarr A., Panaud O., 2001. Representational difference analysis reveals genomic differences between *Q. robur* and *Q. suber* : implications for the study of genome evolution in the genus *Quercus*. *Mol. Genet. Genom.* **265** : 234-241.

# ANNEXES

## Distribution of genomic regions differentiating oak species assessed by QTL detection

Caroline Saintagne, Catherine Bodénès, Teresa Barreneche, David Pot, Christophe Plomion, Antoine  
Kremer\*

INRA

Laboratoire de génétique et Amélioration des arbres Forestiers

69, Route d'Arcachon

33612 Cestas Cedex

France

\*corresponding author

Tel: 00 33 5 57 12 28 32

Fax: 33 5 57 12 28 81

Email: [antoine.kremer@pierroton.inra.fr](mailto:antoine.kremer@pierroton.inra.fr)

**Word count:** 5730

**Running title:** Genomic regions differentiating oak species

**Key words:** Leaf morphology, QTL, *Quercus petraea*, *Q. robur*

## ABSTRACT

Pedunculate oak and sessile oak are two sympatric interfertile species that exhibit leaf morphological differences. We aimed at locating QTLs (Quantitative Trait Loci) of these traits in order to locate genomic regions involved in species differentiation. Fifteen leaf morphological traits were assessed in a mixed forest stand composed of *Q. petraea* and *Q. robur* and in a full-sib pedigree of *Q. robur*. The progeny of the full-sib family were vegetatively propagated in two successive experiments comprising 174 and 216 sibs and assessments were made on two leaves collected on each of the 1080 and 1530 cuttings corresponding to the two experiments. Heritability values of the morphological traits varied between 0 and 0.61 depending on the trait and on the dominance variance, with pubescence of the lamina exhibiting the highest genetic control and obversity the lowest. Traits that exhibited strong species differences in the mixed stand tended also to have higher heritability values. A genetic map was constructed for QTL detection. Composite Interval Mapping with the one QTL model was used for QTL detection. No QTL was detected for two among the 15 characters measured and from one to three QTLs were detected for the remaining 13 traits. In-depth analysis of the QTLs controlling the 5 morphological traits that exhibited the highest interspecific differences in the mixed stand, indicated that they were distributed on six linkage groups, with two clusters comprising QTLs of at least two discriminant leaf traits. These conclusions were reinforced when error 1 for QTL detection was set at 5% at the chromosome level, as up to nine clusters could be identified.

## INTRODUCTION

Pedunculate oak (*Q. robur* L.) and Sessile oak (*Q. petraea* (Matt.) Liebl.) are closely related species with a widely sympatric distribution in Europe. Their taxonomic boundary is not clearly defined and has raised conflicting opinions between botanists. As sessile and pedunculate oak are interfertile in natural conditions (Rushton, 1977), the biological species concept (Mayr, 1963) has been challenged by some authors (Kleinschmit and Kleinschmit, 2000). The ecological species definition is likely more adapted because the two species occupy different ecological niches. Indeed, *Q. robur* preferentially grows on rich, wetter and more alkaline habitats whereas *Q. petraea* occurs on drier and more acidic soils (Becker and Levy, 1990). Nevertheless, in spite of their ecological differences, in many cases the two species coexist in the same mixed stands occupying different microsite niches, and most frequently intermediate forms appear under these circumstances (Olsson, 1975, Rushton, 1979, Dupouey, 1983, Bacilieri *et al.*, 1996).

Interspecific variation in mixed oak stand has raised a general interest in Europe since decades among scientists and foresters. Because of their interfertility and their sympatry, the two species exhibit extremely low genetic differentiation. Various genetic markers have been tested for species differentiation : isozymes (Zanetto *et al.*, 1994, Gömöry, 2000), ribosomal DNA (Petit *et al.*, 1993, Muir *et al.*, 2001), chloroplast DNA (Petit *et al.*, 1993), proteins (Barreneche *et al.*, 1996), anonymous DNA fragments (Moreau *et al.*, 1994, Bodénès *et al.*, 1997, Mariette *et al.*, 2002, Cervera *et al.*, 2000, Aitman *et al.*, 2000, Zoldos *et al.*, 2001). These investigations showed the absence of species specific marker, and that the two species displayed only allele frequency differences. These results differed markedly from phenotypic observations made in mixed stands. As already mentioned, when the two species coexist in the same stands, they show clear ecological preferences. They also differ by leaf morphological characters. Survey on leaf morphological traits made in different European countries showed that the two species exhibited a bimodal distribution for a few leaf morphological traits (Aas, 1993, Dupouey and Badeau, 1993). A recent study using the same techniques over a wide sample of populations confirmed these earlier observations (Kremer *et al.*, 2002) that a few leaf morphological traits show continuous but bimodal distribution in mixed stands. Hence, there is a clear discrepancy between genetic and phenotypic differentiation. An interpretation of such a discrepancy has raised from genetic surveys based on a large set of markers (RAPD and AFLP analysis). When numerous markers were tested, it became apparent that a rare proportion exhibited larger frequency variation between the 2 species. Bodénès *et al.* (1997) screened the genome of the two species, and showed that only 2% of 2800 RAPD fragments actually exhibited significant frequency differences. These results clearly suggested that genomic regions involved in species

differentiation are rare and explain why earlier molecular investigations using a low number of loci proved to be inefficient.

In this contribution, we attempt to investigate the distribution of these informative (species discriminant) genomic regions within the oak genome. Are those regions aggregated or are they distributed throughout the genome? The strategy adopted is a downstream approach from the informative and discriminant phenotypic traits to their underlying loci. As quantitative differences in phenotypes between individuals are due to the segregation of alleles at multiple quantitative trait loci (QTLs), we intend to map the QTLs that control species discriminant leaf morphological traits. With the recent and widespread availability of polymorphic molecular markers, QTL mapping has been used to locate the genomic regions involved in interspecific differentiation in various species: in *Drosophila* (Zeng *et al.*, 2000, True *et al.*, 1997), in tomato (Grandillo *et al.*, 1996) in Monkeyflower (Bradshaw *et al.*, 1995, Bradshaw *et al.*, 1998, Lin *et al.*, 1997), in rice (Cai *et al.*, 2002, Brondani *et al.*, 2002), in maize (Doebley *et al.*, 1993, Westerbergh *et al.*, 2002) in millet (Poncet *et al.*, 2000, Poncet *et al.*, 2002) in sunflowers (Kim *et al.*, 1999). In the present study, we aim at locating QTLs of 15 leaf morphological characters that are traditionally assessed for oak species identification. The QTLs detection was performed in a cloned full sib family of *Q. robur* comprising 278 progeny. Interspecific variation of the morphological characters was first estimated on phenotypic values recorded in natural a mixed stand.

## **MATERIAL AND METHODS**

### **Natural population**

Leaf morphological traits were assessed in a mixed stand of *Q. petraea* and *Q. robur* located in the North-west of France (Bacilieri *et al.*, 1995; Kremer *et al.*, 2002). The study area is square (250m\*250m) and comprised a total of 426 adult trees (older than 100 years) approximately evenly distributed between the two species (199 *Q. petraea*, 215 *Q. robur* and 12 so-called intermediate). The sampling was exhaustive and five leaves were sampled within the upper crown of each tree. The morphological data were earlier analysed (Bacilieri *et al.*, 1995; Kremer *et al.*, 2002) and will be reported here only for the variance components of the traits.

### **Mapping population**

An intraspecific hybrid family of *Q. robur* was obtained by crossing two adult trees. The female parent (accession 3P) was located on the Forestry Research Station of Pierroton (latitude 44.44N, longitude 0.46W ) and the male parent (accession A4) originated from Arcachon (latitude 44.40N, longitude 1.11W ). More than 400 acorns were obtained from the controlled cross, among which 278 produced seedlings. A subset of 94 individuals were previously used for the construction of parental genetic maps (Barreneche *et al.* 1998) and 184 additional full-sibs were genotyped for QTL mapping

in the present study. The 278 seedlings were raised in a seed bed in the nursery until age 3, and they were then transplanted in a clonal bank as stool-beds. The spacing between the stool-beds was 1.5m x 1.5m. After two years following the transplantation, when the seedlings were 5 years old, each stool bed was hedged at the level of the ground in late February. The stool-beds stump sprouted in late April, and when the sprouts were fully elongated, they were cut in cuttings for vegetative propagation. The cuttings consisted in stem segments of about 10 to 12 cm long comprising at least two internodes. They were dipped in a rooting hormone (Rhipozon AA 0.5) and transplanted in small containers (80 cm<sup>3</sup> in volume with a 60/40 mixture of peat/ perlite ) and their leaves were cut in half. The cuttings were then transferred into a greenhouse that was watered by a fog system. The first rooting occurred after five weeks, and in general more than 60 % of the cuttings rooted within two months after installation in the greenhouse. The rooted cuttings were further transplanted in four liters container (with a mixture of 2/1/2 : peat/sand/pine bark) in September. They over-wintered in a greenhouse and were installed outdoor in the nursery for an additional year. This procedure for vegetative propagation was recurrently used over four successive years. The QTL detection reported here was carried out in two experiments.

#### Experiment 1.

174 full sibs were successively propagated in summer 1997 in sufficient numbers and were further transplanted from the nursery to a field test located in Bourran (along the Garonne River) in the South West of France. The experiment design comprised 36 incomplete blocks each containing single vegetative replicates of 30 full-sibs. The whole experiment comprised 1080 cuttings (on average six vegetative replicates per full-sib). The cuttings were planted in the spring 1999, and two fully elongated leaves were collected on each vegetative propagule in summer 1999. Leaf morphological data were adjusted to block effect prior to data analysis.

#### Experiment 2

This experiments comprised 216 full-sibs of the same mapping pedigree. Vegetative propagation took place in the spring of 1998. Two leaves were collected in the summer of 1999 in the nursery on each of 1530 cuttings (on average 7 vegetative replicates per full-sib). The experimental design was considered as fully randomised.

### Assessment of leaf morphological traits

Fourteen morphological traits were assessed on all the leaves collected in the natural populations and in the two experiments. The measurements were made following the procedure indicated by Kremer *et al.*(2002) : lamina length (LL), petiole length (PL), lobe width (LW), sinus width (SW), length of lamina at largest width (WP) number of lobes (NL), number of intercalary veins (NV), basal shape of the lamina (BS), lamina shape (OB=WP/LL), petiole ratio (PR=(PL/(PL+LL))), lobe depth ratio (LDR= (LW-SW)/LW), percentage venation (PV=NV/NL), lobe width ratio (LWR=LW/LL). In addition to the morphological assessments, observation of pubescence (PU) was made at two different spots on the abaxial surface of each leaf : in the central part of the lamina (PUI) and along the veins (PUv). Measurements on the veins were only made on the cuttings and not on the adult leaves. Observations of pubescence were made following Kissling's grading system (Kissling, 1980), ranging from 1 (no pubescence) to 6 (dense pubescence). Pubescence was assessed using a stereo microscope.

### Statistical analysis

#### Estimation of variance components in the natural population

The value of each morphological trait was decomposed according to the following two way ANOVA model using the OPEP Software (Baradat *et al.*, 1995)

$$X = \text{mean} + \text{species} + \text{tree} + \text{leaf} \quad (1)$$

In a second step the subdivision was done separately within each species. Model (2) was successively used for *Q.petraea* and *Q.robur*

$$Y = \text{mean} + \text{tree} + \text{leaf} \quad (2)$$

All sources of variation were considered as random. Consequently the following ratios were computed to account for the proportion of the total variation explained by different sources of variation.

$S = \sigma_{\text{species}}^2 / \sigma_X^2$  derived from model (1) is an estimate of the variation explained by species differentiation, whereas  $T = \sigma_{\text{tree}}^2 / \sigma_Y^2$  derived from model (2) represents the proportion of variation within a species that is due to tree variation within that species.

#### Estimation of variance components in the full-sibs pedigree

All data were analysed using analysis of variance (ANOVA) with the OPEP Software (Baradat and Labbé, 1995), Because vegetative propagation was used to propagated the full-sibs of the mapping

pedigree, the subdivision of the phenotypic value of a cutting could be made with the following model in the two experiments:

$$Z = \text{mean} + \text{clone} + \text{cutting} \quad (3)$$

In a set of full-sib families, the total genotypic variance ( $V_G$ ) is the sum of within ( $\sigma_w^2$ ) and between ( $\sigma_b^2$ ) full-sib family variances

$$V_G = \sigma_w^2 + \sigma_b^2$$

And

$V_G = V_A + V_D$  where  $V_D$  and  $V_A$  are the additive and dominance variance

And where the variance between full-sib families is:

$$\sigma_b^2 = \frac{1}{2} V_A + \frac{1}{4} V_D$$

Given that the trees that were vegetatively propagated were full-sibs, the clonal variance  $\sigma_{clone}^2$  is equal to  $\sigma_w^2$ , and the expectation of the clonal variance  $\sigma_{clone}^2$  in this case, is:

$$\sigma_{clone}^2 = \frac{1}{2} V_A + \frac{3}{4} V_D$$

As dominance variance is generally lower than the additive variance in forest trees (Cornelius, 1994), we consider two extremes values for the dominance variance in order to estimate additive variances and narrow sense ( $h^2$ ) heritabilities ( $V_D=0$  and  $V_D=1/2 V_A$ )

$$\text{If } V_D=0, \text{ then } V_A = 2 \sigma_{clone}^2 \quad \text{and} \quad h^2 = \frac{2\sigma_{clone}^2}{2\sigma_{clone}^2 + \sigma_{cutting}^2}$$

$$\text{If } V_D=1/2 V_A, \text{ then } V_A = \frac{8}{7} \sigma_{clone}^2 \quad \text{and} \quad h^2 = \frac{\frac{8}{7} \sigma_{clone}^2}{\frac{12}{7} \sigma_{clone}^2 + \sigma_{cutting}^2}$$

In addition we also calculated the repeatability of the clonal means ( $R$ ) :

$$R = \frac{\sigma_{clone}^2}{\sigma_{clone}^2 + \frac{\sigma_{cutting}^2}{n}}$$

where  $n$  is the number of vegetative propagules for each clone.

### **Construction of the linkage map**

The mapping strategy used in this study was the so called double-pseudo-testcross mapping strategy (Grattapaglia *et al.*, 1994). Based on the framework map that was obtained on the same pedigree but only on 94 full sibs, (Barreneche *et al.*, 1998), we extended the mapping to 278 individuals and further mapped also additional anchor markers on the total set of individuals. The additional mapping consisted in 38 microsatellites, 6 SCARs ( Bodénès *et al.* 1996) and 13 primers enzymes combinations of AFLP markers (8 Eco/Mse and 5 Pst/Mse). Amplification fragment length polymorphism (AFLP) genotyping was performed as described by Vos *et al* (1995). AFLP markers were named according to the restriction enzyme used, the three selective nucleotides used on the amplification primers, the number of base pair of the band, the quality of the scoring (1 difficult to read to 5 very easy). Microsatellites genotyping was performed using the primers developed by Kampfer (1998) and Steinkellner (1997). AFLP fragments and microsatellites were separated by electrophoresis on a LI-COR (model 4000) DNA sequencer. Primers were labelled with the fluorescent infrared dye IRD800 (LI-COR, Lincoln, Neb.) and fragment analysis recorded by a laser system. Microsatellite and AFLP protocols are described on the web site INRA (<http://www.pierroton.inra.fr/Oakflow/>).

Mapmaker 2.0 (Lander *et al.*, 1987) was used to infer linkage relationships among markers. To allow the detection of linkage of markers in repulsion phase, the data set was duplicated and re-coded. First, markers were divided into linkage group with the group command (parameters LOD=7;  $\theta=0.25$ ). Second, markers were ordered in the linkage group with the first order command (LOD =3,  $\theta = 0.40$ ). The ordered marker sequences were confirmed using the ripple and LOD table commands. Kosambi's (Kosambi, 1994) mapping function was then used to transform the recombination frequency between

linked loci into centimorgan (cM) distances. The graphical presentation of linkage maps and QTL obtained using the MapChart software version2.0 (Voorrips, 2001).

### **QTL detection**

The analytical method used to identify putative QTLs and to estimate their phenotypic effect was the Composite Interval Mapping (CIM, Zeng 1994, Jansen and Stam, 1994) which is an extension of interval Mapping (IM, Lander and Botstein 1989). Interval mapping calculates the likelihood score for a putative QTL located at any position within an interval flanked by two adjacent markers. CIM extends this method by fitting the most significant markers outside the interval into the model, allowing more accuracy and efficiency of QTL mapping (Zeng, 1994). The Multiqtl software (Britvin *et al*, 2001), (<http://esti.haifa.ac.il/~poptheor>) was used with no extended parameters (no LOD normalization and no missing data restoration). The standard deviation for each QTL position was estimated by bootstrap (Visscher *et al*. 1996) with 1000 samplings. Empirical statistical significance thresholds for declaring the presence of a QTL were determined by permutations of the data set (Churchill and Doerge, 1994). Thousand different samplings were generated from the actual data by shuffling the trait values with respect to the marker genotypes. Each individual genotyped was randomly assigned to one of the trait values from the sample. In the single QTL model permutations permitted to test the null hypothesis of no QTL contributing to the trait.

Because the permutations tests were calculated at the chromosome level, we further computed the corresponding type I error rate at the whole genome level. The relationship between type I error rate at the genome level ( $\alpha_g$ ) and type I error rate at the chromosome level ( $\alpha_{chr}$ ) is as follows :

$$\alpha_{chr} = 1 - \left\{ 1 - \left[ 1 - (1 - \alpha_g)^{\frac{1}{M}} \right] \right\}^m$$

where,  $M$  is the total number of markers used for the QTL detection on each map, and  $m$  the number of markers in the linkage group.

The QTL detection was done separately for each parental map in both experiments.

The following ANOVA model was used to subdivide the variance components of the clonal means in the QTL detection:

$$\bar{Z} = \text{mean} + \text{marker} + \text{clone} \quad (4)$$

Where  $\bar{Z}$  is the clonal mean as derived from model (3).

If  $\sigma_{\bar{z}}^2$  is the variance of the clonal means ( $\sigma_{\bar{z}}^2 = \sigma_{clone}^2 + \sigma_{cutting}^2 / n$ ), and if  $P$  is the proportion of  $\sigma_{\bar{z}}^2$  explained by the QTL, then the proportion  $Q$  of the clonal variance explained by the QTL is  $Q$  with  $Q = P/R$

Hence we can compute the proportion  $S$  of the genetic (additive) variance explained by the QTL in the two cases where  $V_D=0$  and  $V_D=1/2 V_A$

When  $V_D=0$ , then  $S= Q/2$

And when  $V_D=1/2 V_A$ , then  $S= 7Q/8$

Finally, we can compute the proportion of the phenotypic variance explained by the QTL ( $PV$ ) with  $PV= S* h^2$

To sum up, since vegetative propagation was used to multiply the full-sibs of the mapping pedigree, we were able to estimate the contribution of QTLs to three variances of interest:

- the contribution to the variance of the clonal mean values ( $P$ ), e.g to the phenotypic variance of clonal means within the full-sib cross
- the contribution to the additive variance ( $S$ ) in the population where the parents were sampled, assuming two contrasting values of the dominance variance. An implicit assumption for estimating  $S$  is that the within family variance estimated in one single full sib family here is representative of the within family variance of any full sib cross within the population.
- the contribution to the phenotypic variance ( $PV$ ) in the population where the parents of the mapping pedigree were sampled.

## RESULTS

### Subdivision of the phenotypic variation in the natural population

All sources of variation contributed significantly to the diversity of leaf morphology as shown by the different ANOVAs computed. However the contribution of the different sources were quite different according to the traits considered (Table 1). For size related traits such as length of the lamina (LL), lobe width (LW) species differences represent only a minor part of the total variation. Conversely more than 80 % of the total variation was due to species subdivision for petiole length (PL) and petiole ratio (PR). To a lesser degree, the venation (NV, 67% and PV, 70%), pubescence (PUI 77%) and sinus width (SW, 67%), basal shape of lamina (BS, 72%) were also among the traits that exhibited important species variation. However, despite the important species variation, the range of

variation within a species significantly overlapped the range of variation for the other species. As an example, *Q. petraea* exhibited a petiole length (PL) varying between 3 and 33 mm in *Q. petraea*, whereas PL varied between 0 and 12 mm in *Q. robur*. For pubescence, *Q. robur* is often reported as glabrous and indeed 84% of the leaves did not exhibit any pubescence, however 16% of the leaves did exhibit hairs on the abaxial surface of the lamina. Similarly, *Q. petraea* has usually only a few intercalary veins. Indeed only 75 % of the leaves did not show any intercalary veins and in *Q. robur* and this percentage amounted to 5% in *Q. robur*. The subdivision of variation within a species indicated that the traits that showed high species differentiation exhibited also an important tree-to-tree variation within each species (Table 1).

### **Subdivision of phenotypic variation in the mapping pedigree**

In experiment 1, data were first adjusted to block effects prior using ANOVA model (3), whereas in experiment 2 data were directly analysed using ANOVA model (3). All phenotypic values were normally distributed as tested by Box and Cox method (Box and Cox., 1964) except PUv. The clonal effect was significant in all ANOVAs except for OB and LWR in experiment 1 (data not shown).

Heritability and repeatability values were slightly higher in experiment 2, as the result of higher environmental variance due to the heterogeneity of environmental conditions in the field test (experiment 1) (Table 2). Pubescence (either PUI or PUv) was the trait with the highest heritability, and OB or LWR exhibited the lowest genetic control in both experiments. Overall morphological traits that were reported to be strongly involved in species differences (PL, SW, NV, BS, PU and their derived variables PR, LDR, PV see Table 1) were also among those that had the largest heritability values, especially in experiment 2.

### **Correlations between leaf morphological traits**

Correlation between tree values derived from ANOVA model (2) in the natural population (Table 3a), or clone values derived from ANOVA model (3) in the mapping pedigree (Table 3b) were computed. Size related leaf characters (LL, SW, LW, and WP) were significantly correlated between each other, mostly as a result of the common size effect (data not shown). We focused our attention here only on those traits that exhibited a high between species variation according to Table 1 (e.g., PL, SW, NV, BS and PU) and discarded PR, LDR and PV that are actually transformed variables of PL, SW and NV. The values of these five characters change significantly from *Q. petraea* to *Q. robur*. As shown in Table 1, *Q. petraea* has larger petioles (PL), less intercalary veins (NV), less pronounced sinuses

(SW), shaper basal lamina of the leaf (BS) and exhibits more pubescence (PU) than *Q. robur*. One can therefore consider these 5 characters to be correlated at the interspecific level.

Correlations among these characters are less pronounced at the intraspecific level. In the natural populations, only 5 correlations among the 10 pairwise associations were significant in *Q. petraea*, and only one in *Q. robur* (Table 3a). In the mapping pedigree, where the computed correlations represent mostly genetic correlations, 3 of the 10 correlations were significant in experiment 1, and 5 in experiment 2 (Table 3b). A more intriguing feature of the significant correlations is that in some cases the sign of the correlation is different between the inter and intraspecific level. Among the 14 significant correlations on Table 3a and 3b, 3 changed sign from the inter- to the intra-specific level. This is illustrated by the correlation between PL and NV. It is well known and shown in Table 1 that leaves of *Quercus robur* have smaller petioles (PL) and more intercalary veins (NV). However within the mapping pedigree in *Q. robur*, clones that had smaller petioles tended to bear less intercalary veins (Table 3b). Correlations of opposite sign between the inter- and intraspecific level also occurred in *Q. petraea* (Table 3a).

#### **Linkage map for QTL detection**

The 13 AFLP primers enzyme combinations provided 201 markers segregating in a 1:1 ratio. Among them, 4% showed a significant deviation ( $P < 1\%$ ) from Mendelian segregation and were further discarded. The genotyping of the offspring included also 38 microsatellites and 6 SCARs. Nine additional RAPD markers were selected from the reference map (Barreneche *et al*, 1998) and were included in this study in order to fill some gaps on few linkage groups. Among the 254 markers obtained, a subset of 128 markers (84 AFLP, 34 microsatellites, 1 SCAR, 9 RAPD) evenly distributed in the genome were finally selected to construct the map for QTL detection. The two maps shared 19 microsatellites. The female framework map was composed of 75 markers and covered a distance of 890cM, with an average spacing between markers of 14 cM. In comparison to the reference map this framework map covered about 75% of the pedunculate estimated genome length (1192 cM). The male framework map comprised 72 markers over a total distance of 929 cM, covering 75% of the genome (estimated total length is 1235cM) with an average marker spacing of 15.5cM.

#### **QTL mapping**

No QTL was detected for 2 traits (LW, OB) among the 15 leaf morphological characters assessed when the probability threshold was set at 5% at the genome level. Genetic variation for LW and OB was low as shown by their heritability values (Table 2).

For the remaining 13 traits, between 1 and 3 significant QTLs were detected which explained from 8 to 40% of the additive genetic variance on average, and from 11.1 to 13.8 % of the clonal mean variance on average (Table 4). Overall the proportion of genetic variance explained by the QTLs detected in this pedigree remains moderate, despite the high heritability values of the traits, especially in experiment 2. There was no obvious relationship between the level of the heritability and the number of QTLs, or the proportion of additive variance explained by the QTLs. For example in the case of pubescence (PUv or PUI) that exhibited the highest heritability (Table 2), only two QTLs were detected only in experiment 1, with each QTL explaining between 9 and 20% of the additive variance (Table 4). For six traits (PL, SW, NL, NV, LDR and PV), QTLs were detected in both experiments. Interestingly these traits (except NL) are strongly involved in species differentiation. When considering only the 5 traits that exhibited the highest species differentiation ( eg. the same traits presented in Table 3 : PL, SW, NV, BS and PU), their QTLs were located on 6 different linkage groups (1, 3, 4, 5, 11 and 12) when bulking the information over both experiments (Table 4 and Figure 1). Some of these QTLs were detected on the same linkage group in both experiments (for PL on LG1, SW on LG 3). Colocalisation of QTLs of at least two different traits involved in species differentiation can be seen on LG 1 (PL and PU), LG 3 (NV, SW). A clustering of QTLs controlling raw or derived traits involved in species differentiation can clearly be seen on LG3 (PV, NV, LDR, SW).

When the type I error rate for the QTL detection was relaxed from 5% at the genome level to 5% at the chromosome level, then the former clusters of QTLs involved in species differentiation were reinforced by the addition of suggestive QTLs for other informative morphological traits (Figure 2):

- on LG1F near 30 cM : PUI1, PL1, PL2, BS2 and PUI1 on LG1M
- on LG3F near 20 cM: NV1, NV2, PL2, SW1, SW2 and BS2, SW1, SW2 on LG3M
- on LG4M near 10 cM: PL1 and PUv2
- on LG6M near 40 cM: PUI2, SW1 and Pul1 on LG6F
- on LG9F near 10 cM: PUv1, PUv2, SW1 and BS2 on LG9M
- on LG12F near 30 cM: PUI1 and NV2

In addition to these clusters, three additional were identified (data not shown):

- on LG5F near 40 cM: NV2 and SW1 on LG5M
- on LG7F near 30 cM: PUv1, PUv2 and PUI2 on LG7M

## DISCUSSION

### QTL mapping as a method for detecting genomic regions involved in oak species differentiation.

As leaf morphology traits are of low economic importance, they were only seldom studied for estimating genetic variances or QTL detection. Most of the fifteen traits assessed exhibited heritabilities varying in the range of other growth and quality traits usually assessed in forest trees (Cornelius, 1994). Leaf pubescence was the most heritable trait and in general there was a trend towards larger heritabilities for those traits that show also higher species differentiation. Using the Composite Interval mapping with the one QTL model, no QTL was detected for only two among the 15 characters assessed. From one to three QTLs were detected for the remaining 13 characters, that contributed from 8 to 40 % on average to the additive variance of the traits. These figures are comparable to those obtained recently on leaf morphological characters in *Brassica oleracea*, where 6 to 40% of the additive genetic variation was explained by QTLs detected by the multiple regression approach (Sebastian *et al.*, 2002). The QTLs explained on average 11.1% to 13.8% of the clonal mean variance, which can be compared to other reported studies on similar morphological traits : from 2 to 37 % in *Brassica oleracea* (Lan *et al.*, 2001) and from 5 to 52 % in *Gossypium hirsutum* and *G. barbadense* (Jiang *et al.*, 2000). However when expressed as a proportion of the phenotypic variance within *Q. robur* (and not within the cross) than the QTLs expressed only from 2.4 to 2.8 %.

The major objective of this contribution was to map genomic regions involved in species differentiation by locating QTLs involved in phenotypic differences between *Q. petraea* and *Q. robur*. We choose leaf morphology traits as they have been explored earlier in many studies for species identification. Indeed, we reanalysed the variation of 15 of these traits and showed that some of them exhibit important variation among the two species. The QTL strategy should however not be considered as an exhaustive way of identifying informative regions. First of all, our mapping pedigree was an intraspecific full-sib family, which may have reduced the chances to detect informative QTLs. Some of these loci may be fixed within a species at different alleles and exhibit complete interspecific differentiation. Hence they would not segregate in a full sib intraspecific family. Ideally an interspecific mapping pedigree (either F2 or backcross) would be necessary to detect these QTLs, as usually done in species such as *Drosophila* (Zeng *et al.*, 2000), *Mimulus* (Lin *et al.*, 1997). However due to experimental constraints related to the length of generations, F2 or backcrosses are not available in oaks. Second, some of the loci may be missed by chance just because they are not polymorphic between the two parents of the mapping pedigree. Finally if QTLs controlling morphological traits are numerous and if each one contributes only for a low part of the genetic

variation of the trait, then it can also be missed because the family size of our mapping pedigree was too low (the so called Baevis effect, Baevis, 1995), despite the vegetative propagation used to increase the power for QTL (Bradshaw and Foster, 1992). Neither of these hypothesis can be completely excluded. However the existing data of genetic diversity within the complex of the two species (Kremer at Petit, 1993; Mariette *et al.*, 2002) suggest that they are highly polymorphic, and that fixation within species has never been met for nuclear markers that were polymorphic within the complex. Even when the genomes of the two species were screened for numerous markers (RAPDs, Bodénès *et al.*, 1996 or AFLPs, Mariette *et al.*, 2002), polymorphism was always shared among the two species. Hence we expected that polymorphism for QTLs involved in species differentiation are likely to be shared as well. High levels of polymorphism in oaks increases the chances to detect QTLs in one single full-sib family, but additional experiments in other pedigrees need to be implemented to complete the detection. Finally larger family sizes would be needed to increase statistical power of QTL detection of morphological traits as our results indicated that most QTLs contributed individually only for a minor part of the phenotypic variation of the trait. Because the offspring in the full sib family were vegetatively propagated, we were able to estimate the proportion of the additive variance explained by the QTLs, and therefore indirectly conclude on the saturation of the additive variance by the number of QTLs identified in our mapping pedigree. Depending on the level of the dominance variance, the QTLs detected contributed on average from 8 % to 40 % of the additive genetic variance (Table 4). Given that this proportion may be inflated due to the Baevis effect, we can conclude that only a reduced part of the informative genomic regions involved in species differentiation are therefore identified by QTL detection.

#### **Distribution of genomic regions differentiating oak species**

When the analysis was restricted to the 5 morphological characters involved in species differentiation, we found that their QTLs were distributed on 6 linkage groups when the type I error rate was set at 5% for the whole genome and on 9 linkage groups when it was set at 5 % at the chromosome level. Recalling that our method provides an incomplete number of informative QTLs, we may conclude that species differentiation occurs at multiple sites within the genome and is not confined to a few spots with limited recombination. Our results in oaks contribute to the heterogeneity of results in the same area that have recently been reviewed by Orr (2001). In a synthesis of QTLs detections oriented towards species differentiation, Orr reported that the number of informative QTLs varied between one up to twenty. Examples of the heterogeneity of results are given in *Drosophilae* (*D. simulans* and *D. mauritiana*, Zeng *et al.*, 2000) or *Mimulus* (Lin and Ritland, 1997). In both cases, depending on the traits measured the number of QTLs can vary from a

few to several, including the occurrence of major genes. Part of the inconsistency could be explained by experimental constraints as QTL detection is highly dependent on sample sizes in mapping pedigrees (Wilcox *et al.*, 1997), as it is also illustrated in our study by the differences observed between the two experiments. Given that sample sizes of mapping pedigrees in the few hundred usually tend to inflate the contribution of the detected QTLs, we may conclude that a larger number of informative genomic regions for species differentiation exist within the genome. In addition, our results based on an underestimate of these regions clearly indicated that species differentiation is spread throughout the genome. However QTLs involved in different discriminant morphological characters tend to be distributed in clusters. No conclusion can be drawn as to whether these clusters actually correspond to one single locus with pleiotropic effects or to linked loci contributing to different morphological traits. A further question is whether interactions among these clusters (epistatic effects), or inter locus disequilibria may also contribute to the overall differentiation among the two species.

#### **Acknowledgement**

The study has been carried out with the financial support from the Commission of the European Communities, Community research programme "Quality of life and Management of Living resources" (Project OAKFLOW QLK5-2000-00960). We are very grateful to Guy Roussel for making the intraspecific cross, to Maria Evangelista Bertocchi for the vegetative propagation of the offspring, to Jean-Marc Louvet and Alexis Ducouso for field collection and assessments of morphological traits and to the crew of the INRA nursery at Pierroton (Olivier Lagardère, Bernard Montoussé, Laurent Salera) for the helpful assistance in field work.

#### **REFERENCES**

- Aas G (1993). Taxonomical impact of morphological variation in *Quercus robur* and *Q. petraea* : a contribution to the hybrid controversy. *Ann Sci For* **50**, Suppl. 1 : 107s-113s.
- Aitman T, Cooper L, Norsworthy P, Wahid F, Gray J, Curtis B, et al (2000). Species status of hybridising oaks. Brief communication. *Nature* **405** : 1016.
- 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 in the northwest of France. *Silvae Genetica* **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.
- Baradat P, Labbé T (1995). Traitement statistique des essais de sélection. In Opep. Un logiciel intégré pour l'amélioration des plantes pérennes. Anonymous, Cirad, Montpellier.

- Barreneche T, Bahrman N, Kremer A (1996). Two dimensional gel electrophoresis confirms the low level of genetic differentiation between *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. *Forest Genetics* **3** : 89-92.
- Barreneche T, Bodénès C, Lexer C, Trontin JF, Fluch S, Streiff R, et al (1997). A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 55 rDNA markers. *Theor Appl Genet* **97**: 1090-1103.
- Beavis, WD (1995). The power and deceit of QTL experiments : lessons from comparative studies. In : Proceedings of the 49th annual corn and sorghum industry research conference, Chicago, IL. pp. 250-266.
- Becker M, Levy L (1990). Le point sur l'écologie comparée du chêne sessile (*Quercus petraea* (Matt.) Liebl. et le chêne pédonculé (*Quercus robur* L.). *Rev For Fr*, **42** : 148-154.
- Bodénès 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* **7** : 433-444.
- Bodénès C, Laigret F, Kremer A (1996). Inheritance and molecular variations of PCR-SSCP fragments in pedunculate oak (*Quercus robur* L) *Theor Appl Genet* **93** : 348-354.
- Box GEP, Cox DR (1964). An analysis of transformation. *J Amer Stat Assn* **39** : 357-365.
- Bradshaw HD, Foster GS (1992). Marker-aided selection and propagation systems in trees : advantages of cloning for studying quantitative inheritance. *Can J For Res* **22** : 1044-1049.
- Bradshaw HD, Otto K, Frewen B, McKay J, Schemske D (1998). Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (*Mimulus*). *Genetics* **149** : 367-382.
- Bradshaw HD, Wilbert S, Otto K, Schemske D (1995). Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). *Nature (London)* **376** : 762-765.
- Britvin E, Minkov D, Glikson L, Ronin Y, and Korol A (2001). MultiQtl, an interactive package for genetic mapping of correlated quantitative trait complexes in multiple environments, version 2.0 (Demo). Plant & Animal Genome IX, San Diego, CA, Abstract C01\_01.html.
- Brondani C, Rangel PHN, Brondani RPV, Ferreira ME (2002). QTL mapping and introgression of yield-related traits from *Oryza glumaepatula* to cultivated rice (*Oryza sativa*) using microsatellites markers. *Theor Appl Genet* **104** : 1192-1203.
- Cai HW, Morishima H (2002). QTL cluster reflect character associations in wild and cultivated rice. *Theor Appl Genet* **104** : 17-1228.
- Cervera MT, Remington D, Frigerio JM, Storme V, Ivens B, Boerjan W, et al. (2000). Improved AFLP analysis of tree species. *Can J For Res* **30** : 1608-1616.

- Churchill GA, Doerge RW (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138** : 963-971.
- Cornelius J (1994). Heritabilities and additive genetic coefficients of variation in forest trees. *Can J For Res* **24** : 372-379.
- Doebley J, Stec A (1993). Inheritance of the morphological differences between maize and teosinte : comparison of results for two F2 populations. *Genetics* **134** : 559-570.
- Dupouey JL (1983). Analyse multivariable de quelques caractères morphologiques de populations de chênes (*Quercus petraea* (Matt.) Liebl. et *Quercus robur* L.) du Hurepoix. *Ann Sci For* **40** : 187-194.
- Dupouey JL, Badeau V (1993). Morphological variability of oaks (*Quercus robur* L, *Quercus petraea* (Matt) Liebl, *Quercus pubescens* Willd.) in northeastern France : preliminary results. *Ann Sci For* **50**, Suppl 1: 35s-40s.
- Gömöry DA (2000). Gene coding for a non-specific NAD-dependent dehydrogenase shows a strong differentiation between *Quercus robur* and *Quercus petraea*. *Forest Genetics* **7** : 167-170.
- Grandillo S, Tanksley SD (1996). QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species *Lycopersicon pimpinellifolium*. *Theor Appl Genet* **92** : 935-951.
- Grattapaglia D, Sederoff R (1994). Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross mapping strategy and RAPD markers. *Genetics* **137** : 1121-1137.
- Jansen RC, Stam P (1994). High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136** : 1447-1455.
- Jiang C, Wright R, Woo S, DelMonte T, Paterson A (2000). QTL analysis of leaf morphology in tetraploid *Gossypium* (cotton). *Theor Appl Genet* **100** : 409-418.
- Kampfer S, Lexer C, Glössl J, Steinkellner H (1998). Characterization of (GA)<sub>n</sub> microsatellite loci from *Quercus robur*. *Hereditas* **129** : 183-186.
- Kim SC, Rieseberg L (1999). Genetic architecture of species differences in annual sunflowers : implications for adaptive trait introgression. *Genetics* **153** : 965-977.
- Kissling P (1980). Clé de détermination des chênes médioeuropéens (*Quercus* L.). *Ber.schweiz.bot.Ges.* **90** : 1-28.
- Kleinschmit J, Kleinschmit JGR (2000). *Quercus robur* - *Quercus petraea*: a critical review of the species concept. *Glasnik Za Sumske Pokuse* **37**: 441-452.
- Kosambi DD (1944) The estimation of map distance from recombination values. *Ann Eugen* **12** : 172-175.

- Kremer A, Dupouey JL, Deans JD, Cottrell J, Csaikl U, Finkeldey R, et al (2002). Morphological variation in mixed oak stands (*Quercus robur* and *Quercus petraea*) is stable across western European populations. *Annals of Forest Science* (in press)
- Kremer A, Petit RJ (1993). Gene diversity in natural populations of oak species. *Ann Sci For* **50** (suppl. 1) : 186s-202s.
- Lan TH, Paterson AH (2001). Comparative mapping of QTLs determining the plant size of *Brassica oleracea*. *Theor Appl Genet* **103** : 383-397.
- Lander ES, Botstein D (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121** : 185-199.
- Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, et al (1987). MAPMAKER : an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1** : 174-181.
- Lin JZ, Ritland K (1997). Quantitative trait loci differentiating the outbreeding *Mimulus guttatus* from the inbreeding *M. platycalyx*. *Genetics* **146** : 115-1121.
- Mariette S, Cottrell J, Csaikl M, Goikoechea P, König A, Lowe AJ, et al (2002). Comparaison of level of diversity detected with AFLP and microsatellite markers within and among mixed *Q. petraea* (Matt.) Liebl. and *Q. robur* L. stands. *Sivae genetica* (In press)
- Mayr E (1963). Populations, espèces et évolutions. Ed Hermann, Paris, 496p
- Moreau F, Kleinschmit J, Kremer A (1994). Molecular differentiation between *Q. petraea* and *Q. robur* assessed by random amplified DNA fragments. *Forest Genetics* **1** : 51-64.
- Muir G, Fleming CC, Schlötterer C (2001). Three divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. *Mol Biol Evol* **18** : 112-119.
- Olsson U (1975). A morphological analysis of phenotype in populations of *Quercus* (fagaceae) in Sweden. *Bot. Notiser* **128** : 55-68.
- Orr A (2001). The genetics of species differences. *Trends in Ecology & Evolution* **16** : 343-350.
- Petit RJ, Kremer A (1993). Ribosomal DNA and chloroplast DNA polymorphisms in a mixed stand of *Quercus robur* and *Q. petraea*. *Ann Sci For* **50**, Suppl. **1** : 41s-47s.
- Poncet V, Lamy F, Devos KM, Gale MD, Sarr A, Robert T (2000). Genetic control of domestication traits in pearl millet (*pennisetum glaucum* L., Poaceae). *Theor Appl Genet* **100** : 147-159.
- Poncet V, Martel E, Allouis S, Devos KM, Lamy F, Sarr A, Roberts T (2002). Comparative analysis of QTL affecting domestication traits between two domesticated x wild pearl millet (*Pennisetum glaucum* L., Poaceae) crosses. *Theor Appl Genet* **104** : 965-975.
- Rushton BS (1977) Artificial hybridization between *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. *Watsonia* **11** : 229-236.

- Rushton BS (1979). *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. : a multivariate approach to the hybrid problem, 2. The geographical distribution of population types. *Watsonia* **12** : 209-224.
- Sebastian RL, Kearsey MJ, King GJ, (2002). Identification of quantitative trait loci controlling developmental characteristics of *Brassica oleracea* L. *Theor Appl Genet* **104** : 601-609.
- Steinkellner H, Fluch S, Turetschek E, Lexer C, Streiff R, Kremer A, et al (1997). Identification and characterization of (GA/CT)<sub>n</sub> - microsatellite loci from *Quercus petraea*. *Plant Mol Biol* **33** : 109.
- True JR Liu JJ Stam LF Zeng ZB, Laurie CC (1997). Quantitative genetic analysis of divergence in male secondary sexual traits between *Drosophila simulans* and *Drosophila mauritiana*. *Evolution*. **51** : 816-832.
- Visscher, P.M., Thompson, R., Haley, C.S. 1996. Confidence intervals in QTL mapping by bootstrapping. *Genetics* **143**: 1013-1020.
- Voorrips RE (2001). MapChart version 2.0 : Windows software for the graphical presentation of linkage maps and QTLs. Plant Research International, Wageningen, TheNetherlands.
- Vos P, Hogers R, Bleeker M, Reijans M, Lee T, Van de Hornes M et al (1995). AFLP : a new technique for DNA fingerprinting. *Nucleic Acids Research* **23** : 4407-4414.
- Westerbergh A, Doebley J (2002). Morphological traits defining species differences in wild relatives of maize are controlled by multiple quantitative trait loci. *Evolution* **56** : 273-283.
- Wilcox P, Richardson T, Carson S (1997). Nature of quantitative trait variation in *Pinus radiata* : insights from QTL detection experiments. Integration of quantitative&molecular genetics. Proceeding of IUFRO genetic of radiata pine,NZ,1-4 december.(FRI Bulletin no203) 304-312.
- Zanetto A, Roussel G, Kremer A (1994). Geographic variation of inter-specific differentiation between *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. *Forest Genetics* **1** : 111-123.
- Zeng ZB (1994) Precise mapping of quantitative trait loci. *Genetics* **136** : 1457-1468.
- Zeng ZB, Liu JJ, Stam LF, Kao CH, Mercer, JM, Laurie CC (2000). Genetic architecture of a morphological shape difference between two *Drosophila* species. *154* : 299-310.
- Zoldos V, Siljak-Yakovlev S, Papes D, Sarr A, Panaud O (2001). Representational difference analysis reveals genomic differences between *Q. robur* and *Q. suber* : implications for the study of genome evolution in the genus *Quercus*. *Mol Genet Genom* **265** : 234-241.

## CAPTIONS OF FIGURES

### Figure 1

#### Title

Distribution of QTL for leaf morphology traits with a threshold fixed to 5% at the genome level

#### Legend

Lamina length (LL), Number of lobes (NL), Petiole length (PL), Number of intercalary veins (NV), Sinus width (SW), Petiole ratio (PR), Lobe depth ratio (LDR), Percentage venation (PV), Length of lamina at largest width (WP), pubescence in the central part of the lamina (PUI), Lobe width ratio (LWR) .

index 1 corresponds to QTLs detected in experiment 1, index 2 corresponds to QTLs detected in experiment 2.

Each QTL is delineated by the position of the highest LOD score and the bootstrap mean value of the highest LODscore. Confidence intervals of their position based on 1000 bootstrap samples are added as lines.

### Figure 2

#### Title

Distribution of QTLs for the most discriminant leaf morphology traits ( $P < 5\%$  at the chromosome level).

#### Legend

Petiole length (PL), Number of intercalary veins (NV), Sinus width (SW), pubescence in the central part of the lamina (PUI), pubescence along the vein (PUv), Basal shape lamina (BS)

index 1 corresponds to QTLs detected in experiment 1, index 2 corresponds to QTLs detected in experiment 2.

Each QTL is delineated by the position of the highest LOD score and the bootstrap mean value of the highest LODscore. Confidence intervals of their position based on 1000 bootstrap samples are added as lines.

**Table 1.** Partitioning of leaf morphology variation in the natural population

Trait	<i>Q. petraea</i>		<i>Q. robur</i>		S
	Mean	T	Mean	T	
LL (mm)	91.4	0.59	77.5	0.64	0.23
PL (mm)	16	0.70	4.4	0.52	0.81
LW (mm)	28.2	0.61	24.5	0.63	0.14
SW (mm)	20.5	0.62	13.4	0.50	0.59
WP (mm)	49.7	0.42	44.4	0.50	0.08
NL	11.2	0.52	8.7	0.50	0.51
NV	0.35	0.32	3.1	0.49	0.67
BS	3.9	0.47	6.8	0.48	0.72
PUI	3.4	0.58	1.2	0.74	0.77
OB %	54.1	0.16	57.3	0.17	0.07
PR %	14.8	0.68	5.4	0.50	0.84
LDR %	26.7	0.47	44.5	0.44	0.67
PV %	3.32	0.34	36.0	0.48	0.70
LWR %	30.8	0.51	31.6	0.40	0.02

$$S = \sigma_{species}^2 / \sigma_X^2$$

$$T = \sigma_{tree}^2 / \sigma_Y^2$$

LL=lamina length, PL=petiole length, LW=lobe width, SW=sinus width, WP=length of lamina at largest width, NL=number of lobes, NV=number of intercalary veins, BS= basal shape of the lamina, OB=lamina shape, PR=petiole ratio, LDR=petiole ratio lobe depth ratio, PV=percentage venation, LWR=lobe width ratio, PUI pubescence on the central part of the lamina, PUv pubescence along the veins.

**Table 2.** Partitioning of leaf morphology variation in the mapping pedigree

Trait	Experiment 1				Experiment 2			
	Mean	Heritability		<i>R</i>	Mean	Heritability		<i>R</i>
		$V_D=0$	$V_D=1/2 V_A$			$V_D=0$	$V_D=1/2 V_A$	
LL(mm)	58.5	0.22	0.13	0.56	69.8	0.25	0.15	0.66
PL(mm)	2.6	0.08	0.05	0.3	3.2	0.34	0.20	0.74
LW(mm)	17.8	0.21	0.12	0.54	20.7	0.25	0.15	0.65
SW(mm)	6.4	0.26	0.15	0.62	8.4	0.33	0.20	0.73
WP(mm)	34.6	0.17	0.10	0.49	41.9	0.24	0.14	0.65
NL	7.7	0.23	0.14	0.58	8.5	0.35	0.21	0.75
NV	5.8	0.17	0.1	0.49	4.2	0.48	0.30	0.84
BS	5.1	0.2	0.11	0.53	4.6	0.27	0.16	0.68
PUv	1.5	0.46	0.28	0.8	1.7	0.61	0.39	0.9
PUI	2.8	0.38	0.23	0.74	2.2	0.48	0.30	0.84
OB%	59.6	0	0	0	61.1	0.12	0.07	0.44
PR%	4.4	0.12	0.07	0.38	4.5	0.23	0.14	0.63
LDR%	62.9	0.18	0.11	0.51	58.9	0.29	0.18	0.70
PV%	76.7	0.04	0.023	0.17	52.9	0.32	0.19	0.73
LWR%	31.0	0.02	0.009	0.07	30.0	0.12	0.07	0.43

*R*: Repeatability of the clonal mean

**Table 3a** Matrix of correlation coefficients between leaf morphological traits computed in the natural population.

	PL	SW	NV	BS	PUI
PL		<b>0.58</b>	-0.11	<b>-0.20</b>	<u><b>-0.21</b></u>
SW	<b>0.23</b>		-0.05	-0.05	<u><b>-0.27</b></u>
NV	-0.01	-0.04		0.09	<b>-0.23</b>
BS	-0.04	0.01	-0.09		-0.001
PUI	-0.05	-0.06	0.07	-0.03	

NL=Number of lobes, PL= Petiole length, NV=Number of intercalary veins , SW= Sinus width, PUI= pubescence in the central part of the lamina

The correlation are based on the tree effect (model 2)

Above diagonal values correspond to *Quercus petraea*.

Below diagonal values correspond to *Quercus robur*

Values in bold characters are significant at  $p \leq 5\%$

Underlined values correspond to correlations that exhibit opposite signs at the inter and intraspecific level (see text)

**Table 3b** Matrix of correlation coefficients between leaf morphological traits computed in the two experiments of the mapping pedigree of *Q. robur*

	PL	SW	NV	BS	PUI	
PL		<b>0.17</b>	<b>0.26</b>	<u><b>0.18</b></u>	<b>-0.20</b>	-0.03
SW	<b>0.43</b>		<b>0.24</b>	<b>-0.46</b>	0.06	-0.05
NV	0.13	-0.05		0.09	<b>0.21</b>	0.01
BS	<b>-0.24</b>	0.11	0.13		0.11	-0.01
PUI	-0.13	-0.13	<b>-0.32</b>	0.11		<b>0.20</b>

NL=Number of lobes, PL= Petiole length, NV=Number of intercalary veins , SW= Sinus width, PUI= pubescence in the central part of the lamina

Correlation are based on the clone effect (model 3)

Values in bold characters are significant at  $p \leq 5\%$

Underlined values correspond to correlations that exhibit opposite signs at the inter and intraspecific level (see text)

Above the diagonal : genetic correlation for experiment 1

Below the diagonal : genetic correlation for experiment 2

Diagonal : genetic correlation between experiment 1 and 2

**Table 4** QTL detection summary for the leaf morphological traits  
Experiment 1

Traits	LG	N	Lod	$\alpha_{ch}$	P1	P2	CI	D	P	$V_D=0$		$V_D=1/2V_A$		
										S	PV	S	PV	
PL	1f	124	3.9	0.004	55.4	46.7	20.2	-0.32	13.6	23	1	40	1.9	
PR	10f	134	3.2	0.005	63.1	60.9	7.7	-0.48	9.7	13	1.5	22	1.5	
LWR	10f	131	2.8	0.005	68.9	65.5	12.9	-2.26	9.9	69	1.1	120	1.1	
SW	3f	126	3.7	0.002	22.3	22.1	6.9	-1.00	15.5	13	3.3	22	3.4	
SW	3m	137	3.6	0.003	19.3	17.2	39.8	-0.81	10.9	9	2.3	15	2.4	
LDR	3f	126	3.7	0.002	20.5	18.2	18.5	4.99	14.5	14	2.6	25	2.7	
LDR	3m	137	3.0	0.003	19.1	18.5	9.2	3.84	9.9	10	1.8	17	1.8	
PUI	12f	110	4.8	0.003	14.9	17.6	12.8	-0.37	17.2	12	4.4	20	4.7	
PUI	1m	108	3.0	0.003	29.3	30.8	16.3	0.33	12.9	9	3.3	15	3.5	
NL	2f	89	3.9	0.005	67	50.7	15.7	-0.94	22.9	20	4.6	34	4.7	
NV	3f	126	4.7	0.002	20.2	15.8	11.9	0.91	16.4	17	2.9	29	2.9	
PV	3f	126	5.7	0.002	20.4	18.3	19	12.19	21.7	66	2.7	115	2.7	
Mean values										13.8	23	2.7	40	2.8

Experiment 2

Traits	LG	N	Lod	$\alpha_{ch}$	P1	P2	CI	D	P	$V_D=0$		$V_D=1/2V_A$		
										S	PV	S	PV	
LL	11f	171	3.2	0.004	2.7	7.3	27.8	-6.64	8.8	7	1.7	12	1.7	
LL	6m	184	2.5	0.003	26	25.3	10.3	5.70	6	5	1.1	8	1.2	
PL	1f	156	4.2	0.004	39	44.2	17.1	-0.41	11.2	8	2.5	13	2.7	
PL	11f	171	3.4	0.004	17.7	19.4	23.3	-0.36	7.9	5	1.8	9	1.9	
SW	3f	162	3.1	0.002	17.9	18.3	5.4	-1.51	8.4	6	1.9	10	2	
SW	3m	160	2.5	0.003	43.2	27.6	33.1	-1.61	9.7	7	2.1	12	2.2	
LDR	2f	116	2.3	0.005	73.3	68.3	26.3	4.45	18.1	13	3.8	23	4	
LDR	3f	162	7.1	0.002	17.9	17.7	3.0	6.14	19.1	14	4	24	4.2	
LDR	3m	160	4.5	0.003	41.1	34.2	21	6.41	8.2	6	1.8	10	1.8	
BS	4m	131	2.6	0.004	86.3	74.4	33.7	0.54	8.4	6	1.7	11	1.8	
WP	11f	171	3.0	0.004	5.2	9.6	30.4	-4.69	8.4	7	1.6	11	1.6	
NL	8m	166	3.1	0.005	11.6	16	34.1	0.71	8.4	6	2	10	2.1	
NV	5f	168	7.4	0.005	31.4	32.2	14.1	-1.00	17.6	11	5.1	18	5.4	
PV	5f	168	6.2	0.005	31.4	30.7	6.1	-15.53	15.7	12	3.4	19	3.6	
Mean values										11.1	8.8	2.4	14	2.6

LG : Linkage group (m: male map; f: female map)

N: number of individuals genotyped for the flanking markers

P1: position of the highest LOD score from the composite Interval mapping

P2: position of the mean value for maximum LOD score after bootstrap analysis (1000 bootstrap samples)

CI: QTL confidence interval (95%) based on the 1000 bootstrap samples

LOD: highest LOD in the interval

$\alpha_{chr}$  : theoretical p-value at the chromosome level corresponding at the p-value of 5% at the genome level

D: size of effect

P : percentage of the clonal mean variance explained by the QTL

S : percentage of the genetic (additive) variance explained by the QTL

PV : percentage of the phenotypic variance explained by the QTL

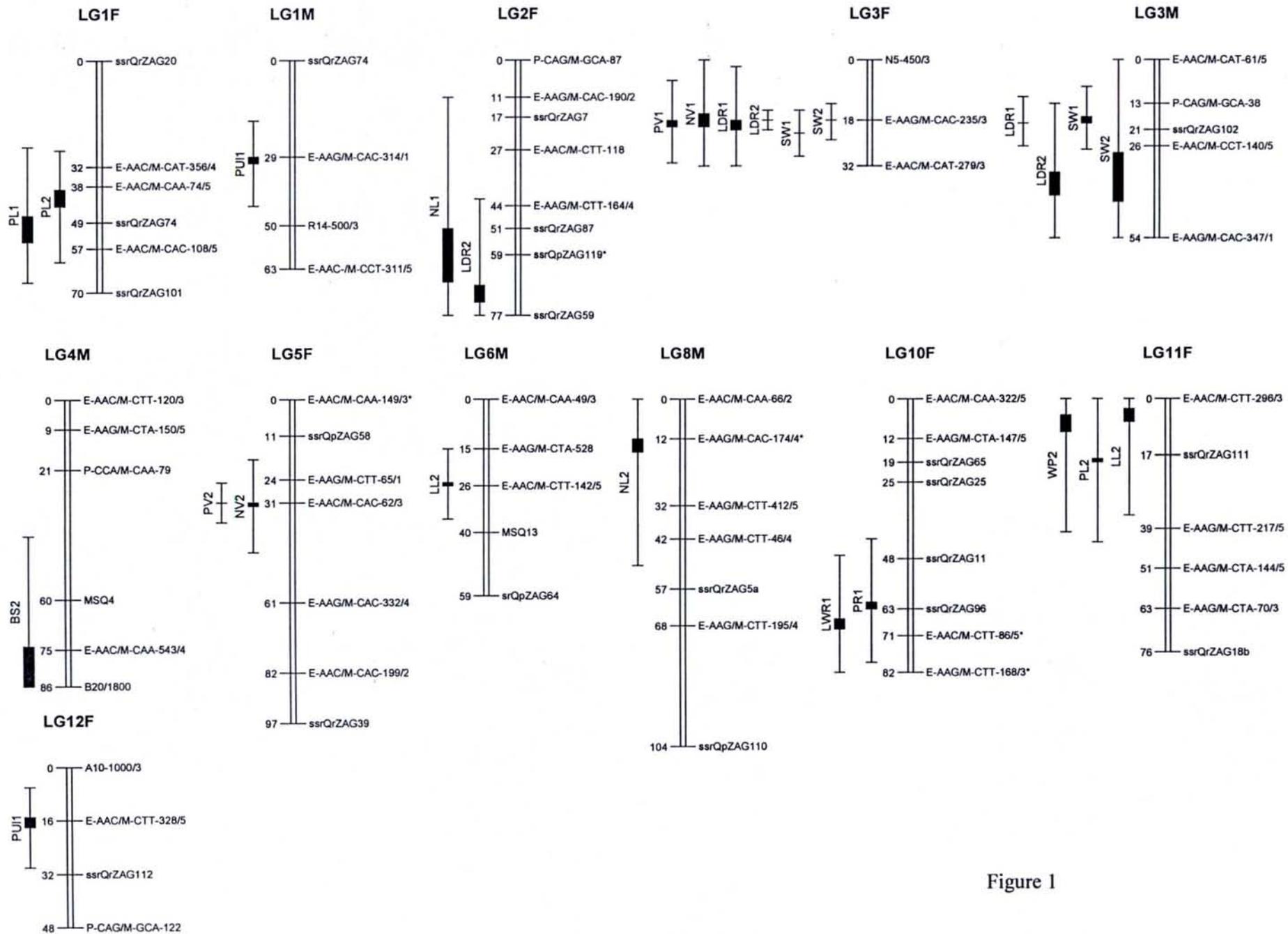


Figure 1

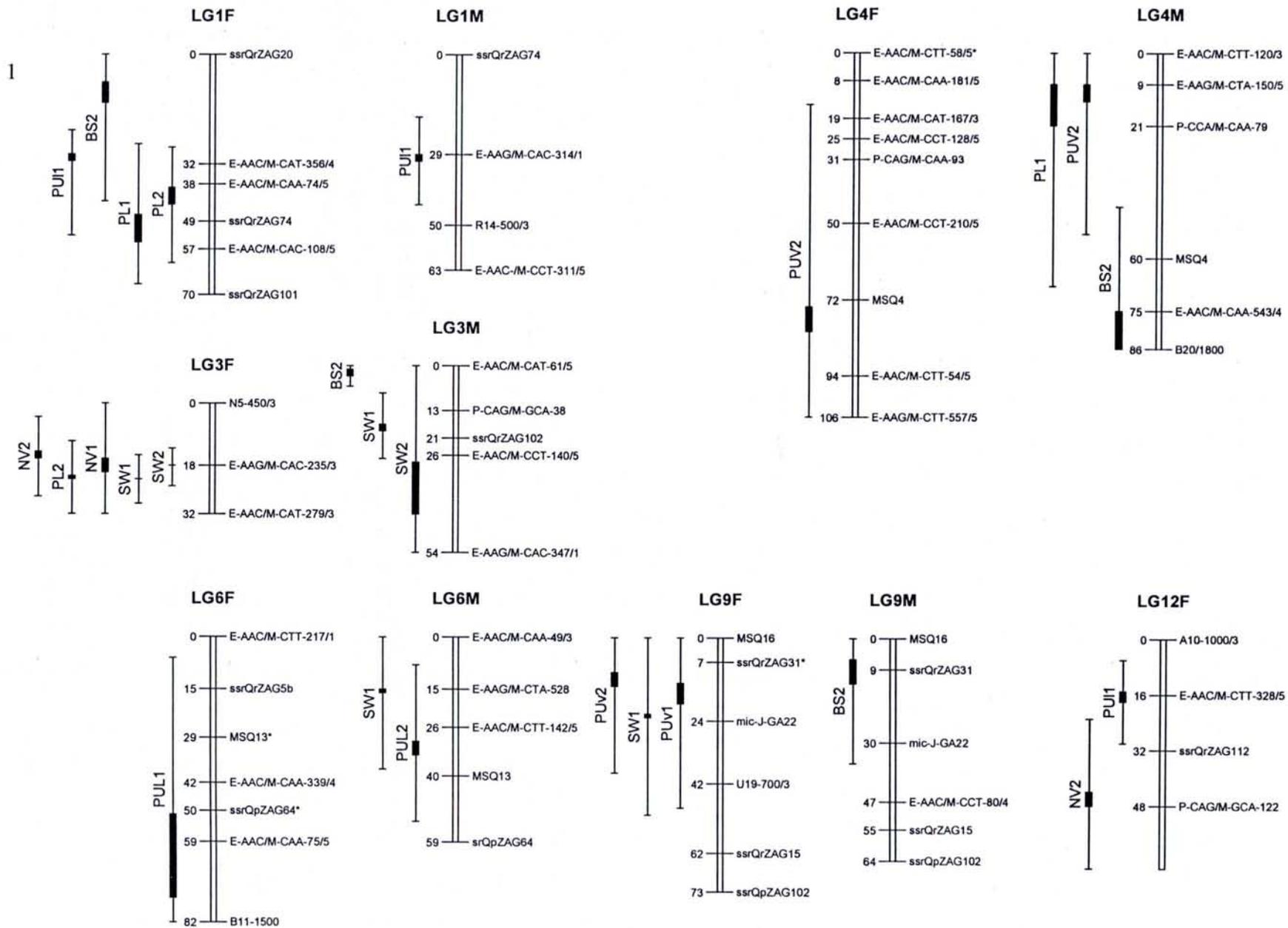


Figure 2

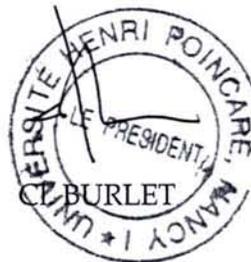
Mademoiselle SAINTAGNE Caroline

DOCTORAT de l'UNIVERSITE HENRI POINCARÉ, NANCY 1  
en BIOLOGIE FORESTIERE

VU, APPROUVÉ ET PERMIS D'IMPRIMER *Thèse N° 835*

Nancy, le *18 juillet 2003*

Le Président de l'Université



**Titre :**

Distribution des régions génomiques différenciant deux espèces proches : le chêne sessile (*Quercus petraea*) et le chêne pédonculé (*Q. robur*).

---

**Résumé :**

Le chêne sessile (*Quercus petraea*) et le chêne pédonculé (*Q. robur*) sont deux espèces inter-fertiles colonisant des espaces communs qui présentent des génomes homogènes tout en maintenant des caractéristiques écologiques et morphologiques spécifiques. Deux approches complémentaires ont été suivies pour étudier l'origine génétique de ces différences phénotypiques. (1) La localisation sur la carte génétique du chêne de QTL (Quantitative trait loci) contrôlant des caractères discriminants les deux espèces et (2) l'estimation, en population naturelle, de la différenciation génétique interspécifique ( $G_{ST}$ ) pour différents marqueurs moléculaires déjà cartographiés. Les deux approches indiquent que les régions impliquées dans la différenciation interspécifique sont dispersées sur l'ensemble du génome avec néanmoins une tendance au regroupement à faible distance. Les différences phénotypiques seraient le résultat de l'effet de plusieurs loci maintenu par une pression de sélection suffisamment forte pour faire face à l'effet homogénéisateur des flux de gènes.

---

**Title :**

Distribution of genomic regions that differentiated two closely related species : sessile oak (*Quercus petraea*) and pedunculate oak (*Q. robur*).

---

**Abstract :**

Sessile (*Quercus petraea*) and pedunculate (*Q. robur*) oak which are two inter-fertile species colonising the same space, display similar genomes, whereas maintain ecological and morphological specific differences. Two complementary approaches have been applied to study the molecular origin of phenotypic differences: (i) Location of QTL, controlling characters which differentiate the two species, in a genetic linkage map of pedunculate oak; (ii) estimation, in natural mixt stand, of the inter-specific differentiation coefficient ( $G_{ST}$ ) for different types of molecular markers, already mapped in the same map. The both approaches suggests that genome regions involved in inter-specific differentiation are dispersed across the genome, although with a tendency to within-linkage group clustering in small distance. Our results suggest that species differentiation is mainly due multi-locus effects maintained by selective pressure strong enough to prevent gene flow from disrupting them.

---

**Discipline :** Biologie Forestière

---

**Mots-clés :**

*Quercus robur*, *Quercus petraea*, différenciation génétique interspécifique, QTL, cartographie,  $G_{ST}$ .

---

**Laboratoire d'accueil :**

INRA- Laboratoire de génétique et Amélioration des Arbres Forestiers  
BP 45, 33 610 Cestas, France.