



**HAL**  
open science

# Variabilité des protéines de l'endosperme du pin maritime révélée par électrophorèse bidimensionnelle : interprétations génétiques, cartographie et relation avec des caractères quantitatifs

Sophie Gerber

► **To cite this version:**

Sophie Gerber. Variabilité des protéines de l'endosperme du pin maritime révélée par électrophorèse bidimensionnelle : interprétations génétiques, cartographie et relation avec des caractères quantitatifs. Sciences du Vivant [q-bio]. Institut National Agronomique Paris-Grignon, 2012. Français. NNT : . tel-02846449

**HAL Id: tel-02846449**

**<https://hal.inrae.fr/tel-02846449>**

Submitted on 26 Aug 2021

**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.

INSTITUT NATIONAL AGRONOMIQUE  
PARIS-GRIGNON  
(INA P-G)

INSTITUT NATIONAL DE LA  
RECHERCHE AGRONOMIQUE  
(INRA)

**THESE**

Présentée par

**Sophie GERBER**  
Ingénieur de L'INA P-G

Pour l'obtention du titre de  
DOCTEUR DE L'INSTITUT NATIONAL AGRONOMIQUE PARIS-GRIGNON

**VARIABILITE DES PROTEINES DE L'ENDOSPERME DU PIN MARITIME  
REVELEE PAR ELECTROPHORESE BIDIMENSIONNELLE :  
Interprétations génétiques, cartographie  
et relation avec des caractères quantitatifs.**

Soutenue le 12 octobre 1992 devant la commission d'examen composée de :

A. GALLAIS, Professeur à L'INA P-G, Président  
D. de VIENNE, Professeur à l'Université Paris XI, Rapporteur  
J-M. ELSER, Directeur de Recherche à l'INRA, Rapporteur  
F. RODOLPHE, Directeur de Recherche à l'INRA  
P. BARADAT, Directeur de Recherche à l'INRA  
A. KREMER, Directeur de Recherche à l'INRA

## **Variabilité des protéines de l'endosperme du pin maritime révélée par électrophorèse bidimensionnelle : interprétations génétiques, cartographie et relation avec des caractères quantitatifs**

### **Résumé**

Ce travail suggère une description du génome du Pin maritime (*Pinus pinaster* Ait.) et envisage l'utilisation de cette description pour l'étude de la variabilité de caractères quantitatifs.

Les performances de croissance des arbres constituant la population de base du programme d'amélioration du pin maritime sont connues grâce à leurs aptitudes générales à la combinaison, évaluées en tests de descendance. Cette étude concerne 18 arbres de cette population, choisis pour couvrir la gamme de variation des performances. La variabilité des protéines contenues dans les endospermes haploïdes (ou mégagamétophytes) de ces pins a été étudiée grâce à l'électrophorèse bidimensionnelle. Une douzaine d'endospermes par arbre ont été analysés et comparés.

À la suite d'une analyse systématique des variations observées, 84 locus sont décrits, impliqués dans des variations de position, de quantité ou de présence/absence des protéines. Des techniques de génétique humaine ont permis de construire une carte qui localise 65 locus sur 17 groupes de liaison et couvre 530 centimorgans au total. Les méthodes d'estimation et de test de la liaison entre deux locus sont discutées. À partir des données de liaison obtenues, la taille du génome du Pin maritime a pu être estimée à environ 2000 centimorgans. Sachant que les pins possèdent de grandes quantités d'ADN, une réflexion sur les différentes descriptions du génome, en terme de cartographie et de quantité d'ADN, est proposée.

La relation entre le poids des mégagamétophytes et les allèles exprimés dans ces organes haploïdes a été étudiée pour chaque locus. L'effet de certains locus a pu être mis en évidence. Le génotype diploïde des pins aux 84 locus a été confronté à leurs performances de croissance grâce à différentes techniques statistiques. Les résultats suggèrent que la variabilité des quantités de protéines pourrait être liée à la variation de caractères quantitatifs. Une réflexion sur l'étude des bases génétiques des caractères quantitatifs, sur l'intérêt et sur les limites de la recherche et de l'utilisation de locus à effets quantitatifs (QTL), notamment dans le cadre particulier de l'amélioration des arbres forestiers, est suggérée.

**Mots-clés :** pin - mégagamétophyte - électrophorèse bidimensionnelle - carte génétique - linkage - taille du génome - QTL |

## **Seed protein variation in maritime pine (*Pinus pinaster* Ait.) revealed by two-dimensional electrophoresis: genetic determinism, mapping and relation with quantitative traits.**

### **Summary**

The present work is intended as a step toward a description of the maritime pine genome (*Pinus pinaster* Ait.). It initiates a study of quantitative traits variation using this description,

The growth performances of the trees composing the breeding population of maritime pine are known thanks to general combining abilities measured in progeny tests. Eighteen trees representing the range of performances were sampled in this population. The variability of the proteins contained in the haploid endosperms (or megagametophytes) of these pines was studied by two-dimensional electrophoresis. An average of 12 endosperms per tree were analysed and compared.

After a systematic analysis of the variation observed, 84 loci were described, responsible for position or amount modifications or presence/absence of proteins. Human genetics techniques were used to build a map. Seventeen linkage groups were detected, which included 65 loci and covered 530 centimorgans. The methods used to estimate and to test for linkage between loci are discussed. The linkage data allowed us to estimate the genome length of maritime pine. It was found to be around 2000 centimorgans. Pines contain one of the greatest quantity of DNA per cell. Some considerations about the relationship between physical and genetic maps in angiosperms and gymnosperms are thus proposed.

The relationship between the weight of megagametophytes and the alleles expressed in these haploid organs was studied for every locus. The effect of three loci was detected. The diploid genotype of 18 pines for the 84 loci was compared to their growth performances thanks to different statistical techniques. The results suggest that protein amount variation could be related to quantitative traits, the usefulness and limits of quantitative trait loci (QTL), particularly for forest tree breeding, are discussed.

**Keywords:** pine - megagametophyte - two-dimensional electrophoresis - linkage map - QTL - genome length

Par sa rigueur et son approche précise des problèmes, François Rodolphe m'a enseigné un usage raisonné des outils biométriques et a contribué de façon fondamentale à l'élaboration de ce travail. Je tiens à le remercier de l'aide qui a été sienne.

La technique biochimique utilisée dans cette thèse a été mise au point par l'équipe du laboratoire de génétique des systèmes végétaux du Moulon. Dominique de Vienne m'a permis de réaliser mes expériences dans ce laboratoire, dans de bonnes conditions tant sur le plan matériel que sur le plan humain. Je voudrais l'en remercier.

Je remercie Jean-Michel Elsen pour sa lecture attentive, critique et constructive.

Je remercie André Gallais d'avoir accepté la direction de cette thèse.

Philippe Baradat a assuré le bon fonctionnement de ce travail et m'a laissé toute liberté pour sa réalisation. Je tiens à l'en remercier.

Les recherches engagées par Antoine Kremer ont été à l'origine de la mise en place du sujet de cette thèse. Il fait aujourd'hui partie du Jury et je l'en remercie.

Nasser Bahrman m'a très souvent aidée et encouragée et je l'en remercie.

Les multiples problèmes informatiques ont été résolus grâce à Thierry Labbe, Patrick Gasqui a répondu efficacement à mes questions, Annie Uteau a recherché les fausses notes avec compétence : je les remercie tous trois.

Mes remerciements vont enfin à tous ceux et celles qui d'une lecture, d'une discussion, d'une phrase, d'un mot ou d'un geste m'ont aidée à achever ce travail.

## SOMMAIRE

<b>INTRODUCTION GENERALE</b>	1
<b>I. MATERIEL ET METHODES</b>	
<b>A. MATERIEL VEGETAL</b>	5
1. Constitution de l'échantillon étudié	5
2. Evaluation des performances de croissance	6
2.1. Mesures en plantations	6
2.1.1. Test de descendance	6
2.1.2. Parc à clones	6
2.2. Mesures sur jeunes plants en conditions expérimentales	7
2.2.1. Stress hydrique	7
2.2.2. Conditions de croissance accélérée	7
<b>B. METHODES D'ETUDES</b>	9
1. Principe de l'électrophorèse bidimensionnelle des protéines dénaturées	9
1.1. Extraction des protéines	9
1.2. Isoélectrofocalisation (IEF)	9
1.3. Seconde dimension	10
1.4. Coloration au nitrate d'argent	10
1.5. Dépouillement des gels	10
2. Etude des variations des spots	11
2.1. Comparaison des variations	11
2.2. Estimation des taux de recombinaison	12
2.3. Utilisation de MAPMAKER	13
<b>II. VARIABILITE DES PROTEINES DE L'ENDOSPERME ET MARQUAGE DU GENOME DU PIN MARITIME</b>	
<b>A. INTRODUCTION</b>	14
<b>B. INTERPRETATIONS GENETIQUES DES VARIATIONS ET REALISATION D'UNE CARTE GENETIQUE (article à paraître dans Theor. Appl. Genet.)</b>	16
1. <i>Summary</i>	16
2. <i>Introduction</i>	17
3. <i>Materials and Methods</i>	17
3.1. <i>Plant material</i>	17
3.2. <i>Protein extraction and electrophoresis</i>	18
3.3. <i>Comparison of the gels</i>	18
3.4. <i>Spot variations and genetic hypotheses</i>	18
3.5. <i>Linkage analysis</i>	19
4. <i>Results</i>	20
4.1. <i>Amount of polymorphism detected</i>	20
4.2. <i>Genetic interpretations</i>	20
4.3. <i>Covariable spots</i>	21
4.4. <i>Genetic mapping</i>	21

5. Discussion	21
5.1. Charge trains	21
5.2. Presence/absence and mobility variation	22
5.3. Covariations of protein phenotypes	22
5.4. Tight linkages	23
5.5. Loci affecting protein amount	23
6. Conclusion	24
<b>C. QUELQUES CAS PARTICULIERS DE VARIATIONS ET LEUR INTERPRETATION GENETIQUE</b>	25
1. Un locus responsable	26
2. Plus d'un locus responsable	28
<b>D. VARIABILITE DES PROTEINES : EXEMPLE DES PROTEINES DE RESERVE</b>	30
1. Gènes des protéines de réserve	31
2. Hérité	32
3. Variations de position	33
4. Allèles nuls et variations de quantité	34
<b>E. ESTIMATION DE LA TAILLE DU GENOME DU PIN MARITIME</b>	36
1. Méthodes d'estimation	36
2. Application aux données du Pin maritime	37
<b>F. CONCLUSION</b>	39
<b>III. ESTIMATION ET TEST DE LA LIAISON ENTRE DEUX MARQUEURS</b>	
<b>A. INTRODUCTION</b>	40
1. Etudes de linkage en génétique humaine	40
2. Etude de linkage chez des espèces végétales pérennes	41
3. Utilisation du logiciel de cartographie	42
<b>B. ESTIMATION DES TAUX DE RECOMBINAISON</b>	43
1. Ecriture de la vraisemblance	43
2. Résultats des maximisations	44
<b>C. DETECTION D'UNE LIAISON ENTRE DEUX MARQUEURS</b>	44
1. Construction du test	45
1.1. Test du rapport de vraisemblances	45
1.2. Test du $\chi^2$	46
2. Comparaison du lod score et du test du $\chi^2$	47
<b>D. DISCUSSION</b>	48
<b>E. CONCLUSION</b>	51

#### IV. CARTOGRAPHIE GENETIQUE ET TAILLE DU GENOME

<b>A. INTRODUCTION</b>	52
<b>B. DESCRIPTIONS DU GENOME</b>	52
1. Comparaison de cartes génétiques	52
2. Taille du génome en terme de quantité d'ADN	54
2.1. Variabilité	54
2.2. Recherche d'une signification adaptative à ces variations	54
2.3. Structure de l'ADN en relation avec sa quantité	55
3. Tailles du génome en centimorgans	55
3.1. Notion de carte saturée	55
3.2. Variabilité des tailles des cartes génétiques	56
3.2.1. Effet du sexe	56
3.2.2. Effet du croisement	57
3.2.3. Variations individuelles	57
3.2.4. Signification de la variation des taux de recombinaison	58
4. Relation entre taille du génome en centimorgans et en paires de bases	59
<b>C. GENOME DES GYMNOSPERMES</b>	61
1. Etudes de liaison chez les gymnospermes	61
2. Quantité d'ADN	62
3. Structure du génome des conifères	62
4. Variabilité des taux de recombinaison	64
4.1. Effet du sexe	64
4.2. Variations individuelles	64
5. Relation entre centimorgans et kilobases	65
<b>D. CONCLUSION</b>	65

#### V. RELATION ENTRE MARQUEURS ET CARACTERES QUANTITATIFS

<b>A. INTRODUCTION</b>	67
1. Sélection des arbres forestiers	67
2. Corrélations entre caractères juvéniles et adultes	68
3. Etude des bases génétiques de la croissance	69
<b>B. ETUDE MULTILOCUS</b>	70
1. Hétérozygotie des arbres	70
1.1. Nombre de locus hétérozygotes	70
1.2. Comparaison des hétérozygoties	71
2. Relation entre l'hétérozygotie et les performances des arbres	73
2.1. Classes de polycyclisme et de vigueur	73
2.2. Relation avec les caractères quantitatifs	73
2.3. Effet de l'hétérozygotie chez les conifères	74
3. Etude simultanée de l'ensemble des marqueurs	75
3.1. Discrimination	75
3.2. Arbres polycycliques et monocycliques	76

<b>C. ETUDE DE L'EFFET INDIVIDUEL DES LOCUS</b>	77
1. Etude de l'effet des locus sur le poids des mégagamétophytes	77
1.1. Méthode	77
1.2. Résultats	78
1.3. Poids de l'endosperme et marqueur	79
2. Relations entre le génotype des arbres et leurs performances	80
2.1. Liaison entre génotype et classes de performances adultes	80
2.1.1. Méthode	81
2.1.2. Résultats	82
2.2 Effet du génotype sur les caractères quantitatifs	82
2.2.1. Méthode	82
2.2.2. Résultats	83
2.2.2.1. Choix d'un niveau de signification	83
2.2.2.2. Répartition des types de locus	83
2.2.2.3. Tri des tests	84
2.2.2.4. Bilan	85
2.2.3 Signification possible des associations	86
<b>D. DISCUSSION</b>	89
1. Les QTL chez des plantes annuelles : intérêts et limites	89
2. Cas particulier des populations panmictiques	92
3. Recherche de QTL chez les animaux domestiques	93
4. Projets chez les conifères	94
<b>E. CONCLUSION</b>	96
1. Place de l'électrophorèse bidimensionnelle	96
2. Perspectives pour les pins	97
 <b>CONCLUSION GENERALE</b>	 99
 <b>REFERENCES BIBLIOGRAPHIQUES</b>	 101
 <b>ANNEXE 1 : Calcul de l'information de Fisher</b>	 122
<b>ANNEXE 2 : Calcul de la puissance du test du <math>\chi^2</math></b>	125
<b>ANNEXE 3 : Relations entre le génotype des arbres et quelques caractères quantitatifs</b>	127

UNIVERSITÄT ZÜRICH

## INTRODUCTION GENERALE

L'idée selon laquelle l'hérédité est associée à des éléments discrets trouve son origine dans les expériences de Mendel qui, en 1865, montre que les caractères très contrastés de deux lignées de pois réapparaissent dans des proportions bien définies à l'issue de l'intercroisement de leurs descendants, sans aucun mélange. La redécouverte de ses travaux intervient vers 1900, et à peu près à la même période, Galton utilise des méthodes quantitatives pour étudier différents aspects de l'hérédité de caractères continus chez l'homme. Cette approche sera développée par Pearson (Bailey 1961). Si des méthodes statistiques nouvelles étaient proposées, les buts poursuivis dans ces études, notamment pour ce dernier auteur, procédaient d'un eugénisme non dissimulé (voir Gould 1984). Une approche similaire chez les végétaux était moins sujette à caution. En 1903 Johanssen s'intéresse à l'hérédité du poids des grains chez des haricots (Gallais 1990). Deux écoles de pensées s'opposent alors, la première s'appuie sur les lois de Mendel, la seconde sur des méthodes biométriques. Fisher suggère la compatibilité des deux approches en 1918 (Bailey 1961). Les variations discontinues et continues transmises par l'hérédité ont en effet en commun les propriétés de ségrégation, de dominance, d'interaction et de linkage, comme le souligne Thoday (1961). Pour cet auteur, la compréhension des caractères continus gagnerait donc beaucoup s'il était possible de les traiter comme les caractères discontinus, en particulier en localisant les facteurs génétiques en cause sur les cartes de liaison. Il propose pour cela l'utilisation conjointe de marqueurs mendéliens et de l'approche biométrique. Ces questions sont toujours d'actualité, puisque, avec le développement des techniques de marquage génétique, les cartes de liaison, et par conséquent une image du génome de plus en plus précise, sont disponibles. Le rapprochement des génétiques quantitatives et mendéliennes, s'il est encore loin d'être atteint, est alors envisageable (Lander et Botstein 1989).

Dans ce contexte, une meilleure connaissance du génome grâce aux cartes génétiques apparaît importante. Si l'on dispose d'un matériel végétal qui se prête à la mise en place, par croisements, de familles génétiquement bien déterminées, la construction d'une carte à partir de ces familles est facilitée. Lorsque l'on s'intéresse à des espèces végétales à longue durée de génération comme les arbres, où l'expérimentation est difficile, la réalisation de cartes est plus critique. Pourtant, une caractéristique de la biologie de la reproduction des gymnospermes, et parmi eux des conifères et des pins en particulier, permet une solution à ce problème.

Chez les angiospermes, la fécondation est double et produit non seulement un embryon diploïde mais aussi un endosperme, tissu nourricier qui entoure l'embryon, généralement triploïde, comportant deux fois l'information du gamète maternel associée à celle du gamète paternel. Si une fécondation similaire a été observée chez des représentants de l'ordre des gnéales, des gymnospermes qui préfigureraient ainsi l'apparition des plantes à fleurs (Friedman 1990, 1991), la fécondation chez les gymnospermes



est toujours "simple" et n'intervient que pour la constitution de l'embryon, diploïde. L'endosperme, ou mégagamétophyte, est issu d'un développement important du gamétophyte femelle et reste donc haploïde, son génotype étant identique à la "moitié" de celui de l'embryon qu'il contient. Le gamétophyte femelle a ainsi deux fonctions, d'une part la production des gamètes femelles et d'autre part la nutrition de l'embryon (Maheshwari et Singh 1967). Le gamétophyte mâle a "envahi" cette dernière fonction chez les angiospermes, et l'endosperme produit est génétiquement plus proche de l'embryon, l'apport maternel restant majoritaire. Des significations évolutives à ces différents types de graines ont été suggérées (Westoby et Rice 1982, Queller 1983, 1984).

L'accès au génotype des gamètes est précieux pour l'étude de la recombinaison. Des études de recombinaisons intralocus ont par exemple été réalisées directement sur des grains de pollen de maïs, à partir de grands effectifs, grâce à une réaction colorée différente des grains portant des mutations à un locus (Nelson 1962, Freeling 1976). Le mégagamétophyte des pins a l'avantage d'être de grande taille. L'électrophorèse bidimensionnelle de l'ensemble des protéines contenues dans cet organe permet alors de détecter de nombreuses variations et d'établir la cartographie génétique des locus correspondants (Bahrman et Damerval 1989).

Cette technique, développée à l'origine par O'Farrell (1975), permet d'analyser des mélanges complexes de protéines extraites en conditions dénaturantes, en les séparant selon leur point isoélectrique en première dimension, et selon leur masse molaire en deuxième dimension. Les protéines sont ensuite révélées par des techniques de coloration non spécifiques. Les deux critères de séparation étant indépendants, la distribution des spots protéiques sur le gel est uniforme. Ces protéines, généralement non identifiées, correspondent a priori à des fonctions variées, elles doivent donc fournir une image du génome assez large (de Vienne et al. 1990). L'électrophorèse bidimensionnelle permet alors d'étudier les variations de l'expression du génome. Damerval et al. (1988) distinguent plusieurs champs d'application de cette technique dans le domaine végétal. Les variations de l'expression du génome sont étudiées lors d'effets induits par l'application de substances hormonales, ou lors de la mise en conditions de stress. La réponse à des pathogènes et la mise en place de symbioses sont également étudiées sous l'angle des profils protéiques. Enfin, des travaux s'intéressent aux modifications de l'expression du génome au cours du développement. Une caractérisation des protéines aux variations intéressantes peut être réalisée à la suite de ces études.

L'étude de la variabilité génétique des protéines grâce à l'électrophorèse bidimensionnelle a donné lieu à un nombre plus limité d'investigations. La divergence entre espèces a par exemple été étudiée par comparaison de profils protéiques en deux dimensions dans le règne animal, chez la drosophile (Spicer 1988), ou les primates (Goldman et al. 1987, Janczewski et al. 1990), et pour les plantes, chez la canne à sucre (Ramagopal 1990) ou le blé (Bahrman et al. 1988, Thiellement et al. 1989). La mesure de la diversité est alors basée sur un grand nombre de produits de gènes. Mais l'une des caractéristiques les

plus intéressantes de l'électrophorèse bidimensionnelle pour le généticien est l'accès aux variations de quantité des protéines qui traduisent l'action de facteurs de régulation. Ces facteurs sont supposés jouer un rôle majeur au cours de l'évolution, du fait du contrôle important qu'ils exerceraient sur les caractères continus (Ayala et McDonald 1980, Hedrick et McDonald 1980, MacIntyre 1982). L'existence de nombreuses variations dans les processus de régulation produirait des phénotypes diversifiés sur lesquels la sélection pourrait jouer (Bush et Paigen 1992). Les résultats obtenus sur le maïs, en utilisant l'électrophorèse bidimensionnelle, vont dans le sens de ces hypothèses puisqu'un effet direct des variations de quantité des protéines sur l'expression du phénotype a été suggéré (Damerval et al. 1987a, Leonardi et al. 1991). Ce type de variations pourrait donc apporter un éclairage nouveau sur les bases génétiques de la variabilité des caractères continus.

Il apparaissait dès lors intéressant de mener une étude de la variabilité des protéines contenues dans les endospermes haploïdes de pins, grâce à l'électrophorèse bidimensionnelle, ces pins ayant par ailleurs été bien identifiés pour leurs performances.

Une espèce comme le Pin maritime se prêtait à une telle approche. Cet arbre couvre une surface de plus d'un millions d'hectares de forêt monospécifique dans les Landes et fournit les volumes les plus importants de l'industrie des bois résineux en France (Durel 1990). Il fait l'objet d'un programme d'amélioration qui a débuté en 1960 par des sélections massales en forêt. Des tests mis en place à la même période permettent à présent une bonne évaluation des performances de croissance des arbres sélectionnés. Un échantillon d'arbres représentatifs de la gamme de variation des performances pouvait alors être constitué pour notre étude.

A partir de cet échantillon, le premier objectif de ce travail a été de tenter une description du génome du Pin maritime. Il s'agissait de réaliser une carte génétique en étudiant les ségrégations de formes protéiques dans les endospermes des différents arbres. Cette démarche présentait deux originalités. La première était liée à l'utilisation de l'électrophorèse bidimensionnelle à des fins de marquage; une définition précise de la notion de locus s'imposait en effet. La seconde provenait de l'emploi de plusieurs arbres sans apparemment pour construire une carte de liaison. Des méthodes adaptées à cette situation ont donc été développées.

Dans ce contexte, deux points méritaient d'être approfondis. La méthodologie de l'étude de la liaison entre marqueurs d'une part, qui peut être abordée de différentes façons d'un point de vue statistique. D'autre part, une approche des notions de cartes génétiques, en liaison avec la taille du génome, sachant que les pins possèdent de très grandes quantités d'ADN.

La caractérisation du génotype des arbres obtenue pouvait alors être confrontée à leurs performances. Des méthodes particulières ont été développées dans ce but. Une réflexion sur l'étude des bases génétiques des caractères quantitatifs, notamment dans le cadre de l'amélioration des arbres forestiers, pouvait ensuite être proposée.

CHAPITRE II

MATHEMATIQUES, NOTION DE LIASON

## I. MATERIEL ET METHODES

### A. MATERIEL VEGETAL

#### 1. Constitution de l'échantillon étudié

Des arbres "plus" sélectionnés en forêt dans les années 1960 constituent la population de base du programme d'amélioration du Pin maritime. Dès leur sélection, des rameaux prélevés sur ces arbres ont été greffés en parcs à clones. Plusieurs copies de ces arbres étaient alors à disposition en plantation. Parallèlement, les graines récoltées sur ces arbres ont été utilisées pour réaliser de premiers tests de descendance (tests de descendance maternelles). Des croisements contrôlés réalisés dans les parcs à clones ont ensuite permis la mise en place de tests de descendance de plein-frères. La croissance en hauteur a été mesurée dans ces tests, soit globalement, soit année par année, à partir de la position des verticilles. Dans ce cas, deux composantes ont pu être distinguées :

- la pousse primaire, correspondant au premier cycle de croissance (d'avril à juillet)
- la pousse secondaire, qui n'existe pas systématiquement, et traduit un deuxième cycle de croissance (de juillet à octobre)

Les arbres "plus" ont pu être classés pour leur vigueur (caractérisée par la hauteur totale) et pour leur aptitude au polycyclisme (caractérisée par la somme des pousses secondaires) grâce aux valeurs moyennes de leurs clones pour ces caractères<sup>1</sup> et grâce à leurs aptitudes générales à la combinaison mesurées dans les tests de descendance<sup>2</sup>, âgés, au moment des mesures, de 12 ans au moins. Ils ont alors été répartis selon trois catégories pour leur vigueur : performances supérieures à la moyenne (+), moyennes (0) ou inférieures à la moyenne (-), et selon leur aptitude au polycyclisme : arbres polycycliques (+) ou monocycliques (-). Dix-huit arbres ont été extraits de cette classification pour couvrir la gamme de variation des caractéristiques de croissance ainsi définies. La constitution de l'échantillon est représentée tableau 1. Nous nous référerons aux arbres en utilisant les numéros, de 1 à 18, indiqués entre parenthèses. Différentes mesures sont disponibles sur ces arbres.

<sup>1</sup> évalués dans le test n°5.44.01

<sup>2</sup> tests n°2.44.1, 2.44.9, 2.44.10, 2.44.16 et 2.44.17

Tableau 1 : Composition de l'échantillon des 18 arbres.

Vigueur	+		0		-	
	+	-	+	-	+	-
Polycyclisme						
Nombre d'arbres	3	3	3	3	3	3
Numéros des arbres "plus"	0014 (1) 3112 (2) 3301 (3)	5301 (4) 5302 (5) 5309 (6)	0016 (7) 0113 (8) 2302 (9)	1312 (10) 0018 (11) 0226 (12)	0147 (13) 0283 (14) 1508 (15)	5306 (16) 3305 (17) 0150 (18)

## 2. Evaluation des performances de croissance

### 2.1. Mesures en plantations

#### 2.1.1. Test de descendance

Parmi les 18 arbres de l'échantillon, 15 sont représentés dans un test de descendance mis en place en 1975<sup>1</sup>. Ce test comprend au total 261 familles de demi-frères issues d'arbres "plus" obtenues par un croisement polycross, chaque clone ayant été pollinisé par un mélange du pollen de 30 arbres. Le dispositif est en blocs incomplets. On dispose, pour 36 à 47 individus des 15 familles issues des arbres de notre échantillon, de la mesure des caractères suivants :

- hauteur totale mesurée à 9 ans (en 1983, H83) et à 13 ans (en 1987, H87),
- circonférence du tronc à 1,30 m à 9 et 13 ans (C83, C87),
- écart à la verticalité basale à 9 ans (ECV83).

Les aptitudes générales à la combinaison (AGC) des arbres-mères pour les 5 caractères ont été estimées selon la méthode des moindres carrés généralisée à des dispositifs non orthogonaux (Searle 1971) en utilisant la librairie de programmes OPEP (Baradat 1980).

#### 2.1.2. Parc à clones

Seize des 18 arbres de l'échantillon sont représentés sous forme de greffes dans un parc à clones<sup>2</sup> réalisé au fur et à mesure des sélections en forêt entre 1962 et 1966. Ces 16 arbres possèdent entre 3 et 19 copies (10 en moyenne). Sur 4 ans (1971-1975) on a totalisé :

- la somme des pousses primaires (SPI),
- la somme des pousses secondaires (SPII).

L'effet année étant important dans le dispositif, les caractères ont été hiérarchisés à l'année dans l'analyse de variance. Les critères de sélection n'ayant pas changé pendant la sélection en forêt, et l'effet éventuel de la provenance des arbres étant absent dans la population landaise, les clones mis en place dans le parc peuvent être considérés comme représentatifs d'une même population de base.

<sup>1</sup> test n°2.44.17

<sup>2</sup> test n°5.44.01

## 2.2 Mesures sur jeunes plants en conditions expérimentales

### 2.2.1. Stress hydrique

Une expérience de stress hydrique a été réalisée sur 16 des 18 familles de demi-frères issues de pollinisations libres en parc à clones<sup>1</sup> par A.Nguyen. Chaque famille était représentée en serre par deux séries de 6 à 12 plants (8 en moyenne) âgés de 2 ans. La première série a bénéficié d'un arrosage normal, alors que la deuxième série a subi une alimentation en eau réduite au moment de la formation du bourgeon déterminant la pousse de l'année suivante. A l'issue de la troisième saison de végétation de ces plants, les moyennes par familles des tailles du premier cycle réalisé (HST) étaient disponibles en conditions normales et stressées (Nguyen, non publié).

### 2.2.2. Conditions de croissance accélérée

Des descendance maternelles, issues de croisements non contrôlés des 18 arbres ont été récoltées dans le parc à clones déjà cité<sup>1</sup>. Onze de ces familles ont fait l'objet d'une expérience en phytotron, en conditions de croissance accélérée, réalisée par M. Lascoux.

Au cours de la première période de croissance, une cinquantaine de plants par famille ont été placés, en chambre de croissance, en condition de lumière continue pendant 9 semaines, les 13 semaines suivantes ayant bénéficié de 16 heures de nuit par jour à basse température pour induire une dormance des plants. La seconde période de croissance, après levée de la dormance, se caractérisait par 9 semaines à 16 heures de nuit par jour, suivie à nouveau d'une induction de dormance (Lascoux 1992).

A la fin de la première période de croissance, des mesures de morphologie ont été réalisées sur les plants :

- longueur totale du plant (LT),
- longueur de la plus longue branche (LLB),
- nombre d'aiguilles secondaires (NAII),
- nombre d'aiguilles primaires, correspondant au nombre total d'entre-nœuds sur le plant (NSU, number of stem unit),
- nombre d'entre-nœuds dans la rosette (dénombrés à l'oeil et au microscope) (ROSET),
- longueur moyenne des entre-nœuds (MSUL, mean stem unit length = LT/NSU).

Ainsi que des mesures de poids secs :

<sup>1</sup> test n°5.44.01

- poids des branches (PB),
- poids des aiguilles primaires (PAI),
- poids des aiguilles secondaires (PAII),
- poids de la rosette (PR),
- poids de la tige (PT).

A l'issue de la deuxième période de croissance, des mesures ont également été réalisées. Au cours de cette période, les plants apparaissent tous polycycliques. Trois cycles successifs ont été considérés individuellement. Pour chaque cycle  $i$  on dispose de mesures morphologiques :

- nombre de branches (NB $i$ ),
- longueur du plant (LT $i$ ),
- nombre d'entre-noeuds (NSU $i$ ),
- nombre d'aiguilles secondaires (NAII $i$ ),
- longueur moyenne des entre-noeuds (MSUL $i$  = LT $i$ /NSU $i$ ).

Ainsi que des mesures des poids secs correspondants :

- poids des branches (PBi),
- poids des aiguilles primaires (PAI $i$ ),
- poids des aiguilles secondaires (PAII $i$ ),
- poids de la tige (PT).

Les plants étaient disposés selon un split-plot. Les aptitudes générales à la combinaison des 11 arbres-mères pour les caractères ont été estimées avec la procédure GLM (General Linear Model) du logiciel SAS (Lascoux 1992).

Le poids de 100 graines ainsi que le taux de germination étaient également disponibles pour l'ensemble des 18 arbres.

L'effet famille était significatif sur tous ces caractères. Au total, nous disposons donc de 49 caractères se référant aux différents arbres de l'échantillon.

## B. METHODES D'ETUDES

### 1. Principe de l'électrophorèse bidimensionnelle des protéines dénaturées

Des électrophorèses bidimensionnelles individuelles ont été réalisées sur les protéines contenues dans les endospermes haploïdes (ou mégagamétophytes) présents dans les graines des 18 pins constituant l'échantillon.

Cette technique particulière d'électrophorèse permet la séparation d'un ensemble de protéines selon leur point isoélectrique, reflet de leur charge, en première dimension, et selon leur masse moléculaire apparente en deuxième dimension.

L'électrophorèse comporte quatre étapes. Les compositions des solutions utilisées à chaque étape sont données tableau 2.

#### 1.1. Extraction des protéines

L'extraction dénaturante de l'ensemble des protéines contenues dans l'endosperme des graines de pin suit un protocole mis au point par Bahrman et Damerval (1989) inspiré par Anderson et al. (1985). La graine utilisée est débarrassée de son tégument, l'endosperme est incisé pour éliminer l'embryon qu'il entoure. Chaque endosperme est alors écrasé dans 6  $\mu$ l de tampon d'extraction (tableau 2) par mg d'endosperme. Le mélange est laissé à température ambiante pendant une heure, puis centrifugé pendant 5 minutes à 10 000g. Le surnageant est prélevé et se conserve à -80°C.

#### 1.2. Isoélectrofocalisation (IEF)

Au cours de l'IEF, sous l'action d'un champ électrique, un mélange de composants amphotères, les ampholytes, établit dans le gel d'acrylamide un gradient de pH à l'équilibre. Les protéines vont migrer dans ce gradient et s'immobiliseront à leur point isoélectrique, où leur charge nette s'annule.

Les gels d'IEF sont coulés dans des tubes de verre de 28 cm de long et 1,2 mm de diamètre. 20 gels d'acrylamide sont réalisés à la fois. Pour la polymérisation, 0.015% de persulfate d'ammonium sont ajoutés à la solution d'acrylamide (tableau 2).

Une quantité de 35  $\mu$ l d'extrait protéique est déposée au sommet de chaque tube. La migration est réalisée avec, à la cathode, de la soude à 50 mM et à l'anode, de l'acide phosphorique à 50 mM. 35 000 Volts heures sont appliqués au total, à raison de 1000 V pendant une heure suivis de 1600 V pendant environ 21 heures.

Tableau 2 : Composition des solutions utilisées pour l'électrophorèse bidimensionnelle.

Tampon d'extraction
-urée 3M -ampholytes (Pharmalytes pH 3-10) 2% -dithiothreitol 1% -FSN-100 (fluorosurfactant, détergent) 4%
Gels d'isoelectrofocalisation
-acrylamide 3.78% -N,N'-méthylènebisacrylamide 0.22% -urée 9.2M -Triton X-100 2% -ampholytes 4% (3% Pharmalytes pH 5-8, 1% Pharmalytes pH 5-6)
Gels de seconde dimension
-acrylamide 10.71% -N,N'-méthylènebisacrylamide 0.29% -saccharose 1% -SDS 0.15M -Tris-Cl <sup>-</sup> pH 8.8, 0.5M
Tampon de migration
-Tris 25mM -glycine 0.2M -SDS 0.1M

### 1.3. Seconde dimension

La seconde dimension est une électrophorèse en gel d'acrylamide en présence d'un détergent, le sodium dodécyl sulfate (SDS). Utilisé à haute concentration, ce détergent se fixe sur les protéines et les charge négativement, de façon environ proportionnelle à leur taille. La mobilité des protéines devient alors fonction de leur masse. Le protocole expérimental utilisé a été mis au point par Damerval et al. (1986, 1987b).

Les gels d'acrylamide sont coulés entre deux plaques de verre et se lient par liaison covalente à un film de Gelbond fixé sur l'une des plaques. Celui-ci facilite la manipulation des gels une fois séchés. Les gels ont une taille de 24cm x 20cm x 1mm. Pour la polymérisation, 0.04% de persulfate d'ammonium, ainsi que 0.07% de TEMED (N,N,N',N'-tétraméthylènediamine) sont ajoutés à la solution d'acrylamide (tableau 2). Les gels d'IEF sont déposés au sommet des plaques, au contact du gel, et recouverts d'une solution d'agarose à 1% pour assurer leur maintien.

Les 20 gels sont placés dans une cuve contenant un tampon de migration (tableau 2), 5 V/cm sont appliqués pendant 16h environ. Du bleu de bromophénol est utilisé pour visualiser le front de migration.

### 1.4. Coloration au nitrate d'argent

Les protéines présentes sur les gels sont révélées par coloration au nitrate d'argent (Granier et de Vienne 1986, Damerval et al. 1987b). La succession des bains, leur composition ainsi que leur durée sont précisées tableau 3.

Après leur coloration, les gels sont laissés dans une solution de glycérol à 2.5% pendant une heure. Recouverts au préalable d'une feuille de cellophane, ils sont alors séchés sous vide.

### 1.5. Dépouillement des gels

Entre 10 et 14 gels par arbre ont été réalisés (12 en moyenne par arbre et 222 au total pour les 18 arbres). La première étape du dépouillement a consisté à comparer tous les gels correspondant à un arbre, en se référant à l'un d'entre eux, un gel "maître", pour noter les différences.

Des niveaux de coloration différents pouvait exister entre des gels ne provenant pas de la même série d'expérience. Pour tenir compte de ces variations un spot était donc toujours considéré relativement à ses voisins.

Tableau 3 : Coloration au nitrate d'argent.

Etape	Composition du bain	Durée
Fixation	ethanol 40%, acide acétique 10%	3 heures
Rinçages	ethanol 10% carbonate de sodium 0.3% carbonate de sodium 0.05%	2 x 10 minutes 2 x 15 minutes 5 minutes
Réducteur	ferricyanure de potassium 0.15% thiosulfate de sodium 0.3% carbonate de sodium 0.05%	1 minute 30
Rinçages	6 bains d'eau desionisée	6 x 10 minutes
Coloration	nitrate d'argent 0.1%	15 minutes
Révélation	carbonate de potassium 4% thiosulfate de sodium 30mM	5 minutes
	carbonate de potassium 4% thiosulfate de sodium 30 mM formaldehyde 0.054%	1 minute 30 13 minutes
Arrêt	acide acétique 1%	10 minutes
Rinçages	5 bains d'eau désionisée	5 x 10 minutes

Tous les spots apparaissant avec deux aspects différents selon les gels du génotype étudié, alternativement présents ou absents, ou avec deux quantités différentes, ont été retenus et repérés par un numéro. Les deux premiers chiffres se réfèrent à une zone du gel, les deux derniers au numéro du spot dans la zone. Un dépouillement final regroupant l'ensemble des gels des différents arbres a permis de retenir les spots aux variations fiables. La détection visuelle de plus de deux classes de quantité pour un même spot étant difficile, seuls les spots présentant deux classes ont été retenus. Des comigrations réalisées en mélangeant en quantités égales deux échantillons déjà analysés du même arbre ou de deux arbres différents ont permis de vérifier les positions relatives de spots proches.

Pour chaque spot repéré variable parmi les gels d'un arbre au moins, la description de tous les gels, c'est-à-dire des 222 gels correspondant aux 18 arbres, a été saisie dans un fichier (figure 1), transformé par la suite pour être analysé (figure 2). Les quelques spots détectés présents/absents parmi les gels de certains arbres et avec des variations de quantités parmi les gels d'autres arbres ont été traités à part.

## 2. Etude des variations des spots

### 2.1. Comparaison des variations

Pour effectuer un premier tri parmi les spots, une comparaison systématique de leurs variations par paire a été effectuée. L'idée était alors de rechercher les spots dont les variations étaient soit complémentaires, soit identiques. Sans tenir compte de l'arbre, les 222 aspects d'un premier spot, x, étaient comparés aux 222 aspects d'un second spot, y. Le nombre de gels dans chaque catégorie était compté :

		spot x	
		0	1
spot y	0	n(00)	n(01)
	1	n(10)	n(11)

En l'absence de données manquantes :

$$n(00) + n(01) + n(10) + n(11) = 222$$

Les couples de spots tels que  $n(10) + n(01) = 0$  étaient parfaitement covariables.

Les couples de spots tels que  $n(00) + n(11) = 0$  présentaient des variations complémentaires et pouvaient être interprétés comme les variants de position d'un même polypeptide. Dans certains cas, plus de deux positions pouvaient exister pour une même protéine à travers les gels des différents arbres. Ces situations correspondaient à des couples de spots tels que  $n(11) = 0$  mais  $n(00) \neq 0$ . Il

Arbre	Spot	Gels :													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1209	1	0	1	1	0	1	1	1	1	1	1	1	1	0
2	1209	1	1	1	1	1	1	1	0	0	1	0	0		
3	1209	0	0	1	1	0	1	0	1	1	1	0	0	0	
4	1209	0	0	0	1	6	1	0	1	1	0	0	0		
etc...															

Figure 1 : Aspect du fichier de saisie.

1 : spot présent ou en grande quantité

0 : spot absent ou en petite quantité

6 : donnée manquante.

Arbre	Numéros des spots :											
	1201	1203	1204	1205	1207	1208	1209	1211	1215	1220	1221	...
1	0	0	0	1	0	0	1	0	1	0	0	...
(première ligne = premier gel de l'arbre n°1)												
etc...												

Figure 2 : Aspect du fichier de travail.

s'agissait alors de tester ce couple de spots par rapport à d'autres spots, qui étaient susceptibles de venir compléter la série. Un polypeptide détecté par exemple avec trois formes alternatives parmi les gels correspondait à un triplet de spots tel que  $n(111) + n(110) + n(101) + n(011) + n(000) = 0$ , en adoptant les mêmes notations que précédemment. La recherche de telles associations, par trois, quatre, et jusqu'à sept spots, a été menée systématiquement à partir des couples de spots "incomplets" tels que  $n(00) \neq 0$ , puis des triplets tels que  $n(000) \neq 0$ , etc...

La construction de tous les tableaux (spot x, spot y) possibles, le repérage, parmi ces tableaux, des situations d'identités ou de complémentarités, ainsi que les tests d'associations entre plus de deux spots, ont été réalisés à l'aide de programmes FORTRAN. A l'issue de cette première étape, les locus supposés responsables des variations observées ont pu être définis. Les étapes suivantes se sont donc appuyées sur un fichier décrivant, pour chaque locus, l'allèle exprimé dans chaque endosperme haploïde des 18 arbres.

## 2.2. Estimation des taux de recombinaison

Le taux de recombinaison entre deux locus pouvait être estimé en utilisant les ségrégations observées parmi les endospermes des arbres hétérozygotes à ces deux locus. Chez l'arbre i, hétérozygote aux locus 1 (A,a) et 2 (B,b), le nombre d'endospermes correspondant à chacun des quatre types gamétiques possibles était comptabilisé :

Arbre i	Locus 2	Locus 1	
		A	a
	B	$n_{11}$	$n_{12}$
	b	$n_{13}$	$n_{14}$

Le nombre total d'endospermes observé chez l'arbre i étant :

$$N_i = n_{11} + n_{12} + n_{13} + n_{14}$$

et le nombre d'endospermes dans une haplophase donnée :

$$n_i = n_{11} + n_{14}$$

Le calcul des valeurs de  $n_i$  et de  $N_i$ , pour chaque couple de locus et chaque arbre double hétérozygote a été réalisé par un programme FORTRAN.



### 2.3. Utilisation de MAPMAKER

Pour réaliser la carte génétique des locus détectés, nous avons utilisé le logiciel MAPMAKER, version  $\beta 2$  sous UNIX, et plus précisément les programmes conçus pour l'analyse de données type CEPH (Centre d'Etude du Polymorphisme Humain) (Lincoln et al. 1990).

Les données type CEPH correspondent à des familles humaines à deux ou trois générations. Notre cas de figure peut être assimilé à un pedigree à deux générations. Un programme FORTRAN nous a permis de constituer, à partir de nos données, un fichier au format type CEPH utilisé par MAPMAKER.

CHIAPIETRE III

WAVELENGTHS OF THE SPECTRUM OF THE SUN  
AND THE SPECTRUM OF THE STARS

## II. VARIABILITE DES PROTEINES DE L'ENDOSPERME ET MARQUAGE DU GENOME DU PIN MARITIME

### A. INTRODUCTION

Quelques travaux se sont intéressés à l'hérédité des modifications de protéines observées grâce à l'électrophorèse bidimensionnelle. Chez le haricot par exemple, Brown et al. (1981) confrontent les profils protéiques des graines de trois cultivars, de trois de leurs croisements F1, et des descendances F2 correspondantes. Ils visualisent ainsi les différences de positions, selon les cultivars, des polypeptides composant deux types de protéines de réserve, et réalisent une étude de liaison génétique des locus à l'origine de ces variations. Des approches similaires existent chez des animaux. A partir de backcross et de lignées recombinantes de souris, Racine et Langley (1980a) décrivent les variations de positions de 9 protéines différentes exprimées dans les reins des animaux. Parmi elles, ils identifient 3 enzymes, et cartographient 7 des 9 locus en cause. La comparaison des profils de parents et d'enfants, de protéines du plasma chez l'homme (Rosenblum et al. 1983) ou de protéines du foie chez la souris (Neel et al. 1985), a permis également de suivre l'hérédité de protéines aux positions différentes, et, dans le deuxième cas, de présences/absences de quelques-unes. Pourtant, dans tous ces exemples, les variations de quantité de protéines, auxquelles l'électrophorèse bidimensionnelle donne accès, ne sont pas évoquées.

L'étude de Bahrman et Damerval (1989) réalisée à partir de 56 endospermes des graines d'un seul pin maritime a permis de montrer qu'il était possible, en comparant les profils en électrophorèse bidimensionnelle des protéines contenues dans ces organes haploïdes, de révéler un grand nombre de variations, touchant non seulement la position des protéines sur le gel ou leur présence/absence mais aussi des variations de quantité des protéines. Les 37 locus correspondant à ces variations ont pu être cartographiés sur 10 groupes de liaison. Donc, d'après cette étude, les cas où les variations de quantités expriment une hérédité monogénique ne semblent pas rares, sachant que l'action des régulateurs en cause est possible à différents niveaux de la synthèse protéique, et selon des processus encore assez mal connus. A partir d'haploïdes doublés d'un croisement de deux lignées d'orge, Zivy et al. (1992) étudient également la ségrégation de formes protéiques différentes en électrophorèse bidimensionnelle. Ils détectent 26 locus, dont un locus responsable de variations de quantités, et cartographient 19 locus sur 7 groupes de liaison.

La réalisation d'une étude similaire, à partir de 18 pins maritimes non apparentés aux caractéristiques de croissance contrastées, devait a priori permettre de révéler un nombre de variations important. Il nous a semblé intéressant, pour cette étude, de souligner le passage entre l'observation de la variabilité des spots protéiques et son interprétation génétique, ce passage n'étant pas toujours immédiat. Il était alors possible de définir avec précision le nombre minimum de locus responsables, et les différents types d'effets des locus sur les protéines. La réalisation de la cartographie génétique des locus décrits a également nécessité l'utilisation d'une technique originale. Ces deux points sont plus particulièrement développés dans l'article en anglais qui suit, et font l'objet de compléments présentés par la suite.

**B. INTERPRETATIONS GENETIQUES DES VARIATIONS ET REALISATION D'UNE CARTE GENETIQUE (article à paraître dans Theor. Appl. Genet.)**

---

**SEED PROTEIN VARIATION IN MARITIME PINE (*Pinus pinaster* Ait) REVEALED BY TWO-DIMENSIONAL ELECTROPHORESIS: GENETIC DETERMINISM AND CONSTRUCTION OF A LINKAGE MAP**

Gerber S.<sup>1\*</sup>, Rodolphe F.<sup>2</sup>, Bahrman N.<sup>1</sup>, Baradat P.<sup>1</sup>

<sup>1</sup> Laboratoire de Génétique et d'Amélioration des Arbres Forestiers, INRA, BP 45, F-33611 Gazinet Cedex, France

<sup>2</sup> Laboratoire de Biométrie, INRA, F-78370 Jouy en Josas, France

**1. Summary**

Proteins from haploid megagametophytes from 18 trees were studied by two-dimensional electrophoresis. A total of 222 seeds, averaging 12 per tree, were analysed individually. One-hundred and fifty protein-spots appeared to segregate in at least one tree. Genetic interpretation were made to define the number of loci responsible for presence/absence, staining differences or position variation of the segregating spots. Complete covariation observed between spots could be the result of either the separation of a single gene product into two or more constituents, very close linkage, or the action of a pleiotropic gene. Human genetics techniques were used to map the 84 putative loci detected. Sixty-five loci were organised in 17 linkage groups, whereas 19 remained unlinked.

**Key words:** pine - two-dimensional electrophoresis - haploid megagametophyte - linkage map.

---

\* To whom correspondence should be addressed

## 2. Introduction

Genetic linkage maps are important tools for studies of genome organization (Polans et al. 1985, Tanksley et al. 1988) and for marker-aided selection (Soller and Beckmann 1983). Restriction fragment length polymorphism (RFLP) and randomly amplified polymorphic DNA (RAPD) provide an unlimited number of markers, and thus saturated linkage maps can be constructed. Maps are now available in an increasing number of plants (Helentjaris et al. 1986, Chang et al. 1988, McCouch et al. 1988, Gebhardt et al. 1991, Landry et al. 1991). The long generation time of tree species provides an impediment to the creation of suitable pedigrees for mapping. Thus, DNA markers are only beginning to be used in tree species, and especially in conifers, for mapping purposes (Neale and Williams 1991, Carlson et al. 1991).

However, the haploid megagametophyte of the seeds of gymnosperms provides a facile opportunity for geneticists to study inheritance and linkage of genes affecting protein phenotypes. Studies concerning isozyme loci are numerous. For example, more than 20 loci were identified in recent studies of *Pinus* (Niebling et al. 1987, Szmidt and Muona 1989), *Cunninghamia* (Geburek and Wang 1990), *Pseudotsuga menziesii* (Adams et al. 1990), and *Thuja* (Xie et al. 1991). The use of two dimensional electrophoresis (2D-PAGE) of the total proteins allows the study of a much larger number of loci than possible with isozyme analysis (Bahrman and Damerval, 1989). The loci detected affect several aspects of protein phenotypes, including presence/absence, modification of mobility and quantity. Quantitative polymorphism is of particular interest, as it has been found to be correlated with quantitative trait variation (Leonardi et al. 1991).

Enlarging the study of Bahrman and Damerval (1989) which was carried out on a seed lot originating from a single maritime pine (*Pinus pinaster* Ait.), the present contribution is based on 2D-PAGE of megagametophytes from 18 trees of the same species. The first aim of this study was to determine the number of loci affecting spot phenotypes. The second aim was to build a genetic linkage map. In order to combine the information provided by the 18 unrelated trees, we used methods developed for mapping in human families.

## 3. Materials and Methods

### 3.1. Plant material

Eighteen trees were sampled from the French breeding population of maritime pine (*Pinus pinaster* Ait.,  $2n=2x=24$ ). Cones were collected on grafts of the trees in a clonal orchard planted near Bordeaux (south-west France). An average of 12 (10 to 14) seeds from each of 18 clones were analysed separately by two-dimensional electrophoresis (a total of 222 seeds).

### 3.2. Protein extraction and electrophoresis

After removing the seed coat and the embryo from each seed, total proteins were extracted from individual megagametophytes following Bahrman and Damerval (1989). The extraction buffer contained 3M urea, 2% ampholytes (pH 3 to 10), 1% dithiothreitol and 4% FSN 100 (detergent) (Anderson et al. 1985). Six  $\mu$ l of buffer were applied per mg of megagametophyte, and 35  $\mu$ l of the resulting homogenate was submitted to isoelectrofocusing in the first dimension followed, in the second dimension, by a sodium dodecyl sulfate electrophoresis performed in polyacrylamide gels bound to Gelbond PAG films (Granier and de Vienne, 1986). Gels were then silver stained (Damerval et al. 1987b) and dried.

### 3.3. Comparison of the gels

The dried gels were compared in a pairwise manner while superimposed on a light box. Due to the small size of seeds, a single gel was generally run per megagametophyte. In a first step, the gels were compared within each genotype, to detect the spots exhibiting apparent segregation. In a second step, the gels from the 18 genotypes, a total of 222 gels, were compared.

To assess the relative positions of spots in different gels of the same genotype, or in different genotypes, pairs of samples were mixed in equal quantity for coelectrophoresis. The resulting gel was then compared with the individual gels from each sample. Each spot was identified by four numbers, the first two digits indicating its approximate position in a given region of the gel, and the second two numbers its number in the region.

### 3.4. Spot variations and genetic hypotheses

The diploid genotype of a tree could be inferred from segregation observed among gametes. With ten gametes per tree, however, segregation distortion could not be tested. When a spot was found with two forms among the gels of a single tree, the tree was assumed to be heterozygote at a locus with two alleles determining the two forms of the spot. Two different types of phenotypes were detected:

- Some spots had two classes of staining intensity. Additional classes may have been present in some cases, but were difficult to assess by eye and were thus discarded.

- Some spots were alternatively present or absent.

Among the present/absence spots, series of spots could be seen that were position variants. These spots were generally close to each other on the gel, and differed by their isoelectric point and sometimes by their apparent molecular mass. A genotype in which two spots of the series were found to be alternatively present but never detected simultaneously on a gel, was interpreted to be

heterozygous at a locus with two alleles determining the two position variants (table 4.A). Additional alleles corresponding to new position variants could be found through genotypes, as well as silent alleles (table 4.B).

To avoid redundancy, we compared the patterns of all pairwise combinations of spots on the 222 gels using FORTRAN programs. Some pairs of spots covaried precisely in their phenotypes. We assumed that a single locus was responsible for the covariation of such pairs, though two tightly linked loci may also have been the cause.

When the covariation in several spots could be explained by the action of a single locus, the spot with the lowest number was usually chosen to give its name to the locus, and written in italics.

### 3.5. Linkage analysis

The linkage map of the loci was calculated using a multipoint linkage analysis technique, based on maximum likelihood, provided by the MAPMAKER computer package (Lander et al. 1987). The 18 trees of our sample are unrelated and thus not heterozygous at the same loci. The linkage between two loci can be studied only in "informative" genotypes, which are heterozygous at these two loci. The number of informative genotypes will vary according to the pair of loci considered. Moreover, the linkage phases of the heterozygous genotypes are unknown, thus a gamete cannot be classified as being in a parental or recombinant phase. Similar situations are found in human genetics, where maps are constructed with several unrelated families with a limited number of children per family (Lander and Green 1987).

We converted the 18 trees and each of their megagametophyte progeny into 18 fictitious human families (table 5). The genotype of the mother of the family was inferred from the segregation observed among megagametophytes. A fictitious father that was recessive homozygote at all loci, was also created. Thus, each gel could be interpreted as providing the maternal genotype of each "child". The human data type option of MAPMAKER was then used to compute our map, using the Haldane function to calculate map distances.

Based on two-point analyses, apparent linkage groups were first constructed authorizing a minimum lod score of 4, and a maximum recombination rate of 0.3. The "best" order of the markers was then determined. The positions of remaining loci were tested relative to these groups. Loci being at least a 100 times more likely to be linked than unlinked were added to the group (Donis-Keller et al. 1987).

**Table 4 :** Examples of mobility variation and its genetic interpretation. A/ Biallelic locus, B/ Triallelic locus and null allele.

A/	Different types of gels observed <sup>a</sup>				
	spot x	spot y			
tree x1	●	∅	or	∅	●
tree x2	●	∅	(all gametes)		
<b>Interpretation :</b> A biallelic locus determines two position variants, spot x (left allele (l)), spot y (right allele (r)).					
tree x1 tree x2	heterozygous genotype l/r homozygous genotype l/l				

B/	spot x	spot y	spot z				
tree x1	●	∅	∅	or	∅	●	∅
tree x2	∅	●	∅	or	∅	∅	●
tree x3	∅	∅	●	or	●	∅	∅
<b>Interpretation :</b> A triallelic locus determines three position variants, spot x (left(l)), spot y (middle(m)), spot z (right (r))							
tree x1 tree x2 tree x3	heterozygous genotype l/m heterozygous genotype m/r heterozygous genotype l/r						
If an additional kind of variation is observed :							
tree x4	∅	∅	∅	or	∅	●	∅
We assume the occurrence of a null allele (0) in this genotype :							
tree x4	heterozygous genotype m/0						

<sup>a</sup> ● : presence of the spot, ∅ : absence.

Table 5 : Creation of a fictitious human family.

Tree x : alleles expressed in the gels for one locus. Spot present (p allele) or absent (a allele).					
gel n°1	gel n°2	gel n°3	gel n°4	gel n°5	gel n°6
p	p	a	p	a	a
Equivalent "human family" : Mother : genotype p/a Father : genotype Y/Y					
Genotype of the children :					
child n°1	child n°2	child n°3	child n°4	child n°5	child n°6
p/Y	p/Y	a/Y	p/Y	a/Y	a/Y

## 4. Results

### 4.1. Amount of polymorphism detected

More than 600 protein spots can be seen on a gel of a haploid megagametophyte, with isoelectric point ranging from 4.5 to 7 and molecular mass from 18 to 94 kDa (figure 3). We found 150 spots which exhibited variation in at least one among the 18 trees studied (figure 4):

- 125 spots were alternatively present or absent,
- 15 spots exhibited staining intensity variation,
- 10 spots exhibited both kinds of variation. <sup>1</sup>

Some spots were organised in "charge trains", including about 2 to 15 spots, with the same molecular mass and different but close isoelectric points. Since they always varied together they were considered as a single spot (see, for example, spot n°1205, 2430, 2402 or 3315 on figure 4).

### 4.2. Genetic interpretations

Among the 125 spots alternatively present or absent, 93 could be interpreted as position variants of the same proteins. Variation in these spots was attributed to 35 polymorphic loci with as many alleles as the number of different positions:

- 19 biallelic loci,
- 12 triallelic loci,
- 2 four-allelic loci including a "silent" allele,
- 1 five-allelic locus including a "silent" allele,
- 1 six-allelic locus.

The allelic products of these loci are presented on figure 5 tied together by a continuous line and by a small-dotted line when an additional silent allele was detected in the corresponding locus.

The variation observed for the remaining 32 present/absent spots and for the 15 spots with two staining intensities were considered to be under the control of 47 biallelic loci.

The pattern of variation of the ten spots expressing both quantitative variation and presence/absence were genetically interpreted, according to the segregation observed within each genotype. If a given protein is detected with two different forms through gels of one genotype, the action of one biallelic locus can be suggested, since the parental genotypes studied here are diploid. Among the ten spots,

some had three forms among the gels of single genotypes. In two cases, we had thus to assume the action of two different loci. In contrast, six cases were interpreted as a result of the action of triallelic loci. These loci associated alleles responsible for quantitative differences, with alleles determining absence or position variants.

A total of 92 loci were identified, accounting for the pattern of variation of the 150 spots.

#### 4.3. Covariable spots

Some spots were found to vary in a precisely identical manner, regardless of tree. Twenty-two spots showed this behaviour (see large-dotted lines in figure 5). Groups of covariable spots were explained by the action of a single locus.

A total of 84 loci were retained for the linkage analysis.

#### 4.4. Genetic mapping.

MAPMAKER estimated the presence of 17 linkage groups that included 2 to 8 loci, whereas 19 loci remained unlinked (figure 6). The 65 mapped loci covered 530 cM.

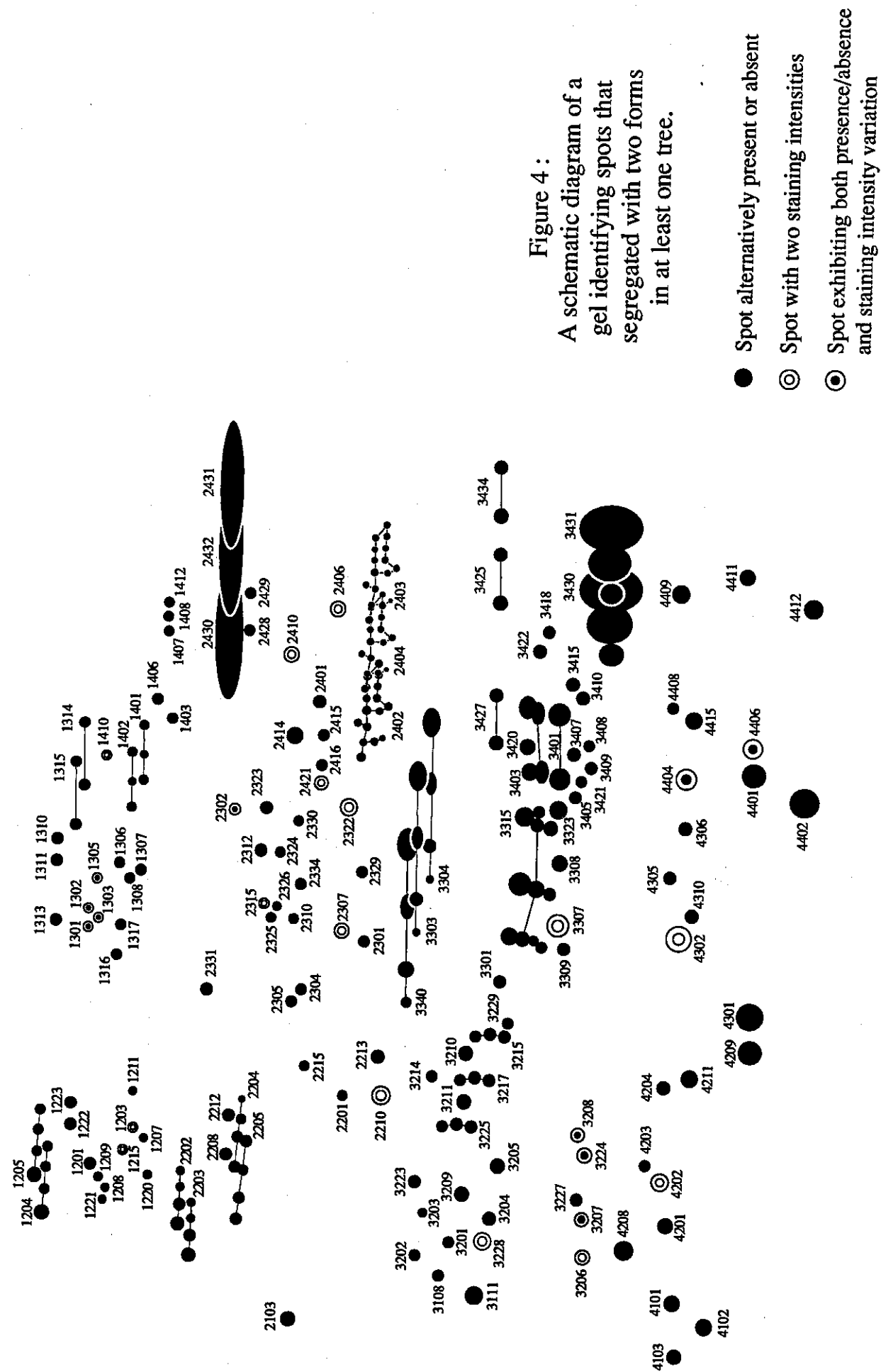
### 5. Discussion

#### 5.1. Charge trains

Two-dimensional electrophoresis of megagametophytes revealed several spots associated in charge trains which covaried as single spots. These charge trains could be either due to post-translational modifications of a protein, or to differences existing in the structural genes. On two-dimensional electrophoresis gels, storage proteins result in patterns similar to the charge trains described here. They are usually constituted by the assembly of subunits, themselves formed by association of several heterogeneous polypeptides, differing by isoelectric points and/or molecular mass (Higgins 1984, Vedel and Delseny 1987). Explanation of this heterogeneity was found at the DNA level, and no case of post-translational modifications were described. In maize, the heterogeneity of one zein fraction was due to the sequence variation in a tandem repetition (Heidecker and Messing 1986). In rye, among 42 repeats for prolamin, only 5 sequences were found identical to the consensus, the other differing as a result of amino-acid substitution and/or deletions (Hull et al, 1991). The spots organised in charge trains on our gels were almost all found with two or three position variants through the gels of the different genotypes (figure 5).



Figure 3 : Two-dimensional electrophoresis gel of proteins extracted from the haploid megagametophyte of maritime pine.



## 5.2. Presence/absence and mobility variation

The presence/absence variation observed for some spots on our gels can have two interpretations. The disappearance of a spot can be due to a "silent" allele, preventing the expression of the protein. However, the absence of a spot may be the result of a decrease in protein quantity behind the level of detection. A quantitative variation would then be interpreted as a presence/absence variation.

The interpretation of mobility variation of spot is less ambiguous. This phenomenon is commonly observed in plants as well as in animals (wheat (Zivy et al. 1983), maize (Leonardi et al. 1987, 1988), and primates (Goldman et al. 1987)). Most of the position variants observed for the proteins present in the megagametophyte of maritime pine corresponded to horizontal shifts, and thus to an isoelectric point modification of the protein. Apparent molecular mass modifications were less frequent (figure 5). The presence, at the same locus, of alleles responsible for position variants and alleles causing the absence of the protein was observed in four cases (loci n°1305, 2305, 2201, and 4209 on figure 5). A combination of two alleles corresponding to different sizes of the protein and a null allele at the globulin-1 gene in maize has been described at the molecular level (Belanger and Kriz 1991). The production of a larger protein was due to an insertion in an exon, whereas another insertion caused a translational frameshift and resulted in a null allele. Similar mechanisms may cause the polymorphism of pine seed proteins discussed here.

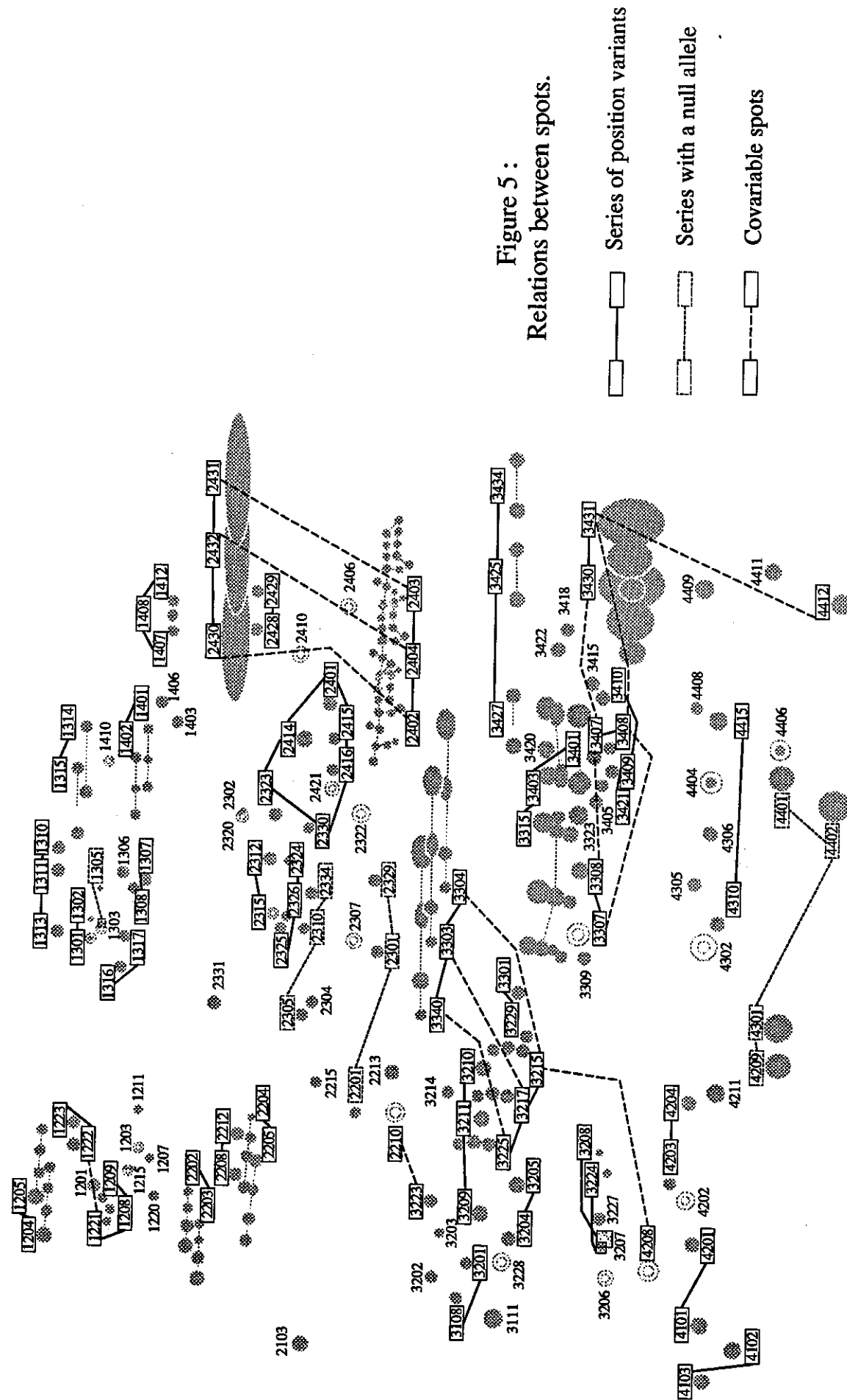
## 5.3. Covariations of protein phenotypes

Several spots had perfectly correlated patterns of variation over all the 222 gels compared (figure 5). Three hypotheses can be proposed to explain this:

-The two proteins come from a single polypeptide, post-translationally modified.

For instance, the left, the middle or the right position variant of one protein ( spot n°2430, 2432, 2431, figure 5) was present on the gels simultaneously with the left, middle or the right position variant of a different protein (spot n°2402, 2404, 2403). The same situation was found with spot n°3340, 3303, 3304 and n°3225, 3217, 3215. Both pairs of triplets correspond to the most abundant proteins of the gels, and probably have a storage function. Each pair may represent large and small subunits of storage proteins, which often result from a post-translational cleavage of a unique polypeptide, assembled afterward by a disulphide bound (Higgins 1984, Collada et al. 1991). According to Allona et al. (1992) glutelins of maritime pine have a dimeric structure. The two subunits, joined by a disulphide bridge, are heterogeneous and are both basic. Spots n°2430 (or 2432, 2431) and 2402 (or 2404, 2403) appear to fit this model.





- A single pleiotropic gene controls the expression of several different proteins. This example was illustrated (with 2D-PAGE) by Gottlieb and de Vienne (1988) for pea. Yeast cells transformed by a single multicopy plasmid revealed unexpected amplification of several proteins on 2D-PAGE, suggesting also the pleiotropic action of a single gene (Thoraval et al. 1990).

- The modifications of the two covariable spots are due to polymorphism at two tightly linked loci. As a consequence, for a heterozygous tree, only parental gametes are observed since the size of our sample was too small to detect rare recombinants. However, the covariations we observed are independent of the tree considered: a given form of a first spot was strictly associated with that of a second spot in all 222 gametes. If the hypothesis of the presence of two linked loci is correct, the 18 trees would be in the same phase (complete linkage disequilibrium) for all the pairs of loci studied.

#### 5.4. Tight linkages

A set of loci affecting proteins of about the same molecular mass but different isoelectric point were found to be tightly linked (n°4302, 4310, 4306 and 4404 on linkage group n°5). In the present case, the phases of the 18 genotypes are different for these loci. An additional locus (n°4211), separated by 3.6 cM from this group, was responsible for the modification of a protein located on the same horizontal line further away on the gel. Several pairs of loci responsible for the variations of proteins with same apparent molecular mass and close isoelectric points were also found to be linked (n°4406-4411, 4408-4409, 3418-3315, 3422-3420, 3307-3309, or 3111-3228).

We also noticed strong linkage between loci affecting spots with no obvious resemblance on the gel, and that exhibited different type of variation (n°3203 and 1305 on linkage group n°1, n°1401 and 2320 on group n°4, n°2103 and 2323 on group n°9, n°2204 and 3214 on group n°14). This corroborates observations made by Bahrman and Damerval (1989), though with loci affecting different spots than we observed.

#### 5.5. Loci affecting protein amount.

The loci responsible for intensity modifications of spots appeared to be dispersed through the linkage groups (figure 6). Leonardi et al. (1987) observed that the loci affecting protein amount could be either closely linked or distant from the structural genes of the affected protein. Bahrman and Damerval (1989) found two loci, the first responsible for the mobility variation and the second for the quantitative variation of the same spot, on two different linkage groups. In our data, the same situation was found in one case, with loci n°1303 and n°1305. In contrast, loci n°2320, affecting the amount, and n°2302, determining the presence or absence of the same spot, were located in the same linkage group, but were separated by 27.8 cM.

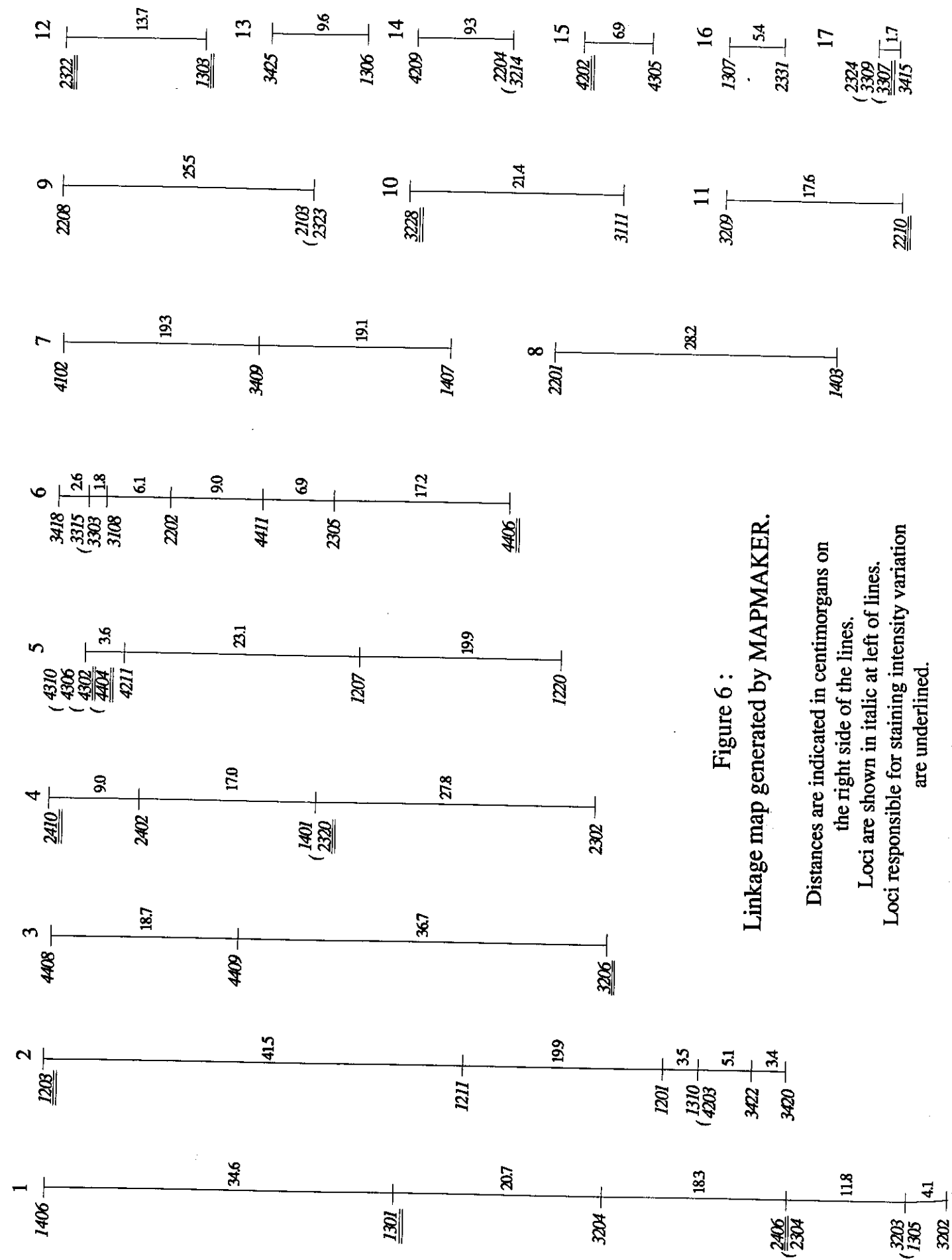


Figure 6 :  
Linkage map generated by MAPMAKER.

Distances are indicated in centimorgans on the right side of the lines.

Loci are shown in italic at left of lines.

Loci responsible for staining intensity variation are underlined.

### 6. Conclusion

Two dimensional electrophoresis of haploid megagametophytes of a pine allowed a large number of protein polymorphisms to be detected. The number of loci causing this variation was estimated. The present study confirmed and enlarged a first investigation of Bahrman and Damerval (1989) made with seeds of a single tree. We also demonstrated that mapping techniques developed for human genetics could be easily applied in a study dealing with several unrelated progenies of gymnosperm trees.

The relationship between allozyme genotype and quantitative traits in conifers has been investigated by several authors (El-Kassaby 1982, Ledig et al. 1983, Strauss 1986). However, no general conclusion can be derived from the different studies, which all included a limited number of loci. Recent data derived from two-dimensional electrophoresis in maize (Damerval et al. 1987a, Leonardi et al. 1991) indicated that protein amount polymorphism could explain part of the phenotypic variation of quantitative traits. The relationship between the growth potential of the 18 trees, evaluated on progeny tests, and their genotype for the genetic markers described is currently under investigation.

## COMPLEMENTS

La comparaison des profils protéiques d'une dizaine de mégagamétophytes haploïdes pour chacun des 18 pins étudiés a donc permis de mettre en évidence les variations de certains spots, d'établir le nombre de locus responsables, et de les cartographier. L'endosperme haploïde des pins rend particulièrement facile l'étude de la variabilité des protéines qu'il contient. Pourtant, l'interprétation génétique de quelques variations a nécessité un traitement particulier, qui est présenté dans un premier chapitre. L'identification de certaines protéines de l'endosperme, en particulier des protéines de réserve, vraisemblablement très représentées dans nos gels, et l'étude fine de leurs modifications pourrait permettre de préciser la nature des variations observées. Nous avons développé dans un deuxième chapitre quelques réflexions sur ces protéines, sur leurs particularités chez les angiospermes et sur les quelques connaissances acquises chez les gymnospermes à ce sujet. Nous suggérerons pour finir une estimation de la taille du génome du Pin maritime à partir des résultats de liaisons génétiques établis dans cette étude, ainsi que dans celle de Bahrman et Damerval (1989).

### C. QUELQUES CAS PARTICULIERS DE VARIATIONS ET LEUR INTERPRETATION GENETIQUE

Comme nous l'avons vu, lorsqu'un spot présentait deux formes ségrégeant parmi les gels d'un arbre, on pouvait supposer l'existence d'un locus dont les deux allèles étaient responsables de ces deux formes : présence ou absence, deux positions ou deux quantités. Certaines variations présentaient pourtant des aspects plus compliqués et des interprétations génétiques particulières ont été nécessaires pour en rendre compte. Six cas particuliers ont été rencontrés, qui impliquent 14 spots repérés sur les figures 4 et 5. Les différents types de gels observés parmi les gamètes de différents arbres sont présentés, et des hypothèses génétiques sont suggérées.

#### 1. Un locus responsable

1)	Types de gels observés			
Spots n° ou n°	3307 2315	3308 2312		
Arbre x1	●	∅	ou	● ●
Arbre x2	uniquement :			● ∅
Arbre x3	uniquement :			● ●

#### Interprétation :

Un locus biallélique induit une variation d'intensité du spot le plus à gauche et la présence/absence simultanée du spot de droite, l'arbre x1 est hétérozygote à ce locus, les arbres x2 et x3 sont les deux homozygotes possibles. Ce phénomène pourrait également être expliqué par une coïncidence, le variant de position du spot de droite aurait des caractéristiques proches du spot de gauche, invariant.

2)	Types de gels observés			
Spots n°	1301	1302		
Arbre x1	●	∅	ou	● ∅
Arbre x2	●	∅	ou	∅ ●
Arbre x3	uniquement :			● ∅

#### Interprétation :

Un locus triallélique explique les deux intensités du spot gauche (allèles  $q_1$  et  $q_2$ ), ainsi qu'un variant isoélectrique (allèle  $v_2$ ).

L'arbre x1 a le génotype ( $q_1, q_2$ )

L'arbre x2 a le génotype ( $q_2, v_2$ )

L'arbre x3 est homozygote ( $q_2, q_2$ )

3)	Types de gels observés		
Spot n° ou n°	4404 4406		
Arbre x1	●	ou	●
Arbre x2	●	ou	∅

Interprétation :

Un locus triallélique explique les deux intensités du spot (allèles  $q_1$  et  $q_2$ ) ou son absence (allèle a).

L'arbre x1 a le génotype ( $q_1, q_2$ )

L'arbre x2 a le génotype ( $q_2, a$ )

4)	Types de gels observés					
Spots n°	3207	3224	3208			
Arbre x1	●	∅	∅	ou	●	●
Arbre x2	●	∅	∅	ou	∅	●

Interprétation :

Les spots 3207 et 3224 covarient comme dans le cas 1), et les spots 3207 et 3208 sont des variants de position. Un seul arbre de type x2 a été observé. Dans cet exemple, on ne peut savoir si un seul locus ou deux locus sont en cause, l'étude de génotypes supplémentaires serait nécessaire. En l'absence de ces données, il a été décidé de ne considérer qu'un seul locus.

## 2. Plus d'un locus responsable

5)	Types de gels observés				
Spot n°	2302				
Arbre x1	●	ou	●	ou	∅
Arbre x2	●		ou		∅
Arbre x3	●		ou		∅

Interprétation :

Un premier locus biallélique est responsable des deux intensités du spot (allèles  $q_1$  et  $q_2$ ).

Un second locus biallélique est responsable de la présence/absence (allèles p et a).

L'arbre x1 est hétérozygote aux deux locus.

L'arbre x2 est homozygote au premier locus (génotype ( $q_2, q_2$ )) et hétérozygote au second (génotype ( $a, p$ )).

Sous cette hypothèse, lorsque l'allèle nul s'exprime, on ne peut pas connaître l'allèle présent au locus responsable des variations d'intensité. On risque de ne pas détecter l'hétérozygotie à ce locus car le nombre de gels observé sera réduit par la présence de l'allèle nul.

6)	Types de gels observés					
Spots n° :	1303	1305				
Arbre x1	∅	∅	ou	●	∅	ou
Arbre x2	∅	●	ou	●	∅	ou

## Interprétation :

Un locus biallélique est responsable des deux intensités du spot 1303 (allèles  $q_1$  et  $q_2$ ).

Un locus triallélique explique les deux variants de position, le spot 1303 (correspondant à l'allèle  $v_1$ ), et le spot 1305 (allèle  $v_2$ ), ou leur absence (allèle  $a$ )

L'arbre x1 est hétérozygote au premier locus (génotype  $(q_1, q_2)$ ), et hétérozygote au second locus (génotype  $(a, v_1)$ ).

L'arbre x2 est hétérozygote au premier locus (génotype  $(q_1, q_2)$ ), et hétérozygote au second locus (génotype  $(v_1, v_2)$ ).

Le même problème que dans le cas précédent se pose. On ne connaît pas l'allèle au premier locus quand l'allèle nul est présent au second locus.

Ces quelques exemples montrent que si l'interprétation génétique des variations observées est généralement immédiate, elle est plus délicate dans certains cas. Malgré cette difficulté, parmi les 6 cas présentés ici, seules 2 séries de spots correspondent à des locus qui n'ont pu être cartographiés, les spots 2312 et 2315 (cas 1) et les spots 3207, 3208 et 3224 (cas 4). Lorsque deux locus sont responsables des variations observées, on peut supposer que deux modificateurs agissant sur une même protéine sont révélés (Klose 1991).

**D. VARIABILITE DES PROTEINES : EXEMPLE DES PROTEINES DE RESERVE**

Les protéines des graines sont distinguées selon leur solubilité et quatre catégories existent. Les albumines sont solubles dans l'eau et comprennent essentiellement des enzymes. Les globulines sont solubles dans des solutions salines alors que les prolamines sont solubles dans des solutions alcooliques. Ces deux catégories représentent des protéines de réserve. Enfin, les glutélines sont solubles en solutions acides ou alcalines et sont probablement des protéines de structure ou de réserve, mais pourraient également être impliquées dans des fonctions métaboliques (Cooke 1988). La distinction des fonctions de ces protéines n'est pas toujours claire : dans les grains de seigle, par exemple, la  $\beta$ -amylase est supposée fonctionner aussi bien comme une enzyme que comme une protéine de réserve (Daussant et al. 1991).

Les protéines de réserve sont donc un constituant important de la graine. Elles sont situées dans les corps protéiques, organites résultant de la modification du réticulum endoplasmique, et sont hydrolysées au moment de la germination. Sur nos gels, on peut supposer qu'elles correspondent aux spots de taille importante et avec un aspect en "train de charge" caractéristique de ce type de protéines visualisées en électrophorèse bidimensionnelle. Anderson et al (1985) supposent que les modifications intervenant dans des gènes dérivés d'un gène ancêtre expliqueraient cet aspect des spots. Ces spots apparaissent très variables dans notre étude, et souvent covariables. De plus les locus responsables de certaines de leurs modifications ont pu être cartographiés.

Certaines protéines de réserve des angiospermes et des gymnospermes présentent des similitudes et sont supposées avoir un ancêtre commun. La séquence en acides aminés d'une partie d'une protéine de *Gingko biloba*, dont les caractéristiques électrophorétiques étaient similaires à celles des légumineuses d'angiospermes (Jensen et Berthold 1989), est apparue proche de séquences de protéines de citrouille, de colza et de pois (Häger et al. 1992). Sur la base de comparaison de séquences d'acides aminés et d'immunodétection, des homologies ont été également trouvées entre les glutélines du Pin maritime et les protéines de réserve du coton, du soja et du pois (Allona et al. 1992). Le même type d'observations chez l'épicéa amène Newton et al. (1992) à supposer que ces protéines ont été conservées depuis la divergence des angiospermes et des gymnospermes. Les protéines de réserve des angiospermes étant très étudiées, on peut alors supposer que certaines de leurs caractéristiques se retrouvent chez les gymnospermes, encore peu connus sur ce point.

### 1. Gènes des protéines de réserve

Les protéines de réserve présentes dans les graines sont codées par des familles multigéniques de tailles variables. Elles sont composées de différentes sous-unités elles-mêmes constituées de plusieurs polypeptides de masses moléculaires et de points isoélectriques légèrement différents (Higgins 1984, Vedel et Delseny 1987). De ce fait elles apparaissent, avec l'électrophorèse bidimensionnelle, comme une association de spots proches quelles que soient les espèces étudiées, par exemple chez le pois (Casey 1979), le blé (Anderson et al. 1985), le hêtre ou le chêne (Collada et al. 1991a, b).

L'organisation des gènes de ces protéines en familles pourrait s'expliquer à plusieurs niveaux. Les crossing-over inégaux et les conversions géniques qui ont lieu dans les familles multigéniques, facilités par le faible nombre ou l'absence d'introns, seraient à l'origine d'une certaine homogénéisation des séquences. Chez la zéine, par exemple, où l'apparition de codons stop est fréquente, un nombre de gènes actifs suffisant serait ainsi maintenu (Heidecker et al. 1991). L'amplification et le regroupement des gènes permettraient une grande production de la protéine correspondante et l'expression coordonnée de l'ensemble des gènes (Polans et al. 1985). Pour les protéines de réserve, nécessaires en grande quantité, cette hypothèse paraît vraisemblable. Pourtant, les mécanismes de régulation des familles multigéniques sont mal connus. On ignore si les membres d'une même famille sont exprimés simultanément ou non (Vedel and Delseny 1987). Les séquences répétées, dont on a parfois suggéré un rôle de régulateur, ne semblent pas être impliquées dans la régulation des protéines de la graine chez le soja (Fisher and Goldberg 1982). La suppression des introns dans un gène de phaséoline (protéine de réserve du haricot), n'a pas affecté la quantité de protéine détectée (Chee et al. 1991). Des mécanismes de régulation traductionnels et post traductionnels ont été mis en évidence dans l'expression des protéines des graines mais n'ont pas été plus précisément décrits (Walling et al. 1986, Chappel et Chrispeels 1986).

Les données concernant les protéines de réserve des gymnospermes sont beaucoup plus réduites que celles des angiospermes. Les études consacrées à ce sujet s'attachent à décrire les corps protéiques, ainsi que les différents types de protéines qu'ils contiennent caractérisées par leur masse moléculaire et leur solubilité (Juo et Stotzky 1970, Gifford 1988, Green et al. 1991). Des approches plus dynamiques s'intéressent à l'évolution de l'expression de ces protéines lors de l'embryogenèse, de la germination, et comparent les différentes étapes et organes. Des similitudes sont à remarquer entre les différentes espèces étudiées (Misra et Green 1990, 1991, Flinn et al. 1991a, b, Gifford et al. 1991, Groome et al. 1991). Ainsi, des gels d'électrophorèse bidimensionnelle réalisés sur graines d'épicéa (Hakman et al. 1990) apparaissent très similaires aux nôtres pour les protéines de réserve, présentes en grande quantité. Trois protéines majoritaires décrites par ces auteurs (de 42, 28 et 22 kD) ont toutes les caractéristiques de 3 spots dominants de nos gels (spot n°2430, 3303 et 3430 de la figure 4).

### 2. Hérité

Les gènes des protéines de réserve présentent des types d'hérité et d'organisation génomique très diverses selon les espèces.

Dans certains cas, les gènes impliqués sont regroupés, par exemple chez des sécalines du seigle (Shewry et al. 1986), des gliadines du blé (Metakovsky et al. 1984), ou chez la phaséoline du haricot (Vallejos et Chase 1991). Dans d'autres cas, plusieurs regroupements de gènes sont localisés sur différents chromosomes, ou différents gènes sont simplement séparés dans le génome. Ces situations sont rencontrées chez la glycinine du soja (Fisher et Goldberg 1982), la légumine du pois (Matta et Gatehouse 1982) ou de la fève (de Pace et al. 1991) et la zéine du maïs (Spena et al. 1983, Heidecker et Messing 1986).

Cette diversité trouve une explication dans la variété des tailles des familles multigéniques. Lorsqu'une famille est de petite taille, elle peut être considérée comme un seul gène. Lorsque la famille s'amplifie, elle apparaît comme un regroupement de gènes ("cluster"). Lorsqu'elle se disperse, par des mécanismes supposés apparentés aux transposons (Spena et al. 1983), les zones du génome impliquées se multiplient. Par conséquent, les familles de petites tailles (moins de 10 membres) présentent des types d'hérité simples qui deviennent nettement plus élaborés lorsque leur taille augmente (Ohta et Dover 1984).

La variabilité et l'hérité des protéines de réserve n'ont pas été étudiées à notre connaissance chez des gymnospermes. Dans notre expérience, quatre locus impliqués dans les variations des spots les plus importants des gels ont pu être cartographiés. Le locus n°2402 est localisé sur le groupe de liaison n°4, le locus n°3307 sur le groupe n°17 et les locus n°3315 et 3303 très fortement liés sur le groupe de liaison n°6 (figure 6). Ces locus sont principalement associés à des variations de position des protéines en question. Nous ne pouvons en déduire une quelconque hypothèse quant à l'organisation des gènes des protéines. En effet, il ne nous est pas possible de savoir si les modifications observées correspondent à des changements du gène de structure, ou si elles interviennent après la traduction, et sont provoquées par un gène distinct du gène de structure. La comparaison des produits de traduction in vitro de deux endospermes présentant des variants de position différents serait un moyen de le préciser.

Il nous a semblé intéressant de tenter d'établir un lien entre la description fonctionnelle de protéines de réserve des pins et des covariations génétiques observées pour quelques protéines de nos gels. Gifford (1988) décrit chez *Pinus monticola* une protéine de réserve insoluble de 50 kD qui, à l'état réduit, correspond à deux protéines de 20 et 30 kD. Cette protéine a les mêmes caractéristiques que la glutéline décrite par Allona et al. (1992) chez le Pin maritime. Les deux protéines correspondent aux deux sous-unités de la glutéline, basiques et très hétérogènes. Elles sont liées par un pont disulphure

lorsqu'elles ne sont pas réduites. On peut faire l'hypothèse que, comme pour de nombreuses protéines de réserve des angiospermes, ces deux sous-unités sont issues d'un seul polypeptide, clivé après la traduction pour donner les deux sous-unités (Higgins 1984).

Nous supposons que la série des spots 2430, 2431 et 2432 correspond à la petite sous-unité et la série des spots 2402, 2403 et 2404 à la grande. Ces spots sont eux-mêmes composés de petits spots qui diffèrent par de petites modifications de charge et de masse : Allona et al. (1992) observent une dizaine de petits spots dans la petite sous-unité et une quinzaine dans la grande, qui traduisent l'hétérogénéité de ces protéines. Nous avons observé que ces deux protéines composites subissent, de façon parallèle, une modification globale de la charge. Selon les allèles présents au locus responsable de cette modification, on observe alternativement les spots 2430 et 2402, 2432 et 2401 ou 2431 et 2403 (figure 5). Si les deux protéines correspondantes étaient issues, comme nous le supposons, d'un même gène de structure, et participaient à la même fonction, leur réactions identiques à un modificateur, dont la nature et le mode d'action (vraisemblablement post-traductionnel) restent à élucider, serait expliquée. Il nous semble que les deux séries de spots également covariables, les spots 3340, 3303 et 3304, ainsi que les spots 3225, 3217 et 3215, pourraient exprimer le même phénomène. De plus, le locus à l'origine de cette variation apparaît génétiquement très lié au locus (n°3315) responsable des modifications d'une autre protéine présente en grande quantité sur les gels correspondant aux spots 3315, 3403 et 3401 (figures 5 et 6).

### 3. Variations de position

Parmi les 150 spots trouvés variables dans notre étude, 93 ont pu être interprétés comme des variants de position, dont les modifications étaient expliquées par le polymorphisme de 35 locus. Ce type de variation affecte donc couramment les protéines de l'endosperme du Pin maritime, et deux à trois variants de la même protéine sont facilement trouvés parmi les gels des endospermes des 18 arbres étudiés. Seules des hypothèses peuvent être formulées sur les mécanismes moléculaires en jeu. Une étude détaillée du polymorphisme d'un gène codant pour la globuline-1 du maïs (Belanger et Kriz 1991) met en évidence des différences alléliques au niveau moléculaire se traduisant par des différences de la protéine résultante. Ainsi, la duplication d'une séquence de 36 paires de bases dans l'un des allèles du gène, produit une protéine comportant 9 acides aminés supplémentaires et d'une masse moléculaire supérieure de 1kD.

Les variations électrophorétiques observées chez les protéines de réserve peuvent aussi être dues à des modifications co- et post-traductionnelles. Ce type de modification peut intervenir sur les acides aminés (méthylations, phosphorylations, etc). Il peut consister à un ajout d'éléments non protéiques au polypeptide, ou à des changements des protéines liés à leur transport ou à leur immobilisation dans les compartiments cellulaires. Chez les protéines de réserve des céréales, l'ajout de glucose, le clivage

protéique et la combinaison des sous-unités constituent les principales transformations post-traductionnelles (Beever 1982). La glycosylation est responsable des hétérogénéités de masse moléculaire des protéines de réserve du soja (Sengupta et al. 1981), et de la phaséoline du haricot (Gepts 1989).

Si l'électrophorèse en présence de SDS est capable de révéler la modification de quelques acides aminés (de Jong et al. 1978), les variants de position observés dans notre expérience correspondent plus fréquemment à des différences de charge, et se traduisent par des déplacements horizontaux sur les gels (figure 5). En effet, le changement d'un seul acide aminé peut suffire à créer un changement de charge (Marshall et Brown 1975). La valeur théorique du pourcentage de remplacements d'acides aminés induisant un changement de charge s'est avérée inférieure (Peetz et al. 1986) ou supérieure (Leunissen et al. 1990) à la valeur observée dans de nombreuses familles protéiques. Ceci s'explique par des grandes différences d'acceptation de ces modifications selon les familles. Ces modifications apparaissent nombreuses chez les familles protéiques représentées dans l'endosperme, les protéines de réserve en particulier, dont l'hétérogénéité pour la charge est générale. Le polymorphisme détectable est donc lié aux différents sous-ensembles protéiques considérés selon les techniques utilisées (Smith et al. 1980).

### 4. Allèles nuls et variations de quantité

Les variations de type présence/absence apparaissent nombreuses : sur les 84 locus décrits dans notre expérience, 34 (40%) sont impliqués dans des variations de ce type. Allendorf et al. (1982), étudient sur des populations de *Pinus ponderosa* et *Pinus resinosa*, une trentaine de locus isoenzymatiques. Ils concluent à la rareté des allèles nuls dans ces populations. Même si l'absence de spots correspond dans certains cas à une quantité en dessous du niveau de détection ou à un variant de position situé hors du gel, les allèles nuls pourraient être plus fréquents dans notre sous-ensemble protéique que parmi les isozymes. Pourtant, Allendorf et al. (1982) n'observent pas avec les isozymes de désavantage sélectif aux homozygotes pour l'allèle nul. De la même façon, chez le maïs et le soja, la faible quantité ou l'absence de certaines protéines de réserve ne semble pas avoir d'incidence sur la viabilité de la graine (Belanger et Kriz 1991, Walling et al. 1986).

Une description fine d'un allèle nul de la globuline-1 du maïs permet de comprendre son fonctionnement (Belanger et Kriz 1991). Cet allèle possède une duplication de 11 paires de bases dans le premier exon qui entraîne un déplacement du cadre de lecture. Pourtant le taux de transcription est normal mais les ARN messagers produits s'avèrent instables. Le même phénomène est observé pour d'autres cas d'allèles nuls, mais à la suite de la modification d'une seule base.

Une instabilité des ARN messagers pourrait également expliquer des variations de quantité dans certains cas. Chez le pois, le locus *r*, étudié par Mendel, a un effet pléiotrope entre autres sur la forme des graines, et sur la production de protéines. Ce dernier effet apparaît clairement lors de la comparaison des profils protéiques en électrophorèse bidimensionnelle de lignées isogéniques différant pour ce seul locus (de Vienne et al. 1988). La production de légumine est réduite dans les cotylédons des plantes de génotype *rr*, et semble liée à une production d'ARN messagers moins stables (Turner et al. 1990). De la même façon, la mutation opaque-2 chez le maïs est à l'origine d'un accroissement de la quantité de zéine. Une augmentation de la transcription des ARN messagers ou leur meilleure stabilité expliquerait cette observation (Geetha et al. 1991). Si la stabilité des ARN messagers était due aux mêmes phénomènes dans ces exemples, certaines variations quantitatives et de type présence/absence pourraient trouver ici une interprétation moléculaire commune.

L'interprétation des variations observées dans les endospermes haploïdes des pins étudiés pourrait être concentrée sur quelques protéines, en particulier des protéines de réserve. Celles-ci pourraient être identifiées par des réactions type anticorps-antigène, déjà réalisées chez le Pin maritime (Allona et al. 1992), et une meilleure description de leur expression et de leur variabilité chez ce gymnosperme serait alors possible.

## E. ESTIMATION DE LA TAILLE DU GENOME DU PIN MARITIME

La réalisation de cartes génétiques saturées, pour lesquelles l'ensemble des chromosomes sont représentés, tous les marqueurs d'un chromosome étant liés génétiquement de proche en proche, est aujourd'hui envisageable, grâce aux marqueurs de l'ADN. La taille d'une carte saturée peut être assimilée à la taille du génome, en centimorgans. Avant qu'une telle carte soit disponible, il est pourtant possible d'estimer la taille du génome à partir de données de liaison partielles (Hulbert et al. 1988, Chakravarti et al. 1991). C'est ce que nous avons voulu réaliser pour le Pin maritime, en utilisant les résultats de notre étude, ainsi que ceux de Bahrman et Damerval (1989). Du fait du développement encore limité du marquage du génome des espèces forestières, une estimation de ce type n'était pas encore disponible.

### 1. Méthodes d'estimation

A partir de résultats de linkage obtenus par la technique des lod scores, décrite en détail chapitre III.C., Hulbert et al. (1988) proposent un calcul pour estimer la taille du génome de l'espèce étudiée. Dans une étude de liaison, on ne pourra étudier la recombinaison qu'entre des couples de locus doubles hétérozygotes, ou paires de locus informatives chez un individu. La méthode de Hulbert et al. (1988) repose sur la comparaison, connaissant le nombre de couples de locus informatifs, du nombre théorique attendu de tels couples ayant un lod score supérieur ou égal à un seuil *Z*, et de ce fait déclarés liés, avec le nombre effectivement observé.

Si *M* est le nombre de paires de locus informatives, et *K* est le nombre de paires parmi elles dont le lod score a été trouvé supérieur ou égal à un seuil *Z*, alors le rapport *K/M* représente la probabilité que deux marqueurs choisis au hasard soient déclarés liés. Cette probabilité peut également être exprimée en fonction de la taille *G* du génome et de la distance *X* entre deux marqueurs pour laquelle un lod score égal à *Z* est attendu.

Pour cela, le génome est assimilé à un seul chromosome de taille *G*. On suppose que les marqueurs sont disposés au hasard sur le génome et suivent une loi uniforme de densité  $1/G$ . On montre alors que la loi de probabilité de la distance *x* entre deux marqueurs tirés au hasard est de densité  $2(G-x)/G^2$ . Quand *x* est petit devant *G*, cette densité est proche de  $2/G$ . La distance *X* entre deux marqueurs correspond à une valeur seuil. Si la distance *x* entre deux marqueurs est comprise entre 0 et *X*, on les déclarera liés. Si cette distance dépasse *X*, ils seront déclarés indépendants. Le produit  $X \cdot (2/G)$  estime alors la probabilité que deux marqueurs tirés au hasard soient déclarés liés et peut donc être égalé à *K/M*. On en déduit l'expression suivante de *G* :

$$G = \frac{2MX}{K} \quad (1)$$



X est relié au taux de recombinaison  $\theta$  :

$$X = - \frac{1}{2} \ln(1-2\theta)$$

et  $\theta$  est la solution de l'équation :

$$Z = n (\theta \log_{10} 2\theta + (1-\theta) \log_{10} 2(1-\theta)) \quad (2)$$

où  $n$  est le nombre de méioses étudiées pour la paire considérée. Dans notre cas,  $n$  varie pour chaque couple considéré. En effet, chaque paire de locus informative le sera uniquement dans un sous-ensemble des 18 génotypes, et le nombre de méioses observées, correspondant au nombre total de gamètes étudiés dans le sous ensemble sera variable d'une paire à l'autre. Hulbert et al. (1988) estiment X pour chaque couple de locus informatif, en fixant une valeur de Z dans l'équation (2). La somme des valeurs obtenues se substitue alors à (MX) dans l'équation (1).

Chakravarti et al. (1991), proposent une variante à cette méthode. Elle consiste à choisir, parmi les couples de locus dont le lod score est supérieur ou égal à Z, la paire qui correspond au taux de recombinaison  $\theta$  le plus élevé. La valeur de la distance X correspondante est alors utilisée dans l'équation (1).

Dans ces deux méthodes, G peut être estimé pour différentes valeurs de Z.

## 2. Application aux données du Pin maritime

On observe dans notre échantillon 89 couples sur les 2711 disponibles dont le lod score est supérieur ou égal à 2, pour une distance maximale de 35 cM. Soixante neuf couples correspondent à un lod score minimal de 3, pour une distance maximale de 29 cM. Pour ces deux valeurs de lod score, avec la méthode de Hulbert et al. (1988) (méthode 1) et la méthode de Chakravarti et al. (1991) (méthode 2), les estimations de la taille du génome sont les suivantes :

	Z	2	3
G (cM) :	méthode 1	1930	1787
	méthode 2	2114	2279

Les données de Bahrman et Damerval (1989), obtenues à partir de 56 endospermes d'un seul arbre ont été réanalysées avec MAPMAKER comme un backcross. Les 666 couples possibles associant les 37 locus deux à deux ( $36 \cdot 37 / 2$ ) sont informatifs. Parmi eux 42 paires ont un lod score minimal de 2 avec une distance maximale de 63.65 cM, 33 ont un lod score minimal de 3 et une distance maximale de

51.08 cM. Dans ce cas,  $n$  est constant pour tous les couples de locus ( $n = 56$ ), l'équation (1) peut être appliquée directement. On obtient les estimations suivantes de G :

	Z	2	3
G (cM) :	méthode 1	1455	1453
	méthode 2	2061	2019

Les estimations obtenues par la première méthode correspondent à des valeurs plus faibles de G pour les deux séries de données. Pourtant, Chakravarti et al. (1991) en comparant les deux méthodes à partir de simulations et de leur application à des données de Morizot et al. (1990), constatent que la méthode de Hulbert et al. (1988) conduit à des surestimations, alors que leur variante donne des estimations plus faibles. Nos résultats sont contraires.

Le choix d'un seuil de 2 ou 3 modifie peu les estimations obtenues à partir des données de Bahrman et Damerval (1989). Les variations sont plus importantes pour ce qui concerne nos données.

Les estimations obtenues à partir de nos données correspondent à une moyenne d'environ 2000 cM. Les données de Bahrman et Damerval suggèrent une taille inférieure, d'environ 1750 cM. Ces deux estimations, obtenues pourtant à partir d'un matériel végétal et d'un nombre de locus différents, sont assez semblables. En première approximation, la taille du génome du Pin maritime serait donc proche de 2000 cM.

Quelques auteurs ont cherché à avoir une idée de la taille du génome des pins. Strauss (1986) à partir de données bibliographiques, suppose qu'un chiasma correspond environ à 50 centimorgans chez différents organismes. Chez les pins, puisque 2.2 à 2.5 chiasmas sont observés en moyenne par bivalents, il en déduit une taille du génome de 1320 à 1500 centimorgans. Neale et Williams (1991) proposent également une estimation rapide basée sur des données d'isozymes. Ils reprennent pour cela un calcul de Guries et al. (1978) qui supposent que 30 groupes de liaisons indépendants existent chez les pins (12 chromosomes multipliés par 2.5 chiasmas par chromosome). Le plus grand groupe de liaison, observé par Conkle (1981) avec des isozyms, ayant 80 cM, Neale et Williams multiplient ces deux nombres pour suggérer une taille du génome de 2500 centimorgans. Ces calculs sont fondés sur des assertions assez imprécises, néanmoins, l'estimation de 2000 cM que nous proposons est intermédiaire entre ces valeurs. Avec cette taille, Neale et al. (1989) calculent que 275 marqueurs seraient nécessaires pour couvrir 90% du génome et espérer détecter des locus à caractère quantitatif situés à environ 10 centimorgans des marqueurs.

## F. CONCLUSION

L'électrophorèse bidimensionnelle des protéines peut donc être utilisée comme un outil de marquage génétique. Pour parvenir à une caractérisation claire des locus responsables des variations touchant les 150 spots retenus dans notre étude, nous avons adopté une démarche qui pourrait être généralisée. Dans un premier temps, des similitudes entre les variations des spots ont été systématiquement recherchées, pour repérer des informations génétiques "redondantes" : certains spots pouvaient être vus comme les variants de position d'une même protéine, d'autres apparaissaient influencés par le même facteur génétique, ou par deux facteurs génétiques très liés. La description de l'effet de chacun des 84 locus finalement retenus, responsables de variations de position, de quantité, induisant une présence/absence des protéines, ou liés à des covariations de spots, est disponible (figure 5). La validité des hypothèses génétiques formulées pourra donc être testée lors d'études ultérieures sur d'autres arbres, ou sur des endospermes supplémentaires des mêmes arbres, notamment pour trancher entre les hypothèses de l'action de deux locus très liés ou d'un seul locus sur des spots différents. L'analyse plus fine de variations particulières, et la recherche d'explications au niveau moléculaire pourront aussi être envisagées.

La carte réalisée, qui localise 65 des 84 locus, couvre 530 centimorgans, soit plus d'un quart du génome, qui comporterait 2000 centimorgans d'après nos estimations. Pourtant, du fait du dépouillement visuel des gels, une grande sévérité dans le choix des variations à retenir a été nécessaire. Les variations décrites représentent donc une sous-estimation de la variabilité existante. L'analyse automatique des gels permettra à l'avenir une détection plus fine des variations, notamment en ce qui concerne les variations de quantité des spots protéiques. On peut en effet supposer que d'un arbre à l'autre, des allèles différents peuvent être exprimés pour des locus induisant des variations de quantité. Compte tenu du nombre d'endospermes par arbre analysés dans cette étude, et du dépouillement visuel, seuls les cas les plus simples ont pu être retenus. La carte pourra donc être étendue à un nombre plus grand de locus, la représentation du génome en sera d'autant améliorée.

Les résultats obtenus ici nous ont amené à approfondir deux questions. D'un point de vue méthodologique, l'estimation et le test de la liaison entre deux marqueurs dans notre dispositif méritait une attention particulière. D'autre part, à la suite de l'estimation proposée, nous nous sommes intéressés aux notions concernant la taille du génome, et notamment chez un gymnosperme comme le pin.

CHAPITRE III

IDENTIFICATION ET LOCALISATION DES Locus

IDENTIFICATION DES Locus

### III. ESTIMATION ET TEST DE LA LIAISON ENTRE DEUX MARQUEURS

#### A. INTRODUCTION

Si la deuxième loi de Mendel montre l'indépendance des ségrégations de différents caractères, on observe dès 1902 des exemples contraires, et Morgan vers 1910, sur la drosophile, définit la notion de linkage. L'idée de localiser les entités physiques qui portent l'information, les gènes ou locus, linéairement les uns par rapport aux autres, est concrétisée en 1913 par Sturtevant, qui cartographie 5 locus chez la drosophile. Les techniques mathématiques liées à l'analyse du linkage et à la cartographie sont développées dans les années 1934, 1936 par Fisher et Haldane (Bailey 1961).

Tanksley et Rick dans un article de 1980, signalent la décroissance de l'intérêt porté aux études de liaison génétique chez les plantes. Un redéveloppement considérable des études sur ce thème interviendra pourtant dans les années suivantes. En effet, l'utilisation des marqueurs de l'ADN, et de la technique du polymorphisme de longueur des fragments de restriction (RFLP) en particulier, ouvrait de nouvelles possibilités. Les marqueurs isoenzymatiques, utilisés depuis 1966 (Lewontin 1991), ne permettaient l'accès qu'à un nombre réduit de locus. Au contraire, les RFLP fournissaient un nombre a priori illimité de marqueurs neutres et codominants. L'étude de l'hérédité de ces marqueurs et leur cartographie génétique a d'abord intéressé la génétique humaine (Botstein et al. 1980, Drayna et al. 1984, White et al. 1985), puis s'est étendue à la génétique des plantes annuelles où environ 300 marqueurs on pu être cartographiés chez le maïs (Helentjaris 1987), les choux (Slocum et al. 1990) ou la pomme de terre (Gebhardt et al. 1991). Cette approche ne fait que débiter chez les plantes pérennes, où les problèmes sont parfois proches de ceux auxquels est confrontée la génétique humaine.

#### 1. Etudes de linkage en génétique humaine

En génétique humaine, au contraire de la génétique des plantes annuelles, où les générations F<sub>2</sub>, les backcross, les haploïdes doublés ou les lignées recombinantes, fournissent un matériel génétiquement bien connu et en quantité non limitante, le matériel d'étude est "imposé" pour la réalisation de cartes génétiques. Les individus sont largement hétérozygotes, et il s'agit alors de combiner des données provenant de familles ou de génotypes très différents, avec des effectifs par familles souvent réduits.

Si deux locus sont proches, sur le même chromosome, leurs allèles auront tendance à être hérités ensemble, en reproduisant l'association allélique (ou haplotype) parentale. Plus ils seront éloignés l'un de l'autre, plus la recombinaison créera de gamètes portant une combinaison allélique différente. La liaison entre deux locus est donc définie à partir de la probabilité de recombinaison entre ces deux locus, estimée par la proportion de gamètes recombinés parmi les gamètes parentaux. Si l'on travaille

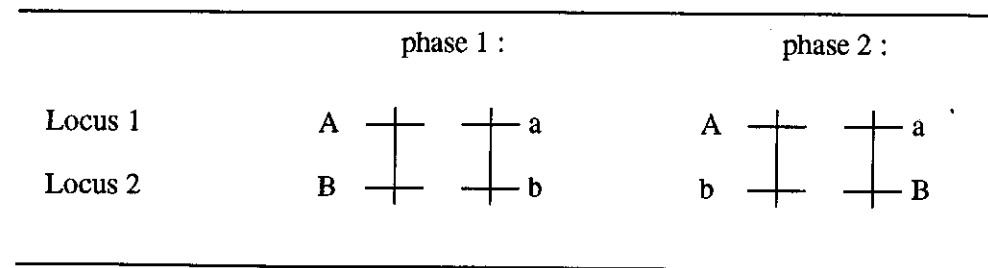


Figure 7 : Les deux phases possibles d'un individu hétérozygote au locus 1 (A/a) et au locus 2 (B/b).

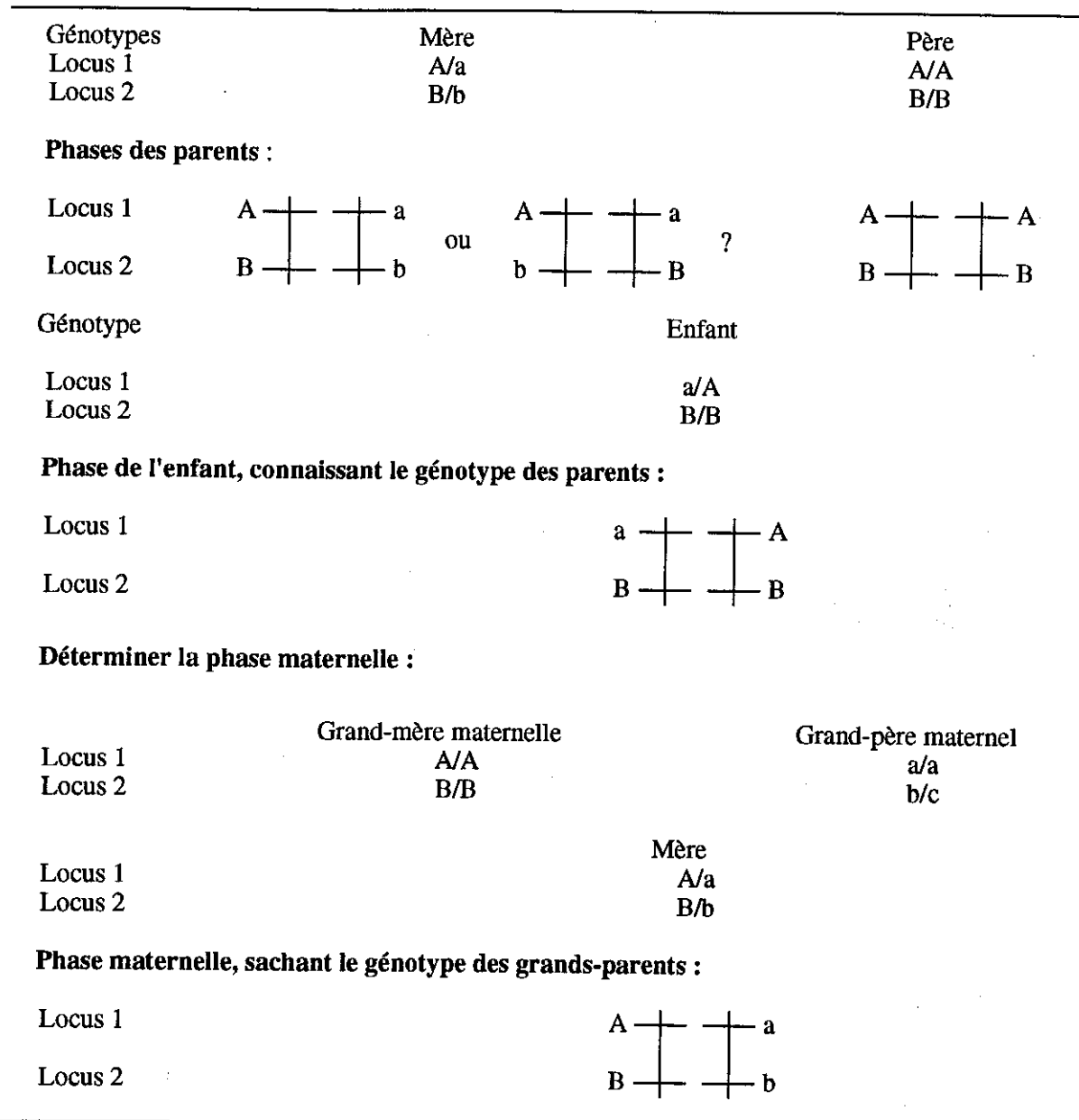


Figure 8 : Un exemple de pedigree où la connaissance du génotype des grands-parents lève l'ambiguïté sur la phase de l'un des parents (White et al. 1985).

sur des individus issus de croisement, la généalogie est parfaitement connue et il n'y a aucune ambiguïté sur la phase parentale relativement à laquelle on identifiera les recombinés. Par contre, sur du matériel non expérimental, un individu hétérozygote à deux locus peut se trouver a priori dans l'une ou l'autre des deux phases possibles (figure 7). On ne pourra pas "compter" les recombinaisons, car on ne connaîtra pas la phase parentale (on ne saura pas si les allèles parentaux se trouvent en cis ou en trans) et on ne pourra donc pas déterminer à quel endroit les crossing-over sont intervenus (Lander et Green 1987). Pour résoudre ce problème, la génétique humaine s'appuie généralement sur des familles où plus de deux générations sont représentées. En effet, le génotype des grands-parents peut, mais seulement dans certains cas, lever l'ambiguïté (figure 8).

De plus, les familles étudiées sont de petite taille, et de ce fait, on est contraint de travailler en combinant les informations issues de plusieurs familles. Chaque locus sera hétérozygote dans une fraction des familles seulement, et les quantités de données disponibles pour chaque paire de locus ne seront pas égales.

Dans le but d'estimer les taux de recombinaison et de cartographier les locus, des méthodes statistiques ont alors été élaborées pour tenir compte de toutes ces contraintes. Celles-ci sont basées sur la méthode du maximum de vraisemblance et étaient déjà employées par Haldane et Fisher dans les années 1930 (Thompson 1990). Ces techniques ont été généralisées par l'intermédiaire de logiciels conçus à l'origine pour les études de familles humaines. Deux logiciels sont principalement utilisés. LINKAGE (Lathrop et al. 1984, Lathrop et Lalouel 1988) est basé sur une formulation générale de la vraisemblance pour des données issues de pedigrees (Elston et Stewart 1971). Ce logiciel commence à être utilisé en génétique animale, où les familles sont souvent constituées de pedigrees complexes, que LINKAGE peut prendre en compte (P. Le Roy, communication personnelle). MAPMAKER (Lander et Green 1987, Lander et al. 1987) repose sur des algorithmes différents qui doivent accélérer la réalisation des cartes avec de nombreux marqueurs, malgré des quantités de données variables selon les locus et des phases inconnues. Une carte recouvrant 95% du génome, et qui localise 403 locus RFLP sur le génome humain a été réalisée grâce à ce logiciel (Donis-Keller et al. 1987). Le logiciel ayant été étendu pour l'analyse de familles F2 et de backcross, il est actuellement très utilisé par les généticiens des végétaux.

## 2. Etude de linkage chez des espèces végétales pérennes

Prenant l'exemple de données issues de gymnospermes, avec des phases inconnues, Nordheim et al. (1983) justifie a posteriori l'utilisation de méthodes bayésiennes pour estimer les taux de recombinaison. Ils constatent en effet des carrés moyens d'erreur plus faibles avec ce type d'estimation et en particulier pour des tailles d'échantillons réduites. Pourtant, l'utilisation d'une loi uniforme a priori pour le taux de recombinaison ne se justifie pas en soi, les techniques de maximum de

vraisemblance traditionnelles nous ont semblé plus à propos. La technique préconisée par ces auteurs a été appliquée pour des locus isoenzymatiques chez des conifères (O'Malley et al. 1986, Harry 1986, Geburek et Wang 1990). Avec la multiplication du nombre de locus à analyser simultanément, l'utilisation de logiciel de cartographie et des méthodes liées au maximum de vraisemblance se généralisent à toutes les espèces. Ainsi, une étude engagée actuellement pour la cartographie de locus RFLP chez un pin (Devey et al. 1991), reprend le principe de la génétique humaine en travaillant à partir de pedigrees. La cartographie pourra alors être réalisée par le logiciel MAPMAKER, avec quelques modifications (Neale et Williams 1991). Dans notre cas, l'analyse des liaisons génétiques à partir des 18 génotypes non apparentés composant l'échantillon pouvait être assimilée à une étude de plusieurs familles humaines :

- les phases parentales des 18 individus devaient être déduites de l'observation d'un nombre limité de leurs gamètes,
- l'étude des distances entre locus s'appuyait sur des quantités d'information très inégales, selon l'hétérozygotie des différents individus pour les locus considérés.

Le logiciel MAPMAKER, et plus particulièrement les programmes écrits pour des familles humaines, pouvait être utilisé directement pour la réalisation de notre carte génétique en assimilant chaque gamète à un enfant, et en créant un père fictif (voir tableau 5).

### 3. Utilisation du logiciel de cartographie

La construction d'une carte génétique avec MAPMAKER comporte plusieurs étapes. Le taux de recombinaison entre les paires de locus pour lesquelles des données sont disponibles est tout d'abord estimé. Un test sévère est alors appliqué à ces estimations pour décider si les locus sont liés ou non. Les groupes de liaison supposés sont alors constitués. Un locus sera ajouté à un groupe s'il est lié avec un locus du groupe (au moins), leur taux de recombinaison ne dépassant pas un seuil fixé, et la valeur du test correspondant étant au moins égale à un minimum, choisi au préalable. L'ordre le plus vraisemblable des locus à l'intérieur des groupes ainsi constitués est alors recherché. Il est possible de tester ensuite, par rapport à ces groupes "forts", la liaison éventuelle des locus restés non liés après la première étape. Pour finir, la distance entre les locus de chaque groupe de liaison est calculée.

Le logiciel MAPMAKER présente un désavantage particulier : il n'est pas possible aux utilisateurs de ce programme de connaître son fonctionnement précis et les méthodes de calcul utilisées. Les publications le concernant ne donnent en effet que peu d'informations à ce sujet (Lander et al. 1987, Lincoln et al. 1990). De ce fait, en amont de la cartographie, nous avons cherché à savoir plus précisément comment les taux de recombinaison étaient estimés par le logiciel. De plus, nous avons

voulu comparer la technique du lod score, utilisée pour décider de l'existence d'une liaison entre deux locus par MAPMAKER, avec un test plus classique.

## B. ESTIMATION DES TAUX DE RECOMBINAISON

La vraisemblance d'un échantillon observé (ici, une ségrégation) peut être écrite pour un dispositif donné en fonction du taux de recombinaison, dont on ignore la valeur. Cette vraisemblance est liée à la probabilité d'avoir observé la ségrégation des deux locus considérés en les supposant liés avec un taux de recombinaison  $\theta$ . L'utilisation de la technique de l'estimateur du maximum de vraisemblance, comme son nom l'indique, consiste à estimer ce taux par la valeur qui rend la vraisemblance maximale. La probabilité d'avoir observé l'échantillon est alors la plus grande possible.

### 1. Ecriture de la vraisemblance

Pour estimer la liaison entre deux locus, seuls sont informatifs les individus doubles hétérozygotes. Soit  $m$  le nombre de tels individus pour une paire de locus donnée. On a observé  $N_i$  gamète de l'individu  $i$ . Parmi ces gamètes,  $n_i$  sont dans la phase 1 de la figure 7. Soit  $p$  la probabilité que la phase de l'individu  $i$  soit de type 1, donc  $(1-p)$  est la probabilité que celle-ci soit de type 2. Soit  $\theta$  le taux de recombinaison. La probabilité qu'un gamète reproduise l'une des haplophases parentales est  $(1-\theta)$ . A l'inverse, la probabilité que ce gamète soit recombinant est  $\theta$ . La probabilité d'observer  $n_i$  gamètes d'une haplophase parmi les  $N_i$  gamètes de l'individu  $i$  est alors :

$$P_i(N_i, n_i, \theta) = C_{N_i}^{n_i} [p (\theta^{n_i} (1-\theta)^{N_i-n_i}) + (1-p) (\theta^{N_i-n_i} (1-\theta)^{n_i})]$$

D'où la vraisemblance pour  $m$  génotypes indépendants :

$$P(N, n, \theta) = \prod_{i=1}^m P_i(N_i, n_i, \theta)$$

La log-vraisemblance, s'écrit donc :

$$L \approx \sum_{i=1}^m \ln [p (\theta^{n_i} (1-\theta)^{N_i-n_i}) + (1-p) (\theta^{N_i-n_i} (1-\theta)^{n_i})]$$

Il reste à chercher les valeurs de  $p$  ( $0 \leq p \leq 1$ ) et de  $\theta$  ( $0 \leq \theta \leq 0.5$ ) qui maximisent la log-vraisemblance. Le paramètre  $p$  peut être ici considéré comme un paramètre de nuisance : sa valeur nous importe peu et pourtant la vraisemblance en dépend. L'estimation du taux de recombinaison  $\theta$  nous intéresse plus particulièrement.

## 2. Résultats des maximisations

Avec 84 marqueurs étudiés, 3486 ( $84 \times 83/2$ ) combinaisons de marqueurs deux à deux étaient possibles. Seuls 2711 couples de locus ont été trouvés informatifs parmi les gamètes d'un génotype (double hétérozygote pour ces locus) au moins. 1521 couples étaient informatifs pour trois génotypes au moins. Un maximum de 12 génotypes a été trouvés informatif pour quelques paires de locus.

Les 2711 maximisations ont été réalisées en utilisant la procédure Minfun de la bibliothèque Statlib du logiciel Splus (Becker et al. 1988), cette procédure permet de rechercher le minimum de fonctions dont les paramètres ont des valeurs bornées et ne nécessite pas le calcul analytique des dérivées. Les taux de recombinaison calculés grâce à cette méthode ont été trouvés égaux aux taux fournis par le logiciel MAPMAKER. Ceci nous a donc permis de vérifier quel était le modèle utilisé par ce logiciel pour estimer les taux de recombinaison, faute de documentation plus précise sur son fonctionnement.

### C. DETECTION D'UNE LIAISON ENTRE DEUX MARQUEURS

Les différents tests utilisés pour détecter une liaison entre deux locus reposent sur un même principe : ils consistent à tester l'hypothèse nulle  $H_0 : \theta = 0.5$  de l'indépendance entre les deux locus, contre  $H_1 : \theta \neq 0.5$ .

Le test utilisé par le logiciel MAPMAKER est appelé lod score ("lod" pour "log of the odds ratio", "logarithme du rapport des avantages"). Les bases de ce test ont été posées par Morton en 1955. Le lod score est égal au logarithme décimal du rapport de la vraisemblance calculée en supposant que les deux locus sont liés sur la vraisemblance calculée sous l'hypothèse de l'indépendance de ces deux locus. Si ce rapport dépasse un certain seuil A, les deux locus seront alors déclarés liés avec un taux de recombinaison égal à l'estimateur du maximum de vraisemblance de  $\theta$ . Le lod score est le test du rapport de vraisemblances.

Un test statistique, type  $\chi^2$ , est caractérisé par le choix d'un seuil de signification défini par l'erreur de première espèce, qui exprime l'importance du risque que l'on prend de rejeter, à tort, l'hypothèse nulle alors qu'elle est vraie. Pour un test de liaison, en se plaçant sous l'hypothèse nulle d'indépendance entre les deux locus, cela correspondrait au risque de décider que ces locus sont liés alors qu'ils ne le sont pas. Si, dans des études très diverses, le même seuil  $A = 3$  est toujours employé, défini au départ, pour des raisons que nous analyserons, par Morton (1955), cela ne correspond pas forcément à des risques de première espèce identiques. Dans le cas de nos données, nous avons voulu comparer le lod score, calculé par MAPMAKER pour chaque couple de locus à un test plus classique, où le risque de première espèce pouvait être précisé.

## 1. Construction du test

### 1.1. Test du rapport de vraisemblances

En prolongement de l'estimation du taux de recombinaison par maximum de vraisemblance, un test du rapport de vraisemblances aurait pu être réalisé. La construction de ce test dépend de certaines propriétés de la log-vraisemblance (Prum 1991). En particulier, soit  $L(\theta)$  la log-vraisemblance du paramètre  $\theta$ . On suppose qu'elle est deux fois dérivable et on note  $L'(\theta)$  et  $L''(\theta)$  ses dérivées première et seconde par rapport à  $\theta$ . On se place en  $\theta^*$ , vraie valeur de  $\theta$ . Sous certaines conditions de régularité, on montre que :

- l'estimateur du maximum de vraisemblance  $\hat{\theta}$  converge presque sûrement vers la vraie valeur  $\theta^*$  du paramètre,
- $1/n L''(\hat{\theta})$  converge presque sûrement vers son espérance en  $\theta^*$ . Celle-ci est notée  $I(\theta^*)$  et est appelée information de Fisher en  $\theta^*$ ,
- $\sqrt{n}(\hat{\theta} - \theta^*)$  converge en loi vers une loi normale  $N(0, I^{-1}(\theta^*))$ , si  $I(\theta^*)$  est inversible.

Sous l'hypothèse  $H_0 (\theta^* = \theta_0)$ , pour  $n$  suffisamment grand,  $\hat{\theta}$  est proche de  $\theta_0$  et on peut écrire un développement limité de la vraisemblance autour de  $\hat{\theta}$ . En particulier en  $\theta_0$  :

$$L(\theta_0) \approx L(\hat{\theta}) + (\theta_0 - \hat{\theta}) L'(\hat{\theta}) + 1/2 (\theta_0 - \hat{\theta})' L''(\hat{\theta}) (\theta_0 - \hat{\theta})$$

Le terme  $L'(\hat{\theta})$  est nul par définition de  $\hat{\theta}$ , on a donc :

$$2[L(\hat{\theta}) - L(\theta_0)] \approx \sqrt{n}(\theta_0 - \hat{\theta})' (-L''(\hat{\theta})/n) \sqrt{n}(\theta_0 - \hat{\theta})$$

Cette différence de vraisemblances (qui équivaut à un rapport) converge donc en loi vers un  $\chi^2$  dont le nombre de degrés de liberté correspond à la dimension de  $\theta$ .

Dans notre cas, la vraisemblance dépend de deux paramètres  $\theta$  et  $p$ . Si l'on fixe  $\theta = 1/2$ , les deux lois mélangées sont identiques et la vraisemblance ne dépend plus de  $p$ .

Le calcul de  $L''(p, \theta)$  est développé en annexe 1. Il montre que cette matrice dépend de deux paramètres et dépend en particulier de  $p$  en  $\theta = 1/2$ . De plus, on peut montrer, sous  $H_0$ , que  $\hat{p}$  converge presque sûrement vers  $1/2$  où  $L''(1/2, 1/2) = 0$ . On ne peut donc pas déduire la loi asymptotique du rapport de vraisemblances en appliquant les résultats classiques.

On pourrait estimer la loi du rapport de vraisemblances sous  $H_0$  par simulation mais cette méthode nécessiterait des temps de calcul très longs pour chaque  $\hat{\theta}$ . De plus, il faudrait un grand nombre de jeux de données simulées pour obtenir une précision correcte de la queue de distribution.

### 1.2. Test du $\chi^2$

En reprenant les notations utilisées dans l'écriture de la vraisemblance, pour deux locus, on observe, pour chaque génotype  $i$  double hétérozygote,  $N_i$  gamètes parmi lesquelles  $n_i$  sont dans une phase donnée. Sous l'hypothèse  $H_0$  de l'indépendance des deux locus ( $\theta = 1/2$ ), la variable  $n_i$  suit une loi binomiale de paramètres  $N_i$  et  $1/2$ . Pour un arbre double hétérozygote la statistique  $S_i$  :

$$S_i = \frac{(2n_i - N_i)^2}{N_i}$$

Suit asymptotiquement (en  $N_i$ ) une loi du  $\chi^2$  à 1 degré de liberté. On ignore si les  $n_i$  (respectivement les  $(N_i - n_i)$ ) gamètes correspondent au type recombiné ou parental. Mais on constate que la statistique  $S_i$  est invariante si  $n_i$  est remplacé par  $(N_i - n_i)$ . Le rejet de l'hypothèse nulle permettra donc de détecter une liaison entre deux locus sans connaître la phase parentale. On considère alors les  $m$  arbres doubles hétérozygotes pour la paire de locus étudiée. Ils sont indépendants, ce qui permet d'écrire la statistique  $S$  :

$$S = \sum_{i=1}^m S_i$$

qui suit asymptotiquement une loi du  $\chi^2$  à  $m$  degrés de liberté, toujours sous l'hypothèse nulle.

La valeur de  $S$  ainsi que la probabilité associée peuvent être calculées pour chaque couple de locus. Ceci a été réalisé pour les 2711 couples par un programme dans le langage Splus (Becker et al. 1988). Pour étudier la puissance du test, on se place dans les conditions les plus défavorables où les répartitions sous  $H_0$  et sous  $H_1$  sont proches et donc difficiles à distinguer grâce au test. Sous l'hypothèse  $H_0$ , nous avons vu que  $n$  suit une loi binomiale  $B(N, 0.5)$ . Sous  $H_1$ ,  $n$  correspond à un mélange de lois binomiales :  $pB(N, \theta) + (1-p)B(N, 1-\theta)$ . Pour une valeur de  $\theta$  donnée, la répartition sous  $H_1$  la plus semblable à la répartition sous  $H_0$  correspond à  $p = 1/2$ . Dans ces conditions, pour  $\theta = 0.47$ , la puissance du test est de 7%. Celle-ci s'améliore pour des valeurs de  $\theta$  plus élevées : pour  $\theta = 0.4$ , la puissance est de 41%, elle atteint 90% pour  $\theta = 0.3$  (annexe 2).

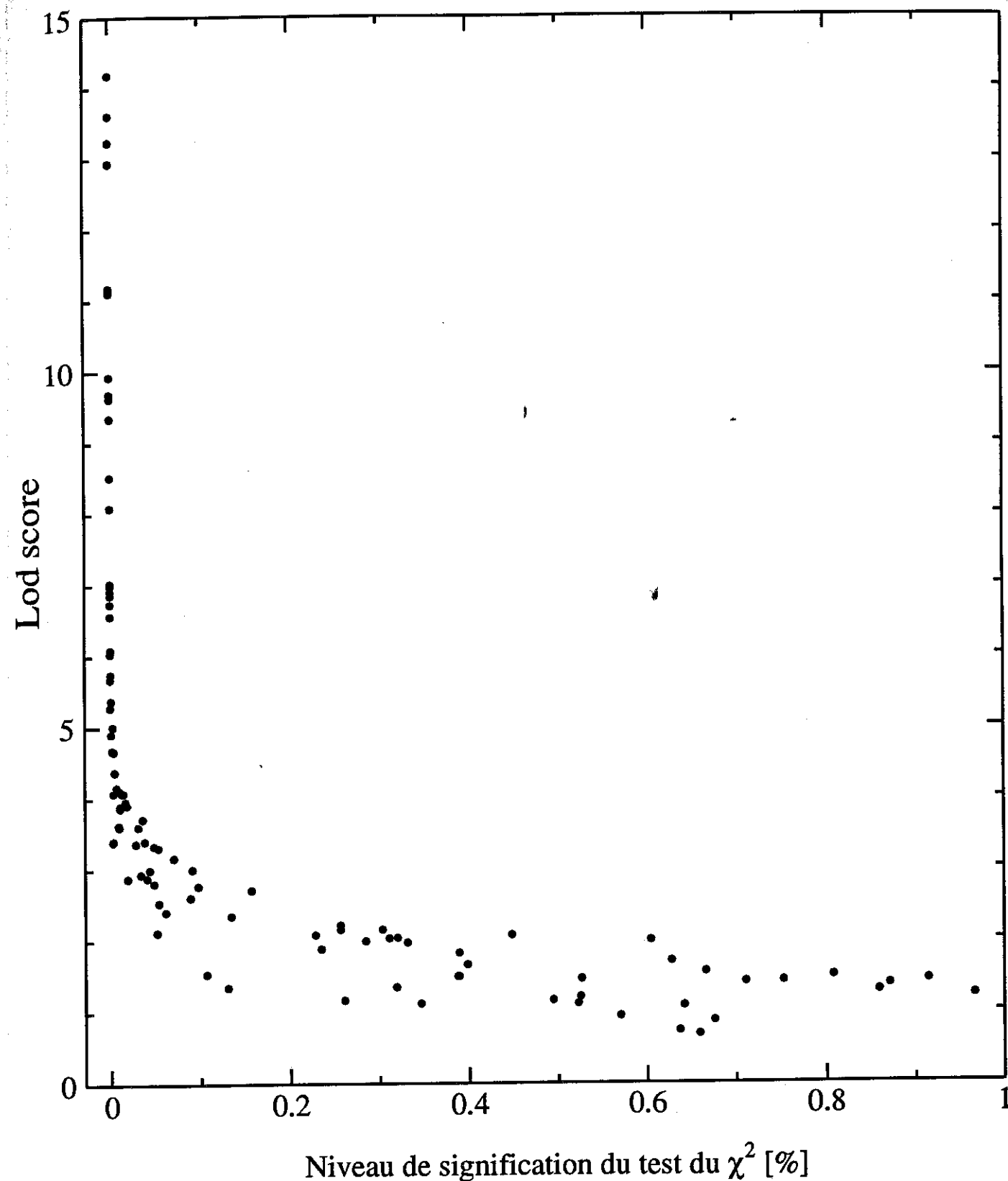
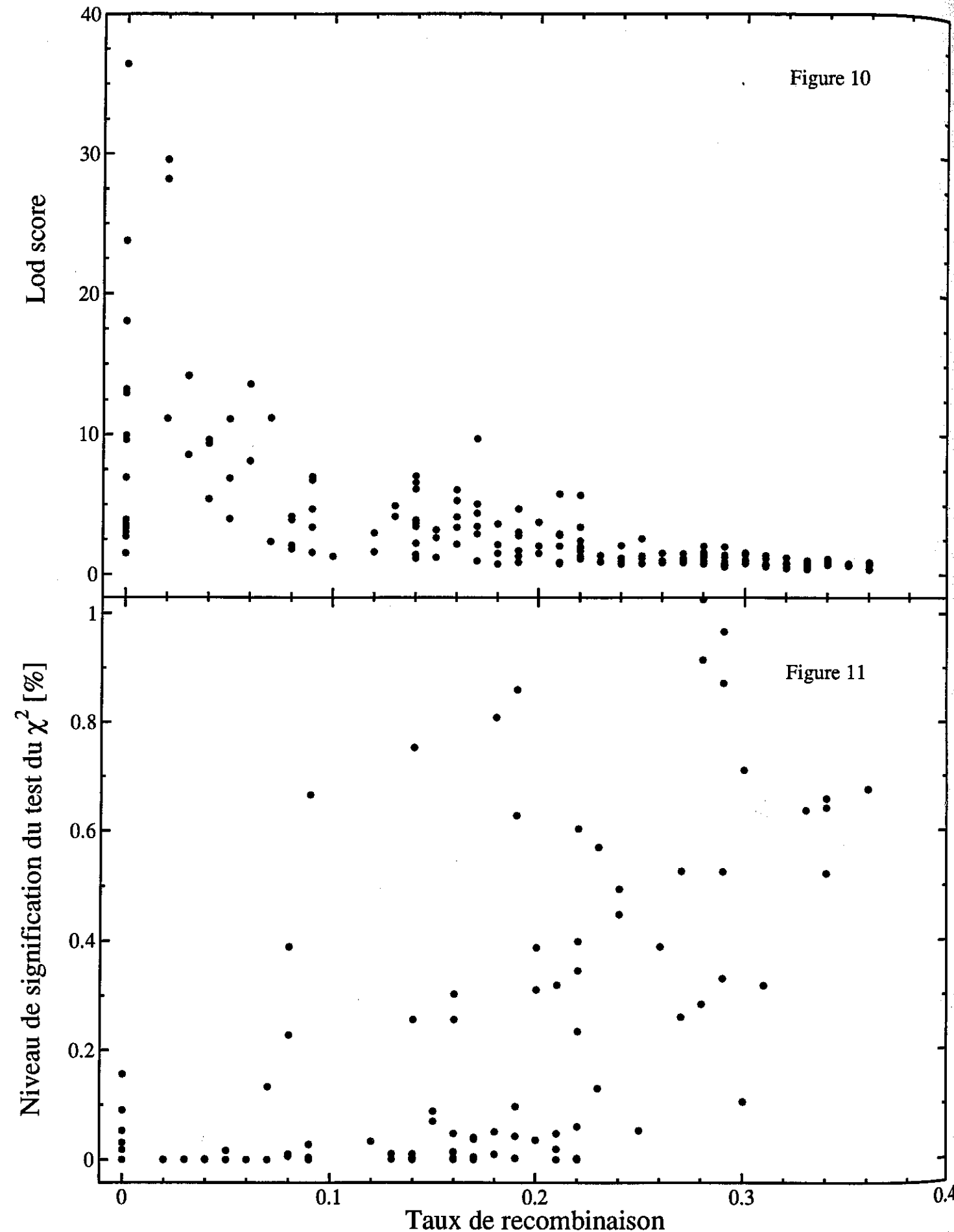


Figure 9 : Relation entre la valeur du lod score et le niveau de signification du test du  $\chi^2$  pour 129 couples de locus.



Figures 10, 11 : Relations entre la valeur du lod score (fig.10), le niveau de signification du test du  $\chi^2$  (fig.11) et la valeur estimée du taux de recombinaison.

## 2. Comparaison du lod score et du test du $\chi^2$

Pour chacune des 2711 paires de locus informatives, on dispose de la valeur du lod score calculée par MAPMAKER et de la probabilité qu'une valeur supérieure à la valeur du test du  $\chi^2$  calculée pour la paire, puisse avoir été observée du seul fait du hasard. On décidera de rejeter l'hypothèse nulle avec un risque de première espèce  $\alpha$  si cette probabilité est inférieure à  $\alpha$ .

Parmi les 2711 couples, 129 correspondent à une probabilité inférieure à 1%. La relation entre le lod score et le niveau de signification du test du  $\chi^2$  pour ces couples est représentée figure 9. Le lod score apparaît proche d'une fonction décroissante du niveau de signification du test du  $\chi^2$ . Cette observation suggère que ces tests sont équivalents ou presque, Cependant, cette équivalence n'est pas stricte : les deux tests méritent d'être comparés.

Si un seuil de 1% pour le test du  $\chi^2$  était jugé suffisamment faible pour rejeter l'hypothèse nulle, on déclarerait liés 60 couples de locus (46%) ayant un lod score inférieur à 3, et parmi eux, 40 avec un lod score inférieur à 2. En effet, le lod score correspond à un test très sévère. Sur les 2711 couples, 89 ont un lod score supérieur à 2, et le niveau de signification maximal du test du  $\chi^2$  parmi ces 89 observations est égal à 0.0045 (0.45%). 69 observations ont un lod score supérieur à 3 pour un niveau maximum de 0.0008, 48 ont un lod score supérieur à 4 pour un niveau maximum de 0.00015. La correspondance entre le niveau de signification du test du  $\chi^2$  et le lod score ne pose pas de problèmes pour les valeurs extrêmes des deux tests, par contre, aux niveaux où se place le choix des seuils, les deux tests concordent moins bien. Si, par exemple, on impose un seuil de 0.0008 au test du  $\chi^2$ , on conservera des couples de locus pour lesquels le lod score est inférieur à 3. Par contre, en choisissant un seuil élevé pour le lod score, on est sûr d'imposer une erreur de première espèce très faible, mais aussi trop faible parfois : certaines liaisons peuvent être rejetées. Avec le test du  $\chi^2$ , il est possible de choisir directement une erreur de première espèce, suffisamment faible pour éviter de déclarer de fausses liaisons, mais moins sévère que le lod score. Par exemple, si l'on souhaite rester en deçà d'un risque global  $\alpha$ , en étudiant  $n$  locus, on imposera un seuil de  $\alpha/(n-1)$  pour chaque test réalisé.

Le lod score est également très "prudent" sur des valeurs du taux de recombinaison trop élevées (figure 10). Avec un lod score minimum de 3, le taux de recombinaison maximum retenu est égal à 0.25. Avec le test du  $\chi^2$  (figure 11) des taux de recombinaison plus élevés (0.30, 0.35) pourraient être conservés. De plus, comme nous l'avons vu, le test est puissant pour des valeurs de  $\theta$  de cet ordre.

L'analyse, avec MAPMAKER, des données de Bahrman et Damerval (1989) concernant 56 megagamétophytes d'un seul pin maritime, assimilés aux descendants d'un backcross, conduit aux mêmes résultats que ces auteurs. Les liaisons trouvées significatives à 0.1 % et même à 1% correspondent à des lod scores supérieurs à 3, seuil utilisé par Lincoln et al. (1990) pour ce type de données. Pourtant, les liaisons trouvées significatives à 5% par le test du  $\chi^2$  correspondent à des



valeurs de lod score un peu supérieures à 2, l'une d'elles étant nulle. Le test du lod score accepte ici l'hypothèse d'indépendance. Donc, dans cet exemple, un seuil de 3 correspond à peu près à un risque de première espèce de 1%, plus élevé que celui calculé pour le même seuil à partir de nos données. Notre dispositif, composé de plusieurs génotypes non apparentés avec un nombre limité de gamètes par génotype, comparé à celui de Bahrman et Damerval (1989) qui ont un bon nombre de données sur un seul génotype, justifie certainement une prudence plus grande pour la détection d'une liaison.

#### D. DISCUSSION

La méthode des lod scores utilisée actuellement dans la majorité des études de linkage, si elle fait référence à l'article de Morton (1955) ne correspond pas exactement au test décrit par cet auteur. Celui-ci propose l'utilisation d'un test séquentiel du rapport de vraisemblances, qui diffère pour plusieurs points de ce qu'on entend aujourd'hui par "lod score".

On choisit au départ une valeur  $\theta_1$ . Pour une paire de locus, le lod score correspond alors au logarithme décimal du rapport de la vraisemblance calculée en supposant  $\theta = \theta_1$  sur la vraisemblance calculée sous l'hypothèse de l'indépendance des deux locus ( $\theta = 0.5$ ). Les lod scores des différentes familles étudiées sont cumulés. Deux valeurs, B ( $B < 1$ ) et A ( $A > 1$ ) sont déterminées en liaison avec les erreurs de première et deuxième espèce. Si la somme des lod scores est inférieure à  $\log_{10}(B)$ , on conclura que la vraie valeur de  $\theta$  est supérieure à la valeur  $\theta_1$ . Au contraire, si cette somme est supérieure à  $\log_{10}(A)$  on conclura que la vraie valeur de  $\theta$  est inférieure à 0.5, donc qu'il y a liaison. Si la somme a une valeur comprise entre  $\log_{10}(B)$  et  $\log_{10}(A)$ , on ne pourra pas conclure. Le principe du test séquentiel est de compléter la somme au fur et à mesure que de nouvelles données sont collectées, et lorsque cette somme dépasse l'une des bornes A ou B, le test est terminé. Sous ces conditions, on peut généralement négliger l'excès par rapport aux bornes, c'est à dire que la somme des lod scores est proche de  $\log_{10}(A)$  ou de  $\log_{10}(B)$  à la fin du test. Cette approximation permet alors à Wald (1947), cité par Morton (1955) de calculer la relation approchée entre A, B,  $\alpha$ , risque de première espèce et  $\beta$ , risque de deuxième espèce :

$$A \approx \frac{1-\beta}{\alpha} \quad \text{et} \quad B \approx \frac{\beta}{1-\alpha}$$

Dans ces conditions, si l'on souhaite un risque de première espèce de 0.1 % et un risque de deuxième espèce de 1% on choisit  $A \approx 1000$  et  $B \approx 0.01$  et donc  $\log_{10}(A) = 3$  et  $\log_{10}(B) = -2$ . Le seuil de 3 utilisé dans de nombreux articles apparaît ici sous la forme initialement suggérée par Morton (1955).

Pourtant, ce test a été beaucoup modifié depuis sa première définition. Pour calculer le lod score, on ne se base plus sur une valeur  $\theta_1$  prédéterminée mais sur l'estimateur du maximum de vraisemblance du taux de recombinaison. Dans ce cas, les formules pour A et B ne s'appliquent plus (Chotai 1984). De plus, la notion de séquentialité a disparu, et on se contente de calculer une valeur globale sur une taille d'échantillon fixe. Chotai (1984) constate que dans ce cas, l'excès par rapport aux bornes n'est plus négligeable, et cela conduit à des erreurs de première et de seconde espèce plus faibles que celles prévues par Wald. La même observation a été faite ailleurs : le test est plus conservateur que ce qu'annonce Wald (Guihenneuc 1990).

Pour trouver une justification plus précise aux seuils de lod score habituellement utilisés dans le cadre du test tel qu'il a été modifié, Chotai (1984) reprend une inégalité établie par Haldane et Smith en 1947 pour leur test de rapport de vraisemblances avec un échantillon de taille fixe. Ces auteurs montrent que le rapport de vraisemblances a une probabilité de dépasser la valeur A inférieure à  $1/A$ . Chotai (1984) montre que cette inégalité s'applique sous certaines conditions au lod score, et que donc l'erreur de première espèce liée à un seuil de lod score de  $\log_{10}(A)$  est environ égale à  $1/A$ . Pour tester cela sur nos données, nous avons comparé les niveaux de signification du test du  $\chi^2$  avec l'équivalent de  $1/A$  ( $1/10^{\text{lod}}$  ici) pour les couples de locus déjà étudiés (figure 12). On constate que la droite correspondant à l'égalité entre les niveaux de signification calculés pour le test du  $\chi^2$  et  $1/A$  majore les valeurs observées. Pour les lod scores les plus élevés (figure 13)  $1/A$  est assez proche de l'erreur de première espèce calculée, mais quand le lod score est plus faible, l'erreur commise est toujours inférieure à  $1/A$ . Ici encore, le test est plus conservateur que prévu.

D'après les différents résultats qui viennent d'être présentés, on peut finalement considérer qu'un seuil de 3 pour le lod score correspond à une erreur de première espèce de 0.1% au maximum. Cette valeur peut sembler faible par rapport aux niveaux habituellement choisis pour les tests. La justification de Morton (1955) est classiquement reprise par les généticiens. Il souligne l'importance, dans une étude de linkage, d'éviter de détecter une liaison entre deux locus indépendants. Donc, le risque de première espèce doit être minimisé. En supposant qu'un organisme possède n chromosomes de longueurs égales, Morton suppose que deux locus tirés au hasard ont une probabilité a priori d'être liés  $\phi = 1/n$ . Si  $\alpha$  représente la valeur du risque de première espèce et P la puissance du test utilisé, alors la probabilité a posteriori que deux locus aient été déclarés liés à tort s'écrit :

$$p = \frac{\alpha(1-\phi)}{(\alpha(1-\phi) + \phi P)}$$

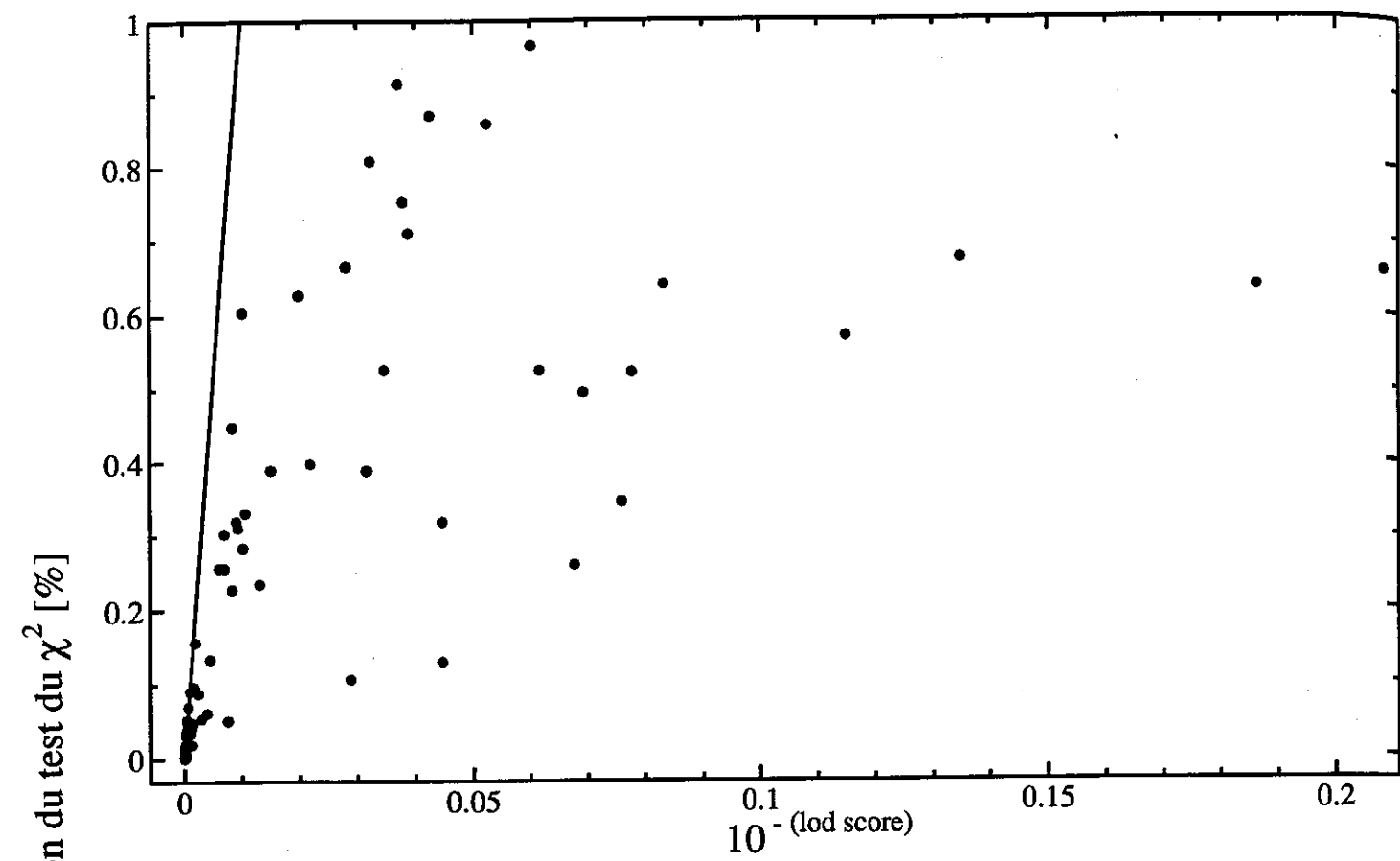


Figure 12 : Relation entre le niveau de signification du test du  $\chi^2$  et la valeur de  $10^{-(\text{lod score})}$ .

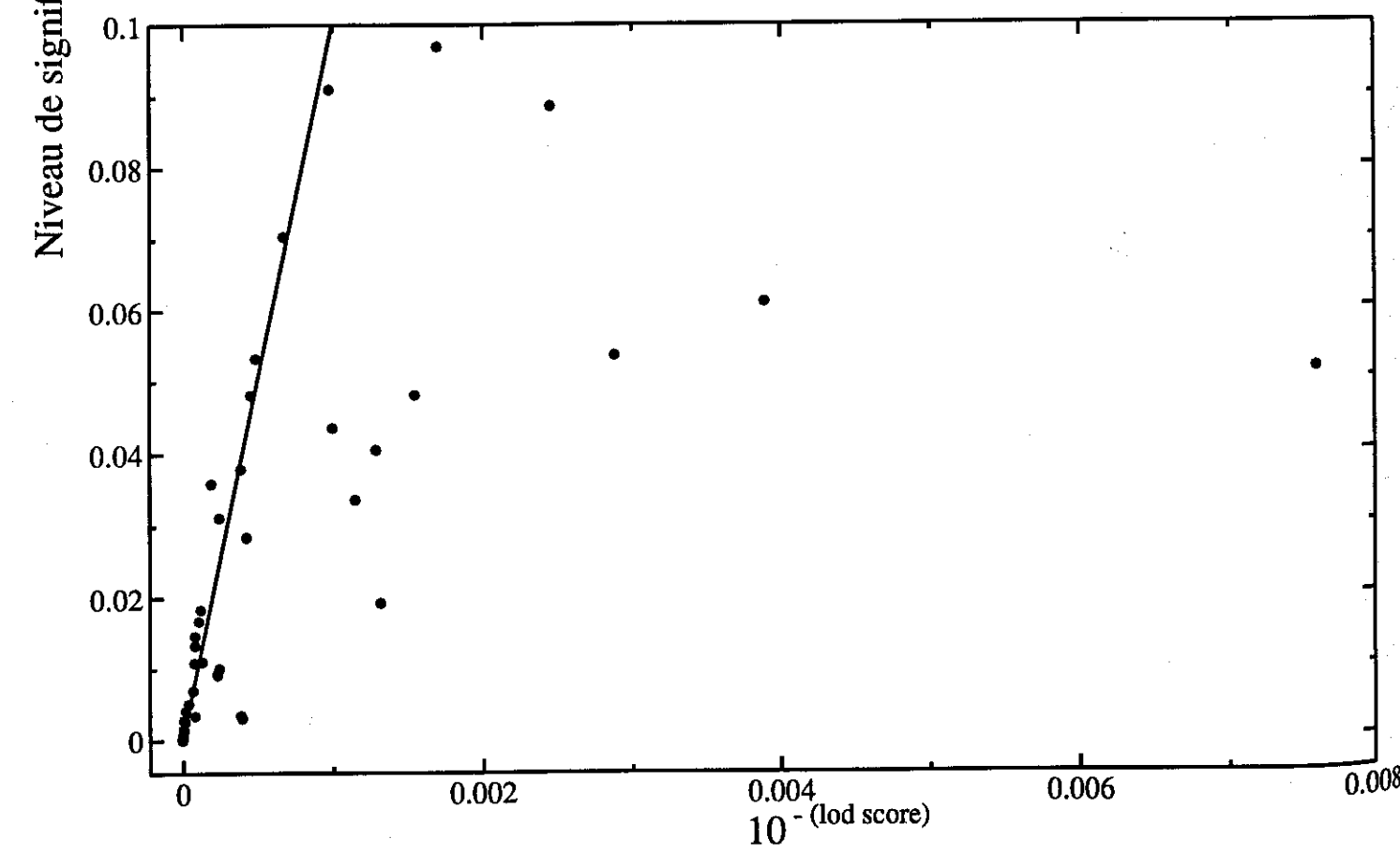


Figure 13 : Points de la figure 12 pour lesquels le niveau du test du  $\chi^2$  est inférieur à 0.1%. La droite représente  $\chi^2 = 10^{-(\text{lod score})}$ .

Si l'on souhaite que la valeur de  $\rho$ , erreur de première espèce a posteriori, soit faible (0.01), avec une valeur de  $\phi$  approchée dans notre cas par 0.083 (12 chromosomes), et une puissance de test de 0.9, il faut donner à  $\alpha$  une valeur très faible (0.0008). Dans ce cas, le choix d'un lod score minimal de 3 se justifie pleinement, puisqu'il correspond à des valeurs de  $\alpha$  du même ordre.

Chotai (1984) critique cette approche, et considère que si le choix d'un niveau de signification est justifié par le fait d'atteindre une certaine erreur a posteriori, il serait plus approprié d'adopter explicitement une démarche bayésienne. Cela éviterait, selon lui, de relier  $\alpha$  à une information a priori de façon arbitraire.

Un niveau de signification faible se justifie également lorsqu'un grand nombre de comparaisons est réalisé, ce qui devient la règle avec le nombre croissant de marqueurs disponibles. Les études de liaison concernent maintenant 100, 200 ou 300 marqueurs simultanément, qui correspondent (au maximum) à 4950, 19900 ou 44850 comparaisons respectivement, le seuil de signification choisi doit en tenir compte. Par exemple, en calculant les lod scores de 81300 paires de locus, Donis-Keller et al. (1987) imposent un seuil de 4 pour déclarer une liaison. Avec ce seuil, ils détectent 3 lod scores liés à des erreurs de première espèce, entre des locus situés sur des chromosomes différents.

## E. CONCLUSION

Il n'est pas immédiat d'associer au lod score un niveau de signification précis dans notre cas. La majorité des études qui ont été consacrées à la comparaison de tests de liaison concerne la génétique humaine, où les risques effectifs apparaissent généralement inférieurs aux risques prévus. La génétique des plantes se base sur des modèles plus simples, grâce au matériel expérimental, et les conclusions ont des chances d'être un peu différentes. D'après notre étude, un lod score de 3 correspond à un risque plus élevé dans le cadre de l'analyse d'un backcross que dans l'analyse de familles non apparentées.

Le généticien préférera ne pas avoir de carte qu'une carte fautive (Morton 1955), et donc choisira un test sévère : le lod score correspond bien à cette exigence. Pour ce qui concerne le logiciel MAPMAKER, des groupes de liaison "solides", sont constitués à partir des locus déclarés liés avec un test sévère. La liaison éventuelle des locus restant par rapport à ces groupes est testée en utilisant des niveaux de lod score moins exigeants (Lincoln et al. 1990). Finalement, la sévérité du test est "tempérée" par l'analyse multipoint qui suit. Mais dans tous les cas, le niveau de signification effectivement utilisé ne peut être annoncé, le choix d'un niveau de lod score est dès lors difficile à justifier. Le seuil "historique" de 3 n'est plus systématiquement utilisé. Messeguer et al. (1991) et Devos et al. (1992), sur des populations F2, choisissent des seuils entre 2 et 3 pour les premiers et entre 2 et 2.5 pour les seconds. A partir de lignées recombinantes, Reiter et al. (1992) choisissent de doubler le seuil habituel, et retiennent un lod score de 6. Seuls Ellis et al. (1992), travaillant aussi sur des lignées recombinantes, justifient le seuil choisi par comparaison avec un test du  $\chi^2$ . Un lod score de 1.5 correspond dans leur expérience à un niveau de 5%, un lod score de 3 étant associé à un niveau de 1%.

Si le désavantage du lod score est une sévérité parfois trop grande pour la détection d'une liaison, son utilisation devient beaucoup plus critique quand des caractères dont l'hérédité est mal connue sont utilisés. Ainsi, la recherche d'associations entre maladies humaines et marqueurs par cette technique rend la signification de niveaux de lod score de 3 encore plus délicate à définir. Dans ce cas, les risques de première espèce ont des chances de ne pas être négligeables (Clerget-Darpoux et al. 1990, Risch 1992).

CHAPITRE IV

LA LIASON GÉNÉTIQUE ENTRE LES GÈNES

ET LES MARQUEURS

## IV. CARTOGRAPHIE GENETIQUE ET TAILLE DU GENOME

### A. INTRODUCTION

Les cartes génétiques construites actuellement chez les végétaux utilisent un nombre croissant de marqueurs diversifiés, et fournissent une image de plus en plus précise du génome dans son ensemble. Ainsi, l'organisation du génome peut être étudiée et la comparaison des cartes génétiques d'espèces plus ou moins éloignées permet de décrire son évolution dans le temps. Sur un plan plus appliqué, les cartes génétiques offrent la possibilité de localiser des gènes intéressants, en particulier les gènes de résistance aux maladies, et d'envisager leur clonage (Martin et al. 1991, Paran et al. 1991, Messeguer et al. 1991). Enfin, l'analyse des caractères quantitatifs grâce aux cartes est capable de détecter des zones du génome impliquées dans les variations de ces caractères (Paterson et al. 1988). Une meilleure connaissance des bases génétiques de tels caractères est donc espérée, et, parallèlement, une meilleure efficacité des programmes de sélection par l'utilisation des marqueurs.

Ces quelques points motivent en grande partie la construction de cartes chez un nombre croissant d'espèces. D'une espèce à l'autre, ces cartes recouvrent pourtant des réalités différentes. Nous avons voulu réfléchir plus particulièrement aux différentes descriptions du génome des angiospermes, en terme de quantité d'ADN et de centimorgans, pour les confronter ensuite aux données acquises jusqu'à présent chez les gymnospermes, et souligner les particularités de ces derniers et les questions qu'ils soulèvent.

### B. DESCRIPTIONS DU GENOME

#### 1. Comparaison de cartes génétiques

L'idée de comparer les cartes génétiques d'espèces différentes trouve des exemples nombreux et anciens chez les animaux. Dès 1920, Dunn compare la liaison entre le gène de l'albinisme et des yeux rouges chez le rat et la souris (Lyon 1990). De nombreux arrangements de gènes homologues entre les cartes des mammifères ont pu être décrits depuis (Gillois 1991). La conservation de groupes de liaison parmi les vertébrés, depuis les poissons jusqu'aux hommes peut permettre de formuler des hypothèses sur leur évolution (Morizot 1990). L'homologie des génomes peut également trouver des applications pratiques. En particulier, la proximité entre les cartes de la souris et de l'homme permet d'utiliser la souris comme modèle pour localiser des gènes responsables de maladies, les résultats sont alors testés sur le génome humain (Jordan 1989). Deux gènes responsables d'un diabète ont été ainsi localisés sur

la carte génétique de la souris et leurs homologues recherchés dans le génome humain (Avner 1991, Todd et al. 1991, Cornall et al. 1991).

Les comparaisons de cartes chez les végétaux ne permettent pas encore d'envisager des interventions de ce type. Tanksley (1983) décrit quelques groupes de liaison de locus isoenzymatiques homologues entre le maïs et l'orge. Plus récemment, des cartes RFLP ont permis des comparaisons à plus grande échelle. Les membres de la famille des solanacées ont le même nombre de chromosomes ( $2n = 24$ ) mais une quantité d'ADN par cellule variable selon les espèces. Les cartes génétiques de la pomme de terre et de la tomate sont apparues très semblables pour l'arrangement des marqueurs. Seuls trois des 12 groupes de liaison ont subi des inversions paracentriques, et les longueurs des cartes respectives, d'abord trouvées significativement différentes (Bonierbale et al. 1988) se sont finalement révélées très proches (Gebhardt et al. 1991). Les conclusions étaient très différentes en comparant la carte de la tomate avec celle du piment (Tanksley et al. 1988). Tous les clones d'ADN copie utilisés pour la cartographie hybridaient aussi bien sur l'une ou l'autre espèce, les mêmes gènes étaient donc présents dans les deux génomes, et bien conservés. Par contre, leurs dispositions étaient extrêmement différentes, certainement modifiées par d'importants réarrangements chromosomiques entre les deux espèces, affectant essentiellement les régions non codantes. Il est intéressant de noter que les quantités d'ADN par cellule des deux espèces diffèrent d'un facteur 4, alors que les quantités trouvées chez la pomme de terre et la tomate sont plus proches. A partir d'un ancêtre commun aux solanacées, une modification de la quantité d'ADN aurait donc été accompagnée par de grands changements du génome entre la tomate et le piment, la tomate et la pomme de terre ayant moins divergé aussi bien pour l'arrangement chromosomique que pour la quantité d'ADN (Bonierbale et al. 1988).

Weeden et al. (1992), comparant les cartes de la lentille et du pois, décrivent une situation intermédiaire entre les deux précédents exemples. Ils observent entre les deux cartes 8 régions bien conservées, qui représentent 40% du génome selon leurs estimations. Comparant de la même façon le maïs et le sorgho, Binelli et al. (1992) concluent à une grande similarité des génomes pour les régions explorées dans leur expérience.

Les modifications observées apparaissent donc d'importance variable selon les espèces comparées. L'incidence phénotypique de telles modifications reste encore à élucider. Doebley et al. (1990) étudient, à l'aide d'une carte RFLP, les zones chromosomiques impliquées dans les changements morphologiques importants entre le maïs et son ancêtre présumé, la téosinte. Certains caractères semblent sous le contrôle d'un gène majeur et de plusieurs modificateurs, d'autres sont plutôt à héritéité polygénique. L'hypothèse selon laquelle les variations dans la régulation des gènes seraient plus pertinentes pour expliquer ces différences (Hedrick et McDonald 1980) ne peut pas être étudiée directement avec les cartes classiques, mais serait intéressante à tester.

Si les cartes génétiques permettent une vision d'ensemble du génome d'une espèce, elles recouvrent des réalités différentes selon les cas, si l'on compare les quantités d'ADN des génomes et les tailles des cartes génétiques correspondantes.

## 2. Taille du génome en terme de quantité d'ADN

### 2.1. Variabilité

Les quantités d'ADN par cellule chez les angiospermes sont très variables selon les espèces. Elles varient d'un facteur 600 entre les extrêmes, *Arabidopsis*, crucifère retenue comme modèle végétal par la biologie moléculaire pour son génome très réduit (0.4 picogramme d'ADN par cellule diploïde, ou valeur  $2C$ ) et une fritillaire (255 pg/ $2C$ ) de la famille des liliacées, dont les représentants ont des génomes de grande taille (Bennet et al. 1982). La signification de telles différences de quantité pose question. Leur association possible avec des valeurs adaptatives particulières a donc été recherchée, et la définition de ces différences en terme de qualité d'ADN a été plus précisément étudiée.

### 2.2. Recherche d'une signification adaptative à ces variations

Les variations interspécifiques des quantités d'ADN ont été trouvées corrélées à la longueur et aux volumes des chromosomes. Cette quantité est également liée aux volumes cellulaire et nucléaire, au taux et à la durée des cycles de mitose (Bennet et Smith 1991). Mais au-delà de ces corrélations, la recherche d'une éventuelle signification adaptative de la variabilité de la quantité d'ADN chez les eucaryotes, ou, en anglais, le "C-value paradox", n'a donné que des résultats partiels et peu généralisables. On constate par exemple que les espèces à temps de génération court ont généralement un faible taux d'ADN, lié à un taux de développement somatique et méiotique plus court. Une plante qui a plus de 30 pg d'ADN par cellule diploïde est, par contre, très probablement une plante pérenne. La réciproque n'est pas vraie, certaines plantes pérennes ayant un génome plus petit (Orgel et Crick 1980). Grime et Mowforth (1982) proposent une explication écologique aux différences des génomes. Ils supposent que les pressions de sélection liées aux climats imposent des contraintes différentes sur la croissance et ont pu favoriser des tailles de génomes associées à des durées de mitoses plus ou moins importantes. On observerait ainsi une tendance à l'augmentation du génome du sud au nord.

Hutchinson et al. (1979), étudiant un hybride entre deux espèces de *Lolium* aux taux d'ADN différents, observent une ségrégation quantitative de ce taux dans la  $F_2$  mais ne constatent aucun effet de ce caractère sur la croissance, le développement ou la morphologie des plantes. Face à une signification adaptative difficile à définir, il ne restait qu'un pas à franchir pour supposer qu'une partie de l'ADN était sans effet phénotypique, "égoïste" et "parasite", profitant des mécanismes de l'ADN pour se

multiplier (Orgel et Crick 1980), en ayant pour seule fonction de survivre dans le génome (Doolittle et Sapienza 1980). La signification, en terme adaptatif, des variations de quantités d'ADN est donc toujours mal comprise.

### 2.3. Structure de l'ADN en relation avec sa quantité

Les techniques de cinétique de réassociation de l'ADN ont permis d'avoir une idée plus précise de la constitution des génomes : elles permettent d'estimer la part de séquences présentes avec un grand nombre de copies par rapport aux séquences uniques ou faiblement représentées dans le génome. On s'est ainsi aperçu que les végétaux possédaient beaucoup plus d'ADN répété que les animaux (Walbot et Goldberg 1979). Comme le nombre de gènes exprimés doit être à peu près constant chez les végétaux supérieurs, malgré des tailles de génomes contrastées, les différences de quantités d'ADN répété, non transcrit, seraient en grande partie à l'origine de ces variations de taille (Vedel et Delseny 1987).

Le génome vu sous l'angle de l'ADN apparaît donc très variable selon les espèces sans que l'on ait une explication précise à cette variabilité. Les cartes génétiques s'envisagent de façon parfaitement déconnectée du niveau de l'ADN. Mais plus leur précision augmente, plus les questions sur l'ADN qui les compose peuvent se poser.

## 3. Tailles du génome en centimorgans

### 3.1. Notion de carte saturée

Pour déclarer, de façon certaine, qu'un sous-ensemble de marqueurs composant une carte représente le génome en entier, il faut être en mesure d'observer qu'en ajoutant des marqueurs supplémentaires, la taille de la carte reste inchangée. Landry et al. (1991) citent un travail sur le soja, pour lequel le passage de 450 à 600 marqueurs correspondait à une taille inchangée de 2700 cM. Les cartes publiées sur cette espèce sont basées sur 150 et 250 marqueurs, et correspondent à des tailles de 1200 et 2150 cM (Keim et al. 1990, Diers et al. 1992). En supposant qu'en moyenne chaque nouveau marqueur cartographié augmente la taille de la carte de façon homogène (dans l'exemple du soja environ 8 cM par marqueur), jusqu'à la saturation, où un plateau est atteint, 70 marqueurs supplémentaires, soient 320 marqueurs au total, auraient permis à la carte de Diers et al. (1992) d'être saturée. Les cartes basées sur 450 ou 600 marqueurs seraient donc "sursaturées", couvrant le génome avec une distance moyenne entre marqueurs diminuée. Dans le cas d'*Arabidopsis*, une première carte, composée de 96 marqueurs, couvrait 501 cM (Chang et al. 1988). Une nouvelle carte, de 111 marqueurs représentait une taille proche, de 493 centimorgans (Nam et al. 1989). Même si l'on peut supposer que la taille du

génome de cette espèce approche les 500 centimorgans, la cartographie d'un nombre de marqueurs en "excès" serait seule capable de préciser cette valeur. Une carte récente composée de 320 marqueurs RFLP et RAPD (fragments d'ADN amplifiés de façon aléatoire) propose en effet une taille de 630 centimorgans (Reiter et al. 1992).

Les génomes de la tomate et du riz semblent à présent bien déterminés, d'après des données de Tanksley citées par Heun et al. (1991). La cartographie de 1000 marqueurs correspond ainsi à un génome de 1600 cM pour la première et 400 marqueurs décrivent 2300 cM chez le second. Dans la majorité des cartes réalisées actuellement, les tailles de génomes proposées restent donc encore des estimations. De plus, les cartes génétiques étant construites à partir du calcul des taux de recombinaison, leur taille dépend des facteurs qui influencent la recombinaison.

## 3.2. Variabilité des tailles des cartes génétiques

### 3.2.1. Effet du sexe

La différence entre les taux de recombinaison mâles et femelles d'une même espèce est un phénomène général, aussi bien chez les animaux que chez les végétaux. Un taux de recombinaison mâle plus important a été observé chez le maïs (Robertson 1984), ou chez *Arabidopsis* (Vizir et Korol 1990). Les quelques études qui permettent de comparer les recombinaisons mâles et femelles à l'échelle de cartes plus complètes concernent des espèces, où, au contraire, les taux de recombinaison femelles sont supérieurs. Chez la tomate, Vicente et Tanksley (1991), comparent deux backcross, réalisés en utilisant l'individu F1 comme mâle ou comme femelle. La carte basée sur les recombinaisons mâles représentait 1097 cM contre 1299 cM pour les recombinaisons femelles. Une observation similaire avait déjà été faite sur la même espèce, où une carte réalisée à partir d'une F2, pour laquelle les recombinaisons des deux sexes sont exprimées, était plus grande qu'une carte obtenue sur un backcross où seule la recombinaison mâle était représentée (Paterson et al. 1991). Chez l'orge, la réduction de la taille d'une carte réalisée à partir d'haploïdes doublés dérivés de culture d'anthères, témoigne du même phénomène (Graner et al. 1991). Dans l'espèce humaine, la carte génétique des autosomes a une taille supérieure de 90% chez la femme (Donis-Keller et al. 1987).

Dans ces exemples, on constate pourtant que la différence des taux de recombinaison entre les sexes d'une même espèce n'est pas uniforme sur la carte, elle apparaît plus marquée dans certains intervalles, absente ailleurs ou même inversée par rapport à la tendance générale à quelques endroits. Néanmoins, les fréquences chiasmiques mâles et femelles observées chez diverses espèces animales et végétales sont bien corrélées, ce qui suggère un contrôle commun des deux phénomènes, exprimé dans des contextes différents (Santos et Cano 1990, Burt et al. 1991).

### 3.2.2. Effet du croisement

Dans certains cas, les cartes génétiques peuvent également être perturbées par le type de croisement utilisé pour les construire. Les croisements interspécifiques permettent, chez des espèces où le polymorphisme détecté est faible, d'augmenter ce polymorphisme et de disposer ainsi de plus de marqueurs. Ceci est réalisé systématiquement chez la tomate, où l'espèce cultivée est croisée à des espèces sauvages, et chez la pomme de terre. Les croisements interspécifiques sont pourtant liés à des perturbations à la méiose, probablement expliqués par la différenciation des génomes mis en présence. On observe alors des distorsions de ségrégation qui nécessitent des traitements statistiques particuliers si les locus concernés sont utilisés pour construire une carte (Mather 1963, Wagner et al. 1992). Parallèlement, les taux de recombinaison subissent une réduction, Rick (1969) observe jusqu'à 86% de réduction chez la tomate pour certains segments, Tadmor et al (1987) constatent une réduction d'un tiers sur un groupe de liaison dans un croisement interspécifique de lentille, Douches et Quiros (1988), chez la pomme de terre, comptabilisent 89% de réduction en moyenne sur 5 intervalles. A plus grande échelle, une carte de la pomme de terre réalisée à partir d'un croisement intraspécifique avait une taille augmentée de 65% par rapport à la taille de la carte construite sur un matériel issu de croisements interspécifiques (Bonierbale et al. 1988, Gebhardt et al. 1991).

### 3.2.3. Variations individuelles

Les taux de recombinaison peuvent également être variables, pour un même intervalle, d'un individu à l'autre. Pour l'estimation des taux de recombinaison, notamment dans notre expérience, on est bien sûr obligé de faire l'hypothèse que le taux est constant quel que soit l'individu (voir l'écriture du maximum de vraisemblance, chapitre III.B.1.). De plus, et plus particulièrement si les effectifs par famille ou par individu étudié sont faibles, les tests qui doivent permettre de décider si l'hétérogénéité d'un ensemble de données est significative ou non, sont souvent très conservatifs (Nordheim et al. 1984). Comme en génétique humaine, on négligera les variations individuelles du taux de recombinaison pour pouvoir construire une carte, si les données proviennent de plusieurs familles ou individus. Ce problème s'est peu posé au départ chez les plantes annuelles, où les cartes étaient construites à partir d'individus issus d'un seul croisement. Beavis et Grant (1991) proposent une carte composite de quatre populations F2 chez le maïs, et soulignent les difficultés liées aux différences entre les quatre cartes d'origine, différences aussi observées entre les cartes de familles de lignées recombinantes chez la même espèce (Burr et al. 1988). Chez le pois, les divergences entre cartes de familles du même type semblent encore plus marquées, et incitent Ellis et al. (1992) à conclure au danger de créer une seule carte pour cette espèce. Tulsieram et al. (1992) se demandent alors si l'ordre des locus sur les groupes de liaison, qui n'est généralement pas perturbé, ne serait pas la meilleure fonction des cartes génétiques.

La difficulté de construire une carte composite, valable pour différents individus d'une population provient aussi, de façon évidente, du fait que l'on ne peut cartographier, pour un individu, que ce qui

est variable. L'individu idéal, qui serait hétérozygote à tous les locus étudiés, fournirait une carte complète. Cet individu n'existant pas, on est contraint de composer avec des individus présentant des hétérozygoties différentes. Cette contrainte, avec laquelle la génétique humaine est tout particulièrement obligée de compter, se pose à la génétique végétale dès lors que plus d'un croisement est en jeu.

Nous n'avons évoqué ici que des causes génétiques aux variations de recombinaison. Des variations liées à des effets du milieu ont été décrites (citées par Tulsieram et al. 1992). Néanmoins, il apparaît difficile de séparer les causes génétiques et environnementales, et des recherches devraient être poursuivies dans ce sens (Tulsieram et al. 1992).

### 3.2.4. Signification de la variation des taux de recombinaison

De nombreux auteurs se sont interrogés sur la signification évolutive de la variabilité des taux de recombinaison observée au sein des populations (revue par Zhuchenko et Korol 1983, Brooks et Marks 1986), ou entre les sexes (revue par Burt et al. 1991). Une réduction de la recombinaison peut permettre la "protection" de certaines associations favorables de gènes, bien adaptés, mis en place par la sélection. Au contraire, son augmentation permet l'incorporation rapide de "bons" allèles et l'élimination de "mauvais" allèles et donc une adaptation possible à des conditions changeantes. La recombinaison peut être également vue comme un moyen de réduire les déséquilibres de liaison. Selon ces hypothèses, la sélection pourra agir dans les deux sens sur la recombinaison, en l'augmentant ou la réduisant sur des parties du génome ou sur son ensemble. La recombinaison étant sous le contrôle de nombreux gènes, cette régulation pourrait intervenir à des niveaux divers (Kaul et Murthy 1985). En étudiant 20 zones encadrant des gènes chez la drosophile, Begun et Aquadro (1992) notent une corrélation positive entre la diversité en nucléotide de l'ADN et les taux de recombinaison. L'importance de ce phénomène dans d'autres espèces et sa signification sur le plan évolutif restent à étudier.

En comparant une trentaine d'espèces de mammifères, Burt et Bell (1987) trouvent une interprétation aux différences des taux de recombinaison. Ils observent une corrélation positive entre la durée de génération et la fréquence des chiasmas. La théorie de la reine rouge fournit une interprétation à cette corrélation. Cette théorie suppose en effet que plus les générations s'allongent, plus l'environnement change et plus la sélection s'accroît. Les espèces ont alors intérêt à créer de la variabilité pour faire face à ces changements. D'autre part, ces mêmes auteurs constatent que les espèces domestiques ont des taux de recombinaison systématiquement supérieurs à ceux des espèces sauvages dans leur échantillon. Ils en déduisent que des taux de recombinaison plus élevés sont indirectement sélectionnés dans les programmes d'amélioration pour leur capacité à supprimer des corrélations négatives entre des traits intéressants.

La description d'un génome en terme de centimorgans apparaît assez relative. Avec le développement des techniques de cartographie en routine, la comparaison d'un nombre croissant d'individus sera possible et la part des différents facteurs intervenant sur la taille des cartes génétiques pourra être plus précisément envisagée. Compte tenu de l'investissement nécessaire à la construction d'une carte, et pour ses utilisations ultérieures, il serait en effet important de savoir dans quelle mesure des résultats peuvent s'appliquer d'une population ou d'un individu à l'autre.

#### 4. Relation entre taille du génome en centimorgans et en paires de bases

Nous avons envisagé les deux descriptions possibles d'un génome : en terme de quantité d'ADN et en terme de recombinaison. Au premier abord, il paraîtrait logique de supposer que plus un génome est grand, plus les recombinaisons y sont nombreuses. Or si les quantités d'ADN varient considérablement, les tailles des génomes en centimorgans ont des ordres de grandeur assez proches, et ne semblent pas affectées par les différences de quantités d'ADN sous jacentes. Seule une tendance a pu être observée : la fréquence des chiasmas, comparée chez différentes espèces, décroît progressivement par unité de longueur lorsque la taille du génome augmente (Rees et Durrant 1986). Les différences d'organisation des séquences d'ADN des petits aux grands génomes pourraient expliquer cette évolution. En effet, il apparaît, au sein d'un même génome, que les taux de recombinaison ne sont pas homogènes partout, certaines séquences étant plus recombinogènes que d'autres.

Chez l'homme, des associations ont été trouvées entre des régions riches en séquences répétées et les zones de chiasmas (Chandley et Mitchel 1988). Chez les végétaux, on suppose que les blocs d'hétérochromatine, composés de séquences très répétées, correspondent à des zones de faible recombinaison (Flavell et al. 1985). L'effet de ces séquences répétées sur la recombinaison pourrait être lié à leur rôle suggéré dans les mécanismes de positionnement et d'appariement des chromosomes à la méiose (Delseny et al. 1986, Vershinin et al. 1990). Charlesworth et al. (1986) retournent le problème et supposent que les zones du génome qui, pour des raisons structurelles, subissent moins de recombinaisons accumulent plus facilement des séquences très répétées au cours de l'évolution. Thuriaux (1977) suggère que la recombinaison est concentrée dans les zones de gènes de structure. Oliver et al. (1992) observent sur un chromosome de levure une coïncidence entre les recombinaisons et les zones de transcriptions, ce qui pourrait étayer l'hypothèse précédente. Pour expliquer les hétérogénéités observées le long du génome, on pourrait alors imaginer un modèle où les gènes de structures et les zones du génome accessibles à la recombinaison seraient séparés par des blocs d'ADN, hétérochromatine ou ADN répété, où la recombinaison serait réduite (Brown et Sundaresan

1991). La proportion de ces blocs et leur répartition dans le génome pourraient être à l'origine des différences de taux de recombinaison par unité de longueur d'ADN observées.

La recombinaison semble également directement affectée par la taille des chromosomes. Chez la levure, le plus petit chromosome présente en moyenne deux fois plus de recombinaisons (en centimorgans par kilobase) que le plus grand (Kaback et al. 1992). Ces auteurs suggèrent que le taux d'interférences chiasmatisques augmenterait avec la taille du chromosome. Cette limitation physique des recombinaisons pourrait expliquer le plateau atteint par les tailles des cartes génétiques lorsque la taille du génome augmente.

La taille d'un centimorgan en nombre de paires de base va donc dépendre de l'espèce considérée. Les variations sont fortes, par exemple un centimorgan correspond environ à 3 kilobases chez la levure (Hulbert et al. 1988), et à 1000 kilobases chez l'homme (Botstein et al. 1980). Parmi les espèces végétales les plus étudiées, un centimorgan vaut environ 140 kilobases chez *Arabidopsis* (Chang et al. 1988), 500 kilobases chez la tomate (Messeguer et al. 1991), et 2000 à 4000 kilobases chez le maïs (Brown et Sundaresan 1991). Ces chiffres ne seraient que des moyennes sur l'ensemble du génome, comme nous l'avons vu. Par exemple, chez le maïs, la transposition peut entraîner des modifications locales importantes des taux de recombinaison (Dooner et Kermicle 1986). Dooner (1986) observe qu'un centimorgan aux alentours du locus bronze chez le maïs correspond à 14 kilobases, valeur plus de 100 fois inférieure à la valeur moyenne pour l'ensemble du génome. Le séquençage du chromosome III de la levure, et la confrontation des cartes physiques et génétiques, fait également apparaître des variations d'un facteur minimum de 10 du nombre de centimorgans par kilobases (Oliver et al. 1992).

Si les cartes permettent de localiser certains gènes intéressants, de résistances par exemple, leur clonage éventuel se heurtera à des difficultés très différentes si l'on s'adresse à *Arabidopsis* à la tomate ou au maïs, notamment pour réaliser des marches chromosomiques : les ordres de grandeurs des intervalles à parcourir, en centimorgans, seront très différents d'une espèce à l'autre en paire de bases, et de ce fait plus ou moins difficiles à réaliser.

Les cartes génétiques sont donc une image très déformée du génome vu sous l'angle de l'ADN, et cela d'autant plus que le génome étudié est grand. Les pins fournissent un exemple particulièrement extrême de grands génomes parmi les végétaux.



## C. GENOME DES GYMNOSPERMES

### 1. Etudes de liaison chez les gymnospermes

L'existence d'un endosperme haploïde dans la graine des gymnospermes explique le grand nombre d'études concernant l'hérédité et la liaison des locus isoenzymatiques exprimés dans cet organe.

La majorité de ces études concerne les pins. Plus de dix espèces ont été étudiées pour une quinzaine de locus dans chaque cas (Guries et al. 1978, Rudin et Eckberg 1978, O'Malley et al. 1979, Eckert et al. 1981, O'Malley et al. 1986, Furnier et al. 1986, Strauss et Conkle 1986, El-Kassaby et al. 1987, Shiraishi 1988, Szmidi et Muona 1989). Le nombre de locus révélé et par conséquent l'étendue des cartes réalisées restent limités. Les études de Conkle (1981) sur plusieurs espèces de pins et Niebling et al. (1987) sur le pin sylvestre restent les plus étendues. Ils décrivent chacun une vingtaine de locus répartis dans 5 à 7 groupes de liaison couvrant jusqu'à 250 centimorgans au total.

Des études semblables ont été également réalisées sur d'autres conifères, le douglas (El-Kassaby et al. 1982, El-Kassaby et al. 1987, Adams et al. 1990), les épicéas (King et Dancik 1983, Boyle et Morgenstern 1985, Ernst et al. 1987, Geburek et von Wuehlisch 1989), le cèdre (Harry 1986), le mélèze (Cheliak et Pitel 1985), le sapin (Neale et Adams 1981), ou le thuya (Xie et al. 1991). A l'issue de ces nombreuses études, il est intéressant de constater que certaines liaisons génétiques détectées sont communes aux différentes espèces de pins et sont observées également chez d'autres genres plus éloignés. En particulier, la liaison étroite entre les locus *Got-1* et *Pgi-2*, signalée par Guries et al. (1978), est retrouvée par Conkle (1981) dans trois autres espèces de pins, et est présente entre autres dans le genre *Picea* (Geburek et von Wuehlisch 1989), chez le cèdre (Harris 1986), ou chez le mélèze (Cheliak et Pitel 1985). La comparaison des génomes des conifères mériterait d'être étendue.

En effet, les conifères, et les pins en particulier, semblent avoir été très conservés au cours du temps, leurs fossiles en témoignent. Prager et al. (1976) s'interrogent sur les bases génétiques de cette évolution apparemment lente. Une étude immunologique comparative des protéines des graines leur permet de constater que le taux d'évolution des gènes de structure de ces espèces est à peu près le même que celui des angiospermes. Ils soulignent par contre que leur caryotype semble avoir évolué beaucoup plus lentement. Ainsi, le nombre de chromosomes ( $2n = 24$ ) apparaît très conservé à travers les différentes espèces, et l'aspect des chromosomes ainsi que les appariements à la méiose lors d'hybridations suggèrent de faibles réarrangements chromosomiques (Saylor 1972). Prager et al. (1976) supposent alors que l'organisation des chromosomes serait plus importante pour l'évolution des organismes que les mutations ponctuelles, et qu'une évolution chromosomique lente serait peut-être le signe d'un changement lent dans l'organisation des gènes de régulation.

La comparaison de cartes génétiques détaillées des conifères, même si elle ne peut répondre directement à cette hypothèse, pourrait permettre d'avoir une meilleure idée des différences ou ressemblances entre les génomes. Comme nous l'avons fait chez le Pin maritime, l'utilisation de l'électrophorèse bidimensionnelle pour réaliser des cartes génétiques chez différentes espèces, puisqu'elle donne accès à des variabilités de quantités de protéines, liées à des phénomènes de régulation, pourrait apporter des éléments intéressants quant à l'hypothèse de Prager et al. (1976).

### 2. Quantité d'ADN

Le génome des gymnospermes est caractérisé par sa grande quantité d'ADN. Le contenu d'ADN par cellule diploïde varie entre 10 et 60 picogrammes selon les espèces. Les pinacées possèdent les plus fortes valeurs parmi les gymnospermes, 40 pg en moyenne, et ce taux est assez constant d'une espèce à l'autre dans le genre *Pinus* (Ohri et Khoshoo 1986, Teoh et Rees 1976). Le Pin maritime possède environ 49 pg d'ADN par cellule diploïde. Par comparaison avec des espèces végétales très étudiées, cette quantité serait plus de 9 fois supérieure à celle du maïs, 12 fois supérieure à celle de la tomate ou plus de 100 fois supérieure à celle d'*Arabidopsis* (Bennet et Smith 1976, 1991). Parmi les espèces d'arbres angiospermes, le frêne a une quantité d'ADN de 3 pg, l'orme et le chêne d'environ 2 pg, le saule et le bouleau autour de 1 pg par cellule diploïde (Bennet et Smith 1991). La taille du génome des pins semble démesurément grande par rapport à toutes ces valeurs. La question de la nature et de la signification de cet excès d'ADN se pose donc particulièrement dans leur cas.

### 3. Structure du génome des conifères

Chez les conifères, les séquences d'ADN uniques, ou présentes avec quelques copies (autour de 3), représentent 20 à 30% du génome (Rake et al. 1980). Le maïs aurait à peu près la même proportion de séquences simple copie, donc la différence des génomes semble mal expliquée par un excès de séquences répétées chez les conifères. Une hypothèse, pour des grands génomes de ce type, suggère que la plus grande part des séquences d'ADN présentes en quelques copies est composée d'anciennes séquences répétées ayant beaucoup divergé (Flavell 1980, Kriebel 1985). Seule une très petite partie de ces séquences seraient transcrites (Kriebel 1985). La structure du génome des conifères est assez mal connue, mais les quelques résultats obtenus jusqu'à présent laissent présager une organisation différente de celle des angiospermes.

Dans les quelques cas où des séquences codantes de l'ADN ou des séquences d'acides aminés de protéines ont été comparées entre des angiospermes et des conifères, ces séquences apparaissent bien conservées. Des homologies de 70 à 90% ont été ainsi trouvées pour l'alcool déshydrogénase et l'aldolase (Kinlaw et al. 1987, Kinlaw et al. 1990), pour la chlorophylle A/B binding protein (Jansson et Gustafson 1987), pour des protéines impliquées dans le photosystème II (Jansson et al. 1990), pour

la petite sous-unité de la RubisCO (Yamamoto et al. 1988, Hutchison et al. 1990) ou pour la grande, codée dans le chloroplaste (Hipkins et al. 1990), ou encore pour les séquences codantes des gènes ribosomiques (Cullis et al. 1987). Les protéines citées ici en exemple font partie de processus clés du fonctionnement de la plante, ce qui a pu contribuer à leur conservation au cours de l'évolution. Leur structuration dans le génome, par contre, a pu être modifiée. Deux cas ont été plus particulièrement étudiés chez des pins.

Le premier concerne l'alcool déshydrogénase (ADH). Les profils obtenus par hybridation d'une partie de l'ADN copie d'un gène d'ADH sur l'ADN génomique de pins sont apparus plus compliqués que ce qui était observé habituellement chez les angiospermes. Les nombreuses bandes détectées suggèrent que les gènes d'ADH des pins sont plus grands ou plus nombreux que ceux des angiospermes. Les séquences étant très proches, une taille plus grande de ces gènes supposerait la présence d'introns plus nombreux ou plus grands. Si ces gènes étaient plus nombreux, une grande partie serait inactive, puisque, dans les études isoenzymatiques chez les conifères, les auteurs retiennent le plus souvent un seul et au mieux deux locus pour cette enzyme (Harry et al. 1989, Kinlaw et al. 1990). On pourrait également faire l'hypothèse que les gènes sont exprimés différemment selon les organes ou dans le temps, et dans ce cas, le nombre de locus mis en évidence par des études isoenzymatiques, réduites à un organe et à un stade de développement, ne refléterait pas le nombre total.

La seconde étude concerne les gènes ribosomiques. Ces gènes correspondent à des séquences répétées de l'ADN, et leurs caractéristiques, bien décrites chez les angiospermes (Vedel et Delseny 1987), se retrouvent dans le génome de *Pinus radiata*, sauf pour deux d'entre elles. Chez les angiospermes, la longueur de l'unité répétée des gènes 18S-25S est variable selon les espèces, la plus grande (12000 paires de bases) ayant été trouvée chez le blé. Le pin présente une unité plus de deux fois supérieure, de 27000 paires de bases (Cullis et al. 1987). Chez des épicéas, des tailles d'unités encore plus importantes ont été décrites, de 35000 à plus de 43000 paires de bases (Bobola et al. 1992). De plus, les gènes 18S-25S et 5S sont présents dans le génome du pin à des sites apparemment nombreux, alors qu'ils ne sont localisés qu'à quelques endroits chez les angiospermes (Cullis et al. 1987). La présence dans le caryotype des pins d'un nombre important d'organismes nucléolaires conforte cette observation (MacPherson et Fillion 1981).

Les gènes de l'ADH et les gènes ribosomiques, représenteraient donc une part plus importante du génome des pins que de celui des angiospermes. Le nombre de gènes fonctionnels parmi eux reste encore à étudier. L'observation du génome à l'échelle moléculaire devrait permettre une meilleure compréhension de la nature de l'excès d'ADN trouvé dans ces génomes, encore peu décrits.

#### 4. Variabilité des taux de recombinaison

##### 4.1. Effet du sexe

La comparaison entre recombinaison mâle et femelle d'un même arbre peut être réalisée chez les pins. L'étude des marqueurs exprimés dans les endospermes haploïdes des graines d'un arbre donné permet un accès direct à la recombinaison femelle. La recombinaison mâle s'étudie à partir d'un croisement contrôlé dans lequel le pollen du même arbre a été utilisé. La soustraction du profil de l'endosperme haploïde à celui de l'embryon d'une même graine issue de ce croisement permet de reconstituer l'apport paternel. La méiose mâle peut alors être étudiée par cette technique à partir d'un nombre suffisant de graines.

Une seule étude de ce type existe. Elle a été faite chez *Pinus radiata* et est réduite à une seule liaison impliquant deux locus isoenzymatiques, Adh-2 et Pgi-2. Le taux de recombinaison mâle a été trouvé supérieur de 43% en moyenne sur cet intervalle chez 6 individus. Un deuxième intervalle (Got-2 Pgi-2) étudié sur un seul arbre, montrait la même tendance (Moran et al. 1983). Nos données de recombinaison chez *Pinus pinaster* ont été obtenues sur endospermes, et traduisent donc la recombinaison femelle. Si la recombinaison mâle s'avérait supérieure sur des groupes de liaison plus étendus, ce qui reste à étudier, la taille du génome définie par les méioses mâles serait supérieure à l'estimation que nous avons obtenue.

##### 4.2. Variations individuelles

La comparaison du taux de recombinaison entre les locus Got-1 et Pgi-2, dont la liaison, comme nous l'avons vu, est caractéristique des conifères, a été faite chez des épicéas. Des variations ont été détectées selon les arbres considérés (Barrett et al. 1987, Ringius et Innes 1990). Des taux de recombinaison variables selon les individus ont également été signalés pour quelques intervalles isoenzymatiques chez divers conifères, sans que les facteurs responsables, génétiques ou environnementaux, aient pu être mis en évidence (Rudin et Eckberg 1978, Strauss et Conkle 1986, Perry et Knowles 1989, Gebureck et von Wuehlisch 1989). Pour ce qui concerne notre étude, nous n'avons pu comparer la carte obtenue à partir des 18 arbres avec celle de Bahrman et Damerval (1989) que de façon limitée. Certaines variations ainsi que certaines covariations, et notamment celles des

protéines les plus importantes du gel ont été observées dans les deux études. Par contre, une comparaison de distances entre locus n'a pas été possible. En effet les locus communs n'étaient pas assez nombreux. On peut supposer que ces différences sont liées à un effet d'échantillonnage, Bahrman et Damerval (1989) ayant construit leur carte à partir d'un seul pin d'origine italienne, et notre carte ayant été construite grâce à 18 pins des Landes.

Une comparaison entre arbres menée sur un grand nombre de locus reste donc à réaliser. Elle permettrait de mesurer l'importance du facteur individuel, et d'estimer la possibilité de transférer une carte construite pour un individu à un autre individu.

### 5. Relation entre centimorgans et kilobases

Selon Rake et al. (1980), le génome des pins comporte entre 10 et  $20 \cdot 10^6$  kilobases. D'après nos estimations 2000 centimorgans couvriraient la totalité du génome. Dans ces conditions, un centimorgan représenterait de 5000 à 10000 kilobases chez les pins. L'espace physique entre deux marqueurs d'une carte saturée sera considérable. La localisation de gènes intéressants et leur clonage chez les espèces végétales bien étudiées ne fait que débiter. Cet objectif paraît lointain chez les pins, mais lorsqu'il sera atteint, devra adopter des techniques particulières pour être réalisable. Chez l'homme, l'amplification de fragments d'ADN particuliers à partir de spermatozoïdes permet d'envisager l'estimation de taux de recombinaison très faible, sur des distances physiques réduites (Li et al. 1988, Cui et al. 1989, Arnheim et al. 1991). Cette technique pourra être transposée à l'ADN des mégagamétophytes des graines de pin pour réaliser des cartographies précises de zones intéressantes. La corrélation entre carte physique et carte génétique pourra se faire grâce à la technique de l'hybridation *in situ*, qui commence à être utilisée chez les conifères (Brown et al. 1992).

### D. CONCLUSION

Les gymnospermes apparaissent à la dernière période de l'aire primaire (-280 millions d'années), soit 140 millions d'années avant les premières angiospermes. L'étude plus approfondie du génome de représentants de ce groupe pourrait permettre d'aborder deux points particuliers.

Malgré leur ancienneté, les gymnospermes semblent avoir été très conservés, et il serait intéressant de mesurer l'expression de cette conservation au niveau du génome. D'autre part, les gymnospermes, et les pins en particulier pourraient être un modèle d'espèce à grand génome, notamment pour la recombinaison, l'étude de sa variabilité au sein du génome et entre individus. La possibilité de comparer des gamètes par le biais des endospermes haploïdes des graines donne en effet facilement accès à la recombinaison. Les techniques récentes qui permettent de révéler le polymorphisme de fragments d'ADN amplifiés aléatoirement ou RAPD (random amplified polymorphism DNA),

fournissent rapidement un grand nombre de locus (Williams et al. 1990, Caetano-Anollés et al. 1991). Elles sont à présent appliquées à l'ADN contenu dans les endospermes de pins, d'épicéas ou de sapin Douglas. Une carte réalisée par cette technique avec 60 endospermes d'un pin (*Pinus taeda*) localise 191 marqueurs sur 12 groupes de liaison, certains groupes mineurs ayant été ignorés, et couvre 1687 centimorgans, ce qui, d'après notre estimation de la taille du génome, représenterait plus de 80% du génome (Grattapaglia et al. 1992). Des cartes de ce type seront également bientôt disponibles chez d'autres conifères, en particulier chez un épicéa, *Picea glauca* (Carlson et al. 1991, Hong et al. 1992). Au niveau diploïde, des cartes sont construites à partir de pedigrees en utilisant les RFLP (Devey et al. 1991). Chez *Pinus taeda*, une carte localisant 91 marqueurs sur 17 groupes de liaison couvre environ 590 centimorgans (Neale et al. 1992).

Un résultat préliminaire concernant la carte génétique d'un croisement interspécifique entre deux pins (*Pinus caribaea* \* *Pinus elliottii*) réalisée avec des marqueurs RAPD localise 217 marqueurs sur 3837 centimorgans, sur 18 groupes de liaison (Dale et al. 1992). La taille de cette carte semble étonnamment grande, et double quasiment les 2000 centimorgans de notre estimation. Il apparaît d'une part que les marqueurs RAPD nécessitent une certaine prudence d'utilisation, notamment pour le choix des polymorphismes à retenir. D'autre part, le logiciel MAPMAKER n'est pas adapté à la construction d'une carte à partir des données d'endospermes d'un seul pin, et conduit à des résultats parfois curieux. Les cartes RAPD de *Pinus taeda* et du croisement interspécifique seront donc réexaminées grâce à des programmes informatiques plus adaptés. Malgré ces réserves, il semble que le croisement interspécifique donne lieu à une carte significativement plus grande que la carte intraspécifique (O'Malley, communications personnelles). Ceci pourrait constituer une originalité supplémentaire des gymnospermes par rapport aux angiospermes.

Ces différents travaux, qui sont pour la plupart encore inachevés, témoignent de la multiplication des études consacrées au génome des gymnospermes. Si les connaissances concernant ces génomes sont encore très partielles, on peut supposer que de nouvelles données viendront rapidement les compléter. Parallèlement, l'intérêt porté à une meilleure compréhension du déterminisme génétique des caractères quantitatifs chez ces espèces devrait apporter des résultats nouveaux.

## V. RELATION ENTRE MARQUEURS ET CARACTERES QUANTITATIFS

### A. INTRODUCTION

L'amélioration du Pin maritime a débuté entre 1960 et 1974 par des sélections phénotypiques en forêt. Le massif landais a été parcouru pour repérer, au sein des peuplements homogènes âgés de plus de 30 ans, des arbres "plus". Ceux-ci, comparés à leur 30 voisins les plus proches étaient jugés supérieurs pour leur volume et leur forme. La population d'amélioration de départ a été ainsi constituée de 550 arbres, copiés sous forme de greffes en parc à clones. Des descendants de ces arbres ont été installés en plantations multilocales. Les plantations les plus anciennes concernent des familles de demi-frères issues de pollinisations libres en forêt de ces arbres. Les plus récentes sont constituées des familles de pleins-frères obtenues à partir de croisements contrôlés, réalisés en parc à clones selon des plans factoriels, hiérarchiques ou diallèles.

Des études détaillées du déterminisme génétique de la croissance en hauteur, et des différents facteurs intervenant dans son expression ont conduit à un classement des arbres "plus". Basé sur les comparaisons des valeurs propres de ces génotypes en parc à clones et de leurs aptitudes générales à la combinaison mesurées dans les tests de descendance âgés de 12 ans au moins, le classement retient deux critères. D'une part la croissance en hauteur elle-même, pour laquelle la part de variance génétique additive dans la variance génétique totale est importante (Kremer 1981b) et d'autre part la capacité de certains génotypes à faire plus d'une pousse par an, ou polycyclisme, dont l'expression est héritable et stable dans le temps et l'espace (Kremer 1981a). Les 18 arbres de notre échantillon traduisent la gamme de variation de ces deux caractères. La classification de ces arbres fournit un matériel de choix pour les études particulières à réaliser dans le cadre général de la réflexion sur la sélection chez les arbres forestiers.

#### 1. Sélection des arbres forestiers

La sélection des arbres vise à l'amélioration de différents caractères, la production en volume en particulier, au moment de l'exploitation. Or, chez le Pin maritime par exemple, les arbres sont jugés exploitables vers 50 ans. La sélection intervient donc forcément bien avant l'âge d'exploitation, à partir de caractères évalués sur des arbres jeunes. Le système de test dans lequel est réalisée la sélection est loin du système dans lequel on attend la réponse à cette sélection : il s'agit d'un cas typique de sélection indirecte (Gallais 1990). Pour qu'une sélection indirecte, basée sur des caractères mesurés sur des arbres jeunes soit efficace, il faut, selon Jiang (1987) :

- que le caractère juvénile soit plus héritable que le caractère adulte, ce qui est probable, sachant que les conditions environnementales sont mieux contrôlées dans des tests précoces,
- que l'intensité de sélection soit plus forte au stade jeune, ce qui est également réalisable, un grand nombre d'individus pouvant être mis en tests précoces,
- que la corrélation entre le caractère mesuré au stade jeune et le caractère mesuré au stade adulte soit forte.

C'est ce dernier point qui pose problème. La croissance en hauteur est un bon prédicteur de la croissance en volume, plus héritable et simple à mesurer (Kremer 1981a). Pourtant, l'observation d'un manque de corrélation entre ce caractère mesuré sur les arbres jeunes et le même caractère mesuré sur les arbres adultes est générale. Dans le cas du Pin maritime, les résultats de Kremer (1992) montrent, d'une part, que la sélection phénotypique des arbres "plus" en forêt a été efficace mais, d'autre part, qu'une sélection pour la hauteur totale d'arbres âgés de moins de 12 ans pourrait entraîner des gains génétiques nuls. En effet, les corrélations génétiques entre la hauteur totale mesurée à 50 ans et les hauteurs mesurées jusqu'à 10 ans sont négatives ou nulles. Ces corrélations deviennent positives et augmentent linéairement avec l'âge à partir de 10 ans (Kremer 1992). L'étude de la croissance des arbres aux stades juvéniles et adultes apparaît donc importante pour la sélection.

## 2. Corrélations entre caractères juvéniles et adultes

Les arbres jeunes et les arbres adultes correspondent à des environnements "internes" et "externes" très différents (Schmid 1992). Ceci pourrait expliquer le manque de corrélations juvénile-adulte (Kremer et al. 1991). Les plantules sont soumises à des effets maternels par l'intermédiaire de l'endosperme de la graine et se trouvent dans les conditions favorables de la pépinière la première année. A l'installation en forêt, l'environnement devient plus sévère. Parallèlement, les plantes subissent une maturation, et on peut faire l'hypothèse que les états juvéniles et adultes témoignent de programmes génétiques différents qui modifient l'expression de la croissance (Poethig 1990).

Les effets maternels, mesurés par l'effet du poids de la graine sur les premières phases de croissance des plantules ne semblent pas systématiques. Il sont parfois observés (Khalil 1985, Toon 1991) mais peuvent également être absents (Kremer et Larson 1983, Bridgwater 1990). Chez le Pin maritime des effets maternels (effets réciproques mesurés dans un croisement diallele) ont été mis en évidence sur les caractères mesurés en première saison de végétation seulement (Guignard 1983). Lorsque ces effets sont détectés, ils disparaissent toujours assez rapidement, et ne semblent donc pas pouvoir expliquer à eux seuls la différence entre caractères jeunes et adultes.

Pour étudier de plus près l'expression de la croissance au cours de la vie de l'arbre chez le Pin maritime, une première approche a consisté à décomposer ce caractère complexe en composantes plus simples, dans le but de se rapprocher du fonctionnement cellulaire (Kremer et Roussel 1982, Kremer 1984). L'accroissement total est vu comme une somme d'accroissements annuels. Chaque accroissement annuel est composé d'une pousse primaire et d'éventuels cycles supplémentaires. Chacune de ces pousses est considérée dans une dernière étape selon le nombre d'entre-noeuds présents multipliés par leur longueur moyenne. Ces différents niveaux ont été étudiés au cours de tests rétrospectifs, où le comportement de jeunes plants descendant des arbres "plus" bien connus au stade adulte, a été analysé dans des conditions de croissance accélérée, qui doivent hâter la maturation (Lascoux 1992). Un grand nombre de réactions des plants, tant sur la morphologie que sur le développement ont pu être observées. Des différences prononcées entre familles étaient mises en évidence. De plus, le classement adulte des familles a pu être retrouvé sur plusieurs caractères, sans faire ressortir un caractère plutôt qu'un autre. Si ces caractères peuvent être utilisés pour une sélection précoce, il semble pourtant important d'essayer de mieux comprendre les mécanismes biologiques sous-jacents (Lascoux 1992). Compte tenu de la difficulté à obtenir des résultats très concordants sur les corrélations juvénile-adulte, notamment du fait des problèmes méthodologiques liés à leur estimation (Kang 1991, Lascoux 1992), l'étude des bases génétiques du développement apparaît indispensable (Namkoong et al. 1988).

## 3. Etude des bases génétiques de la croissance

L'étude de l'hérédité de caractères complexes comme la croissance à différents moments de la vie de l'arbre peut être réalisée grâce à l'utilisation des marqueurs du génome. Les croissances juvénile et adulte seront considérées comme des caractères différents. Ces caractères pourront être disséqués en distinguant dans le génome les facteurs génétiques ayant des influences mineures ou majeures sur leur expression (Stuber 1989a). Un exemple de cette approche concernant la croissance existe chez le maïs. Edwards et al. (1992) étudient les croissances précoces et matures, et deux de leur composantes, le nombre d'entre-noeuds et la longueur de ces entre-noeuds. Grâce à une cartographie RFLP, ils détectent des régions qui affectent la hauteur totale de la plante, distribuées sur tout le génome. Ces régions ont des comportements variés sur les deux mesures de croissance, et sur les deux composantes de la hauteur : elles affectent soit l'un des caractères, soit l'autre ou les deux simultanément. Les programmes génétiques liés aux croissances juvénile et adulte chez le maïs pourraient donc être relativement indépendants mais chevauchants, comme le suggère Poethig (1990), puisque des zones génomiques communes y sont impliquées. Si, avec les cartes génétiques denses, des locus à effets quantitatifs ont pu être détectés et cartographiés chez des espèces comme le maïs, la recherche de tels locus chez des arbres pose des problèmes méthodologiques particuliers, et n'a pas encore donné de résultats.

Nous possédons une description du génotype des arbres constituant notre échantillon pour un bon nombre de locus, impliqués dans des variations de position, de présence/absence et de quantités de protéines. Nous connaissons également les performances de ces arbres évaluées dans différentes conditions et à plusieurs âges. Nous proposons donc, à travers une étude très prospective, une confrontation des deux niveaux d'information, en tenant compte des différents types de locus. Nous tenterons dans un premier temps une approche multilocus du problème, en étudiant l'hétérozygotie des arbres, et l'ensemble des locus simultanément, l'effet individuel des locus étant abordé dans un second temps.

## B. ETUDE MULTILOCUS

### 1. Hétérozygotie des arbres

#### 1.1. Nombre de locus hétérozygotes

A l'issue de l'analyse des variations des protéines exprimées dans les endospermes des 18 pins de notre échantillon, et de leur interprétation génétique, nous disposons d'une description du génotype de chacun des arbres à 84 locus. Nous pouvons déduire de l'observation des 10 à 14 endospermes haploïdes par arbre, si l'arbre était homozygote ou hétérozygote à chaque locus. Un arbre hétérozygote à un locus donné aurait été déclaré à tort comme homozygote si un seul des deux allèles avait été observé parmi l'ensemble des  $k$  endospermes analysés chez cet arbre. La probabilité d'un tel événement est  $P = (1/2)^{k-1}$  (Morris et Spieth 1978). Dans notre cas, les probabilités de fausses classifications sont comprises entre 0.2% pour 10 endospermes et 0.012% pour 14 endospermes, donc restent assez faibles.

Les 84 locus détectés hétérozygotes chez un arbre au moins peuvent être classés dans quatre catégories selon leur effet sur les protéines (Klose 1982) :

- 28 locus (33%) sont responsables de variations de positions de certaines protéines (une position  $i$  correspondant à un allèle  $v_i^*$ ),
- 15 locus (18%) sont liés à la grande quantité (allèle  $q_2$ ) ou à la petite quantité (allèle  $q_1$ ) de spots,
- 31 locus (37%) induisent la présence (allèle  $p$ ) ou l'absence de spots (allèle  $a$ ),
- 10 locus (6%) associent les différents types d'allèles des cas précédents (ces locus correspondent aux cas présentés chapitre II.C.).

\* L'identification des différents types d'allèles telle qu'elle est définie ici sera conservée par la suite.

Les deux premières catégories de locus correspondent à des variations claires de spots. Des variations de charge ou de poids dans le premier cas, liées soit à une modification du gène de structure soit à une modification post-traductionnelle de la protéine. Des variations de la quantité dans le deuxième cas, qui mettent en évidence la variabilité d'un facteur de régulation, dont l'intervention sur la protéine est possible à de nombreuses étapes, de la transcription du gène à sa traduction et au-delà (voir Leonardi 1989). Comme nous l'avons déjà signalé, la troisième catégorie de locus correspond à des variations ambiguës. La variabilité du gène de structure peut être révélée si un allèle nul est en cause par exemple. Mais l'absence du spot peut aussi correspondre à une quantité en dessous du seuil de détection, et la variabilité d'un facteur de régulation, distinct du gène de structure serait révélée dans ce cas. La comparaison de plusieurs lignées pour plusieurs organes chez le maïs a permis de mettre en évidence des variabilités de ce type (10% des variations de type présence/absence initiales), lorsqu'un spot absent dans un organe était détecté dans un autre organe (Leonardi 1989).

En moyenne, 32 locus sur les 84, soit 38%, sont trouvés hétérozygotes chez chaque arbre de l'échantillon (tableau 6). L'hétérozygotie pour les locus responsables de variations de position est de 36% en moyenne, pour les variations de quantité de 33%, pour les présences/absences de 39%.

#### 1.2. Comparaison des hétérozygoties

A la suite de travaux sur la drosophile, l'homme et la souris, les niveaux d'hétérozygotie étudiés avec l'électrophorèse bidimensionnelle avaient été trouvés systématiquement plus faibles que ceux obtenus avec les électrophorèses isoenzymatiques classiques (Leigh Brown et Langley 1979, McConkey et al. 1979, Racine et Langley 1980b). Les hypothèses émises pour expliquer ces observations sont de deux ordres. L'électrophorèse à deux dimensions aurait une sensibilité inférieure à l'électrophorèse classique. Par exemple des modifications de la conformation, détectées pour les isoenzymes, seraient détruites par l'extraction dénaturante en deux dimensions (Walton et al. 1979). La seconde hypothèse suggère que les isoenzymes seraient un sous-ensemble protéique biaisé, plus polymorphe. L'électrophorèse en deux dimensions donnant accès à des types protéiques plus variés, ce biais serait éliminé. Des études ultérieures ont pourtant montré qu'une amélioration de la technique accompagnée de l'étude de l'hérédité des variations observées, conduisait à des mesures de niveaux d'hétérozygotie plus comparables (Rosenblum et al. 1983). La comparaison des hétérozygoties fournies par les deux techniques reste pourtant délicate.

Les valeurs moyennes d'hétérozygoties que nous avons obtenues sont relatives à un échantillon d'arbre très réduit, mais sur un plus grand nombre de locus que dans le cas d'études isoenzymatiques. Selon Nei (1978), ce dispositif est adapté si l'hétérozygotie moyenne est faible.

Tableau 6 : Hétérozygotie des 18 arbres pour les différents types de locus.

Arbre (**)	Nombre de locus hétérozygotes				total
	vp <sup>1</sup>	q <sup>2</sup>	pa <sup>3</sup>	s <sup>4</sup>	
1 (++)	10	4	16	5	35
2 (++)	9	7	8	4	28
3 (++)	4	7	13	5	29
4 (+)	12	3	14	4	33
5 (+)	7	5	10	6	28
6 (+)	10	5	18	5*	38
7 (+0)	11	5	13	7	36
8 (+0)	9	7	13	5	34
9 (+0)	10	4	15	5	34
10 (-0)	14	6	6	5*	31
11 (-0)	8	7	14	6	35
12 (-0)	12	2	10	5*	29
13 (+-)	11	1	12	9*	33
14 (+-)	8	5	14	6*	33
15 (+-)	11	5	11	4	31
16 (-)	10	7	11	6	34
17 (-)	8	5	7	4	24
18 (-)	10	3*	12	4	29
Moyenne	10	5	12	5	32
Nombre total de locus dans chaque catégorie :					
	28	15	31	10	84

<sup>1</sup> locus liés à des variations de position de spots

<sup>2</sup> locus liés à des variations de quantité de spots

<sup>3</sup> locus liés à des présences/absences de spots

<sup>4</sup> locus associant différents types de variations

\* une donnée manquante.

\*\* classes de polyclisme et de vigueur

Les valeurs d'hétérozygotie habituellement trouvées sur la base d'études isoenzymatiques chez les populations de conifères sont parmi les plus élevées du règne végétal (environ 0,27, Mitton 1983, Furnier et al. 1991). L'échantillon de locus que nous avons considéré est biaisé dans la mesure où seuls les spots ayant été détectés avec deux formes différentes selon les gels d'un arbre au moins ont été conservés. Contrairement aux isoenzymes, il n'est pas possible de prendre en compte les locus monomorphes, d'autant que, comme nous l'avons montré, la correspondance spot/locus n'est pas immédiate : 150 spots variables correspondent à 84 locus dans notre étude. Pour tenir compte de ce biais, il est possible d'apporter une correction à l'estimation de l'hétérozygotie.

Si  $h$  représente l'hétérozygotie, et donc la probabilité pour un locus quelconque d'être hétérozygote,  $(1 - (1-h)^I)$  représente la probabilité qu'un locus tiré au hasard dans l'expérience soit hétérozygote chez l'un des  $I$  arbres étudiés au moins. Parmi les  $N$  locus détectés,  $n$  locus en moyenne sont hétérozygotes chez chacun des  $I$  arbres. On a alors :

$$\frac{n}{N} = \frac{\hat{h}}{1 - (1 - \hat{h})^I}$$

Avec  $n = 32$ ,  $N = 84$  et  $I = 18$ , la correction apportée à l'estimation  $\hat{h}$  de l'hétérozygotie est négligeable ( $\hat{h} \approx n/N = 0,38$ ).

On peut également supposer qu'il existe dans l'échantillon un biais en faveur des locus les plus polymorphes, et ceci particulièrement pour les locus responsables de variations de quantité. Un spot détecté variable parmi les gamètes de plusieurs arbres était plus susceptible d'être conservé qu'un spot variable chez un seul arbre, compte tenu de la prudence à adopter lors du dépouillement des gels pour ce type de variations. De plus, seuls des locus bialléliques ont pu être mis en évidence : l'existence de plus de deux quantités pour un même spot ne peut être établie visuellement : l'hétérozygotie a été sous-estimée, du moins pour les variations quantitatives.

Environ un tiers des locus étudiés est hétérozygote chez chaque arbre, et les différences entre classes de locus sont faibles. La valeur d'hétérozygotie proposée est assez proche des résultats obtenus sur isoenzymes. Le biais qui existe dans notre échantillon de locus en faveur du polymorphisme est peut-être du même ordre que celui qui existe avec les isoenzymes. Néanmoins, en tenant compte des remarques précédentes, on peut supposer en première approximation que la vraie valeur de l'hétérozygotie, révélée en électrophorèse à deux dimensions, doit être inférieure.

## 2. Relation entre l'hétérozygotie et les performances des arbres

### 2.1. Classes de polycyclisme et de vigueur

Nous avons voulu tester si les classes de polycyclisme et de vigueur dans lesquelles se répartissent les 18 arbres étaient ou non en relation avec leur niveau d'hétérozygotie.

Pour cela, des tests de Kolmogorov-Smirnov pour la comparaison de deux échantillons, adaptés à des échantillons de tailles limitées, ont été utilisés (Tomassone 1986). Un exemple de ce test, concernant la comparaison entre arbres polycycliques et arbres monocycliques pour le nombre total de locus hétérozygotes est donné tableau 7 (voir la répartition du tableau 6). L'hypothèse nulle suppose que les distributions des deux échantillons (arbres polycycliques et monocycliques) sont identiques. Avec des échantillons de 9 individus, on rejeterait l'hypothèse nulle à 5% si la valeur absolue de la plus grande différence entre les deux répartitions était supérieure à 6, d'après les tables. Dans le cas présenté ici, l'hypothèse nulle ne peut donc être rejetée. La même conclusion s'impose à la suite des tests réalisés pour chaque catégorie de locus définie plus haut : l'hétérozygotie et le polycyclisme apparaissent indépendants.

Le même type de comparaison a été réalisé pour les trois classes de vigueur deux à deux, et là encore, on peut conclure à la ressemblance des arbres plus ou moins vigoureux au regard de leur hétérozygotie, quel que soit le sous ensemble de locus considéré.

### 2.2. Relation avec les caractères quantitatifs

Les caractères quantitatifs décrivant les 18 arbres ont été confrontés aux mesures d'hétérozygotie par le calcul de corrélations de Kendall. Ce coefficient de corrélation mesure la liaison entre les deux variables comparées, indépendamment de l'échelle de ces variables, mais en tenant compte des positions relatives des différents points (Tomassone 1986).

Une association intéressante a pu être observée. Il s'agit d'une liaison apparaissant entre le nombre total de locus hétérozygotes chez l'arbre mère et la moyenne familiale (demi-frères) des hauteurs du premier cycle de croissance réalisé à la suite d'un stress hydrique. Ces deux caractères sont corrélés négativement (-0.52) avec un niveau de signification inférieur à 1% (figure 14).

Tableau 7 : Comparaison du nombre total de locus hétérozygotes avec les classes de polycyclisme.

H	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
$n_+$	0	0	0	0	1	2	2	3	3	5	7	8	9	9	9
$n_-$	1	1	1	1	2	4	4	5	5	6	7	8	8	8	9

$\sup |n_- - n_+| = 2$  (non significatif)

H : nombre total de locus hétérozygotes

$n_+$  : Cumul du nombre d'arbres polycycliques observés dans chaque catégorie

$n_-$  : Cumul du nombre d'arbres monocycliques observés dans chaque catégorie



### 2.3. Effet de l'hétérozygotie chez les conifères

La croissance est un facteur important dans la valeur sélective des arbres. La dépression de consanguinité est très marquée chez ces espèces, donc si l'hétérozygotie mesurée sur un échantillon de locus estime l'hétérozygotie du génome en général (ce qui est discuté, voir Chakraborty 1987), les individus les plus hétérozygotes sont supposés moins affectés par la consanguinité, donc supérieurs. Ceci constitue l'hypothèse de dominance pour expliquer l'avantage des hétérozygotes, l'hypothèse alternative généralement suggérée étant celle de la superdominance (Smouse 1986). Lors d'autofécondations chez les conifères l'effet de l'hétérozygotie apparaît clairement, mais l'avantage aux hétérozygotes est beaucoup moins net en croisement (Strauss 1986). Une corrélation négative entre croissance et hétérozygotie s'explique mal avec les hypothèses précédentes. Pourtant, Strauss et Libby (1987) observent une corrélation négative entre des caractères liés à la croissance et les niveaux d'hétérozygotie les plus élevés chez *Pinus radiata* pour 27 locus isoenzymatiques, ce qui rejoint notre observation. De fait, l'étude des relations entre l'hétérozygotie et le phénotype conduit assez souvent à des résultats contradictoires (Chakraborty 1987) et ceci est vrai pour les conifères (Ledig et al. 1983, Bush et al. 1987, Bush et Smouse 1991). Le genre *Mytilus*, qui fait référence, fournit des résultats plus constants, l'hétérozygotie pour des locus isoenzymatiques a un effet significatif, entre autres sur le taux de croissance, mais elle concerne un sous-ensemble d'enzymes impliquées dans le métabolisme énergétique, qui auraient un effet direct sur le caractère. Cet effet est apparu encore amplifié dans des conditions de stress (Koehn 1991). Chez un ver de terre, Diehl (1988), n'observe pas de corrélations entre le taux de croissance et l'hétérozygotie à 13 locus impliqués dans le métabolisme des sucres, mais en condition de stress, il détecte des corrélations, notamment négatives pour un sous-ensemble de 6 locus. Des variations génétiques proches de la neutralité dans un environnement normal pourraient avoir un potentiel sélectif exprimé en conditions extrêmes (Hartl 1989).

Chez le Pin maritime, le stress hydrique lors des sécheresses estivales dans les Landes est un facteur important pour la croissance (Guyon et Kremer 1982), la réponse de jeunes plants à ce stress apparaît, dans notre expérience, influencée négativement par l'hétérozygotie du parent commun. Notons que les familles repérées comme les plus vigoureuses (voir tableau 6) à l'âge adulte sont aux deux extrêmes du graphe pour les hauteurs (familles 1, 2, 3, 4, 5 et 6). Ce caractère, comme l'hétérozygotie, n'est donc pas lié aux performances adultes, et paraît donc difficile à relier à la valeur sélective des individus, surtout à partir d'un échantillon de petite taille. La corrélation observée pourrait être testée à partir d'une mesure obtenue sur un plus grand nombre de descendants par arbres, et pour l'ensemble des 18 arbres, mais serait mieux estimée sur un plus grand nombre de familles. Si la corrélation était confirmée, le mécanisme en jeu resterait à établir.

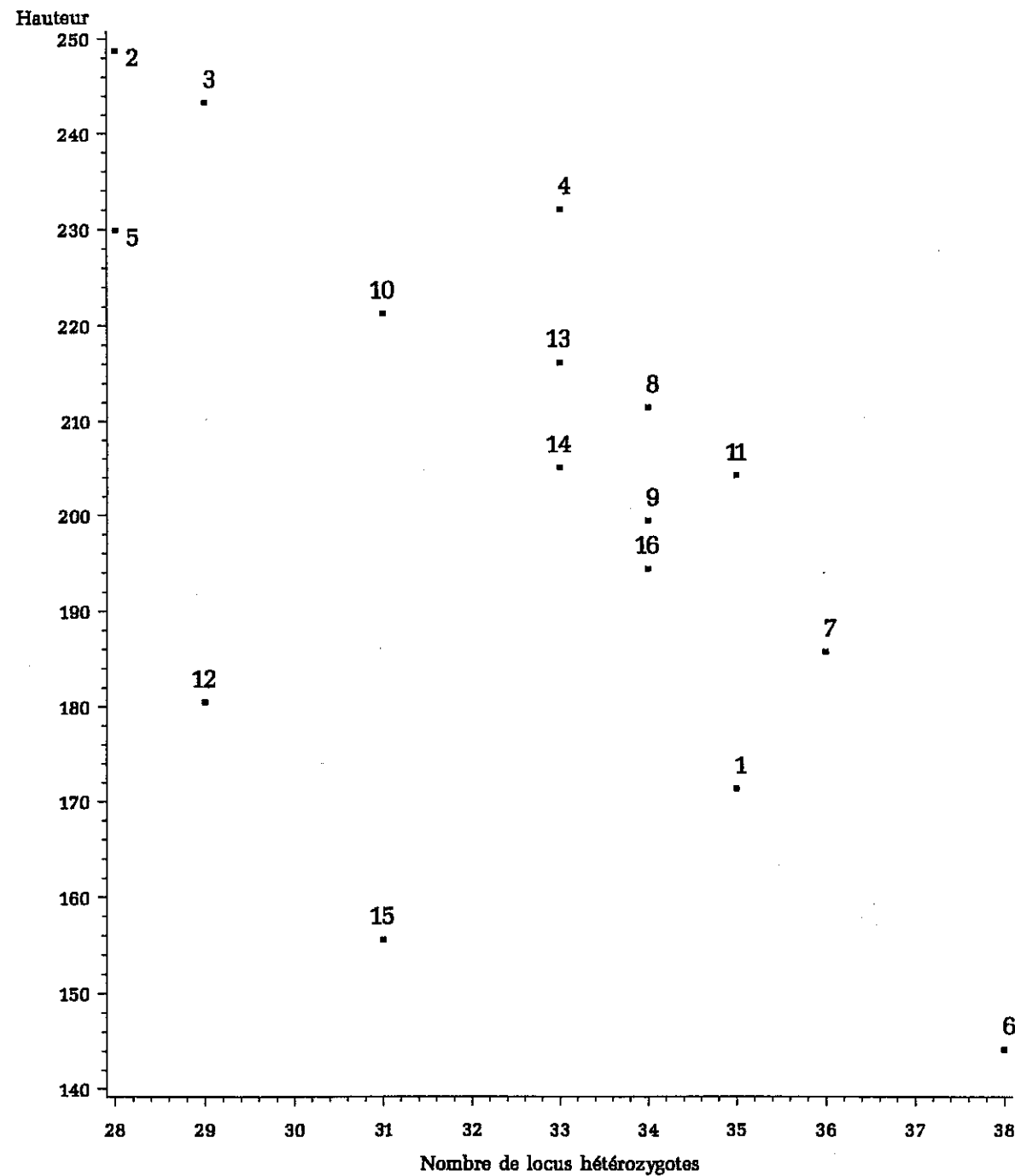


Figure 14 : Relation entre la hauteur mesurée après un stress hydrique et le nombre de locus hétérozygotes de l'arbre mère pour 16 des 18 familles.  
(Coefficient de corrélation de Kendall = -0.52, P = 0.006)

L'effet de l'hétérozygotie dans notre échantillon est réduit à une seule relation significative avec un caractère de croissance particulier. Sur un échantillon de taille réduite, ceci n'est pas surprenant, d'autant que les résultats obtenus sur de plus grands effectifs, mais avec un plus faible nombre de locus conduisent généralement à des conclusions contradictoires.

### 3. Etude simultanée de l'ensemble des marqueurs

Pour prendre en compte de façon simultanée le génotype de chacun des arbres de l'échantillon aux 84 locus détectés, une analyse multivariable était possible. Notre but était d'étudier la répartition des arbres et de rechercher l'éventuel regroupement d'arbres de classes de performances semblables. Pour cela des analyses factorielles des correspondances ont été réalisées. Ces analyses s'appuient sur des tables de contingence dont les lignes correspondent aux arbres de l'échantillon et dont les colonnes représentent les locus, chaque locus ayant autant de modalités que de génotypes différents. Les locus ont été distingués selon les catégories définies au chapitre 1.1. La procédure CORRESP du logiciel SAS (SAS Institute 1990) a été utilisée. L'analyse incluant l'ensemble des locus présentait des particularités. Du fait du grand nombre de génotypes par locus, 260 au total, l'analyse des correspondances se prêtait mal à l'interprétation dans ce cas. Néanmoins, compte tenu des résultats obtenus, il nous a paru intéressant de conserver cette analyse.

#### 3.1. Discrimination

Le premier plan de l'analyse factorielle des correspondances incluant le génotype des 18 arbres aux 84 locus est représenté figure 15. L'arbre 10 se distingue de tous les autres. La droite horizontale figurée sur le schéma définit deux zones. La zone basse comprend 7 arbres monocycliques, et deux arbres polycycliques. La zone haute comprend, symétriquement, 7 arbres polycycliques et 2 arbres monocycliques. Cette différence de répartition est-elle significative? Un test du Fisher exact nous permet de le tester (Siegel 1956) :

		Zone	
		basse	haute
Arbres	polycycliques	2	7
	monocycliques	7	2

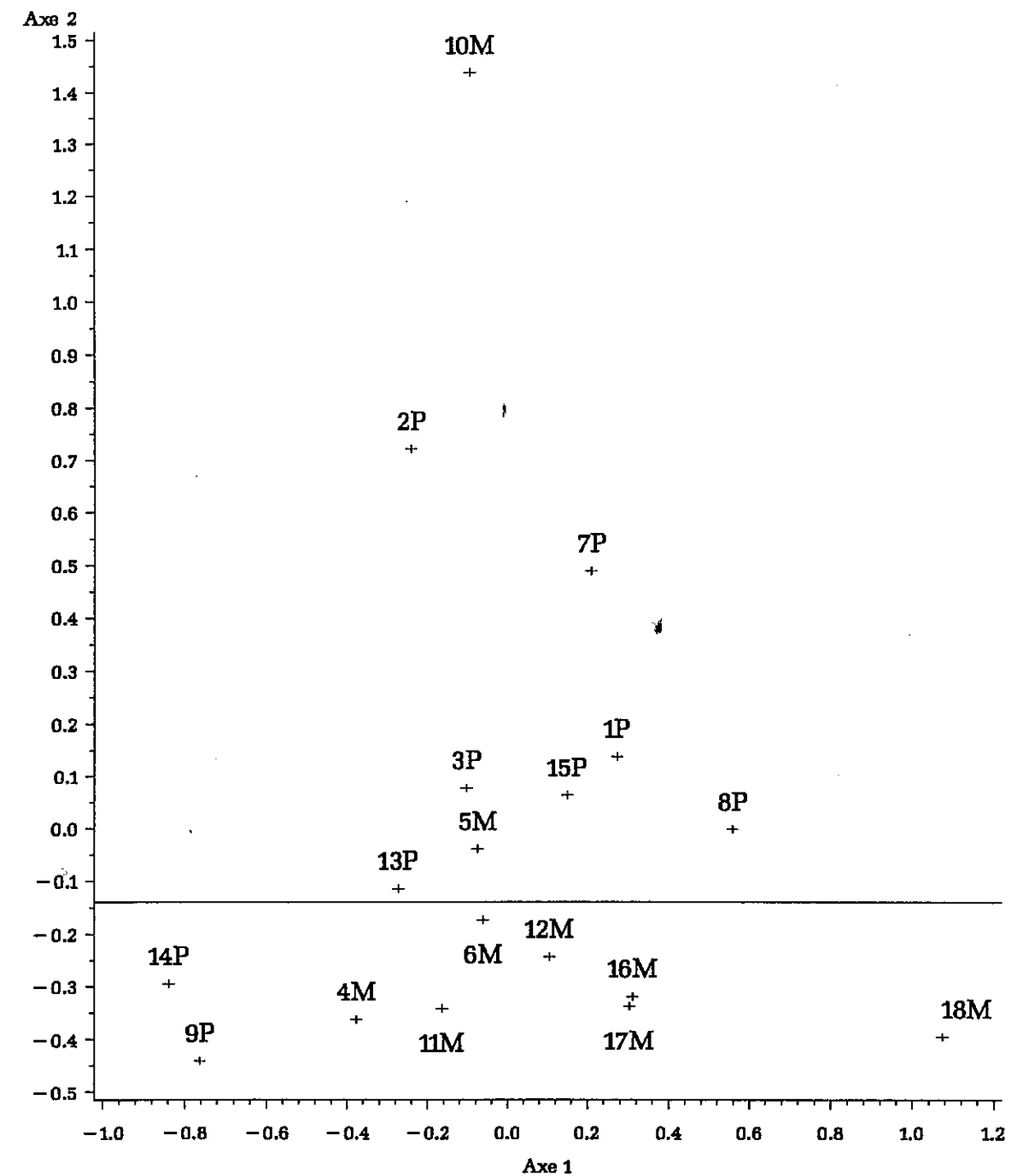


Figure 15 : Premier plan de l'AFC du génotype des 18 arbres aux 84 locus.

nM : arbres monocycliques  
 nP : arbres polycycliques  
 n : numéro de l'arbre

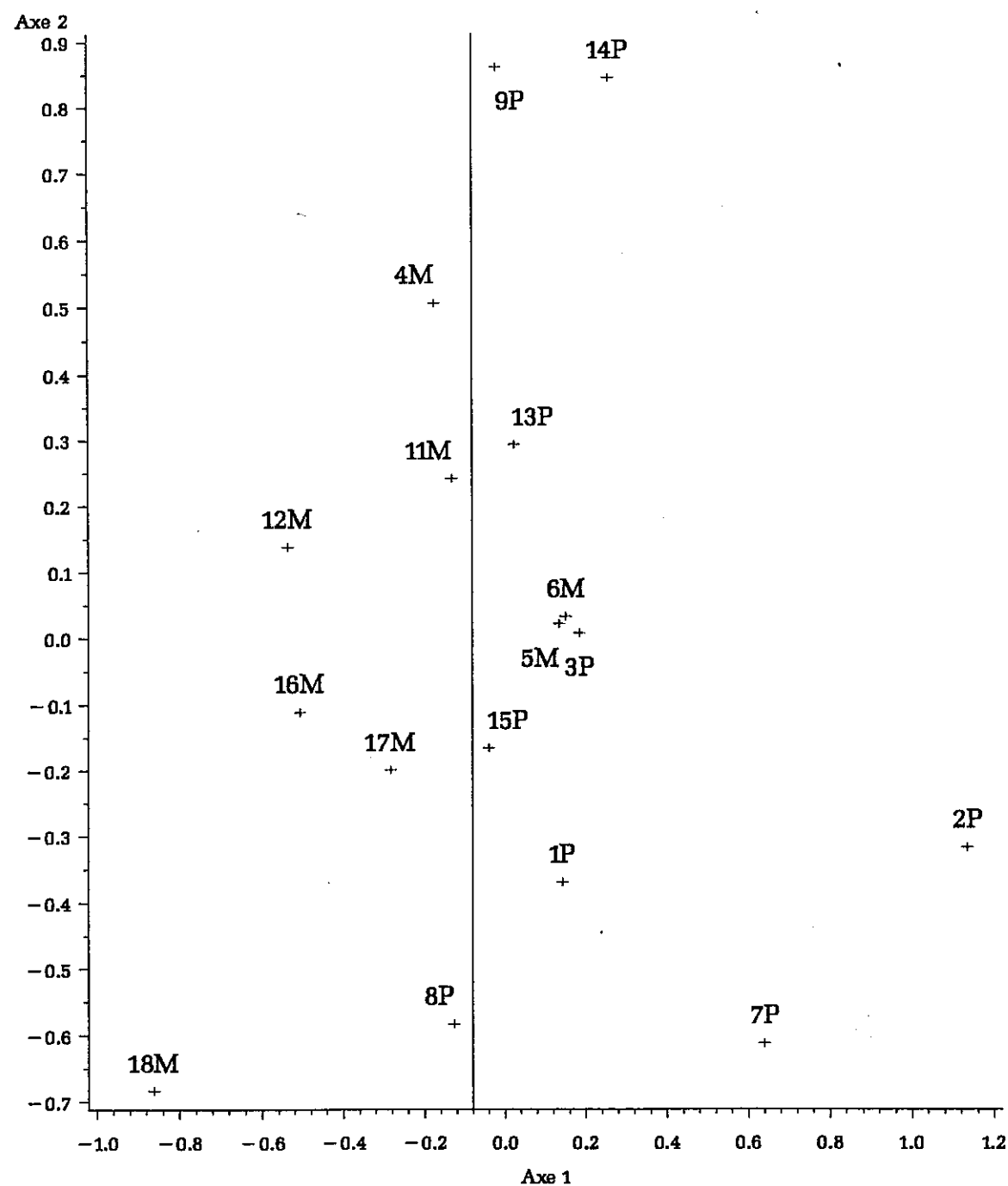


Figure 16 : Premier plan de l'AFC sans l'arbre 10.

nM : arbres monocycliques  
 nP : arbres polycycliques  
 n : numéro de l'arbre

L'hypothèse nulle de l'indépendance entre l'appartenance à une zone et la classe de performance peut être rejetée avec un niveau de signification de 5% selon ce test. Notons que les arbres 9 et 14 s'individualisent des monocycliques dans la zone basse : une ligne courbe permettrait une séparation des deux groupes encore meilleure. La discrimination sur l'axe 1 s'explique essentiellement par la présence chez l'arbre 10 de génotypes "rares" à plusieurs locus. Nous avons donc réalisé une seconde analyse en considérant cet arbre comme un point supplémentaire (Volle 1985). Une ligne verticale sépare les arbres monocycliques des arbres polycycliques, les arbres 5, 6 et 8 restant mal classés (figure 16).

### 3.2. Arbres polycycliques et monocycliques

Dans les deux analyses, la discrimination observée le long des axes est due principalement aux arbres ayant des positions extrêmes : les arbres 10, 7 et 2 dans le premier cas, 2, 7 et 18 dans le deuxième. La contribution des génotypes/locus à ces axes est très dispersée. Une somme des contributions partielles égale à 0.8 permet généralement de mettre en évidence les points ayant contribué le plus au positionnement de l'axe (Volle 1985). Dans les deux analyses, cette contribution est partagée par environ 80 génotypes/locus. On retrouve parmi eux les génotypes "rares" uniquement représentés chez les arbres aux positions extrêmes.

Au vu de ces résultats la différence entre arbres polycycliques et monocycliques semble donc liée à des différences de génotypes répartis entre de nombreux locus. Cette discrimination n'a pas été retrouvée en étudiant les différentes catégories de locus séparément.

Les arbres polycycliques et monocycliques se distinguent sur plusieurs points. La chronologie ainsi que la durée de fonctionnement de leurs méristèmes ne sont pas identiques (Kremer et Roussel 1982). En conditions de croissance accélérée, les arbres polycycliques mettent en place des structures à caractère "adulte" plus tard que les arbres monocycliques : ces arbres conserveraient au cours de leur vie une certaine forme de juvénilité, en effet, les plantules expriment toutes du polycyclisme durant les premières années de leur croissance (Lascoux 1992). Il apparaît en fait que les arbres polycycliques sont avantagés pour la croissance par rapport aux arbres monocycliques (Kremer 1981a). De plus ils produisent un nombre plus élevé de graines, mais plus légères que celles des arbres monocycliques. Ces derniers présentent un développement racinaire aux premiers stades de la croissance qui leur permet d'être compétitifs (Guignard 1983). Les deux types d'arbres semblent donc avoir adopté des stratégies très différentes, qui se traduisent vraisemblablement au niveau génétique.

Si ces différences ont des bases génétiques certaines, la distinction observée sur la base du génotype des arbres pour l'ensemble des locus n'en est pas moins surprenante. En effet lorsque la relation entre le phénotype et le génotype à de nombreux marqueurs est étudiée, la liaison qui peut exister entre un sous-ensemble de locus et le caractère a peu de chances d'échapper au "bruit de fond" créé par les locus sans intérêt (Lewontin 1984). En cela, les études multilocus sont délicates. Nous nous sommes donc intéressés à une approche locus par locus.

### C. ETUDE DE L'EFFET INDIVIDUEL DES LOCUS

La recherche de locus à effet quantitatif (quantitative trait loci ou QTL) repose sur l'étude conjointe, pour les descendants d'un croisement type F2 ou backcross, de la ségrégation pour des marqueurs et de l'expression de caractères quantitatifs chez les mêmes individus. Dans notre cas, la ségrégation pour les marqueurs a été observée sur les mégagamétophytes, et les caractères disponibles se rapportent aux arbres. Nous connaissons pourtant le poids de chaque mégagamétophyte. Pour ce caractère, relatif au niveau haploïde, la recherche de QTL est possible, bien que nous ne soyons pas dans le cadre d'un croisement classique puisque les 18 arbres n'ont pas de lien d'apparentement. Pour ce qui concerne les caractères liés à la croissance nous ne pouvons étudier qu'un effet du génotype des arbres, au niveau diploïde cette fois, sans envisager les ségrégations.

#### 1. Etude de l'effet des locus sur le poids des mégagamétophytes

Chaque mégagamétophyte utilisé pour réaliser un gel avait été pesé au préalable. L'étude d'une relation entre le poids d'un mégagamétophyte et l'allèle exprimé dans cet organe haploïde était donc possible pour chaque locus.

##### 1.1. Méthode

La première étape consiste à analyser ensemble les endospermes d'un même arbre  $k$ , hétérozygote au locus considéré, en réalisant une analyse de variance à un facteur ("allèle"), à deux niveaux. Le poids du mégagamétophyte  $i$  portant l'allèle  $j$  s'écrit :

$$P_i = \mu + I_{j(i)} + e_i$$

$\mu$  est la moyenne du caractère

$I_{j(i)}$  représente l'effet du  $j^{\text{ème}}$  allèle au locus étudié,  $j$  étant fonction de  $i$

$e_i$  est la résiduelle.

Ces analyses ont été réalisées avec la procédure GLM (sommes des carrés de type III pour tenir compte des effectifs déséquilibrés) du logiciel SAS (SAS Institute 1990). L'homogénéité des variances résiduelles correspondant aux différents arbres hétérozygotes a alors été testée par un test de Bartlett. Dans le cas où l'hypothèse nulle n'était pas rejetée, une analyse de variance incluant les  $H$  arbres hétérozygotes aux locus considérés pouvait être réalisée. Si  $SA_k$  et  $SE_k$  représentent respectivement les sommes des carrés du facteur "allèle" (à un degré de liberté) et de la résiduelle (à  $(n_k-2)$  degrés de liberté, où  $n_k$  est le nombre de mégagamétophytes de l'arbre  $k$  étudiés) issues de la première analyse de variance, on a :

$$SA = \sum_{k=1}^H SA_k \quad \text{et} \quad SE = \sum_{k=1}^H SE_k$$

Ces sommes de carrés possèdent  $H$  degrés de liberté pour la première et  $N = \sum_k (n_k-2)$  pour la seconde. La statistique  $F$  :

$$F = \frac{SA/H}{SE/N}$$

suit, si le locus étudié et le poids des endospermes sont indépendants, une distribution de  $F(H,N)$ . Les liaisons entre ces deux caractères sont ainsi recherchées. Le test de Bartlett ainsi que les analyses de variances finales ont été programmés avec le logiciel Splus (Becker et al. 1988).

Cette procédure d'analyse permet d'éliminer l'effet "arbre", qui est effectivement très important, et de tenir compte du fait que la présence du même allèle chez deux arbres différents ne traduit pas la même situation génétique, les arbres étant sans apparentement.

##### 1.2. Résultats

Lorsqu'un grand nombre de tests est réalisé, il est probable de trouver un certain nombre de tests significatifs au niveau  $\alpha$  du seul fait du hasard, qui correspondent à des erreurs de première espèce. Si  $T$  est le nombre de tests réalisés, cet événement suit une loi binomiale de paramètres  $T$  et  $\alpha$ . La probabilité  $P(K)$  de trouver au moins  $K$  tests significatifs s'écrit donc :

$$P(K) = 1 - \sum_{i=0}^{K-1} C_T^i \alpha^i (1-\alpha)^{T-i}$$

Nous avons réalisé autant de tests que de locus disponibles, soit 84. Trois locus ont été trouvés avec un effet du facteur allèle significatif à 1% (tableau 8). Les tests de Bartlett correspondants apparaissent

non significatifs. L'effet de chaque locus a été testé sur un minimum de 7 arbres hétérozygotes, donc sur plus de 70 mégagamétophytes.

La probabilité de trouver au moins 3 effets du facteur allèle significatifs à 1% du seul fait du hasard est égale à 5% dans notre cas de figure. Cette probabilité est faible, mais ne nous dit pas si, parmi les tests, l'un ou l'autre correspond à une erreur de première espèce. Les locus concernés, n°4209, 3204 et 2323 sont situés sur trois groupes de liaison distincts et sont responsables de variations de position de protéines (voir figures 5 et 6).

Tableau 8 : Effet de 3 locus sur le poids des mégagamétophytes.

Locus	Groupe de liaison	Nombre d'arbres hétérozygotes	Niveaux des tests	
			F	Bartlett
4209	14	15	0.016	0.11
3204	1	7	0.010	0.08
2323	9	10	0.008	0.39

### 1.3. Poids de l'endosperme et marqueur

Lorsqu'une descendance, par exemple une F2, est étudiée, deux allèles identiques observés chez deux descendants à un locus marqueur donné ont la même origine. Si un QTL est proche de ce locus et est homozygote pour deux allèles différents chez les deux parents, le fait pour un descendant d'avoir hérité l'allèle maternel (paternel) au locus marqueur aura des chances d'être accompagné de la présence de l'allèle maternel (paternel) au QTL. On détectera alors une différence pour le caractère quantitatif entre les descendants qui diffèrent pour le génotype au marqueur. Ce raisonnement est à la base du principe de la recherche de QTL. La détection de situations de ce type sera optimisée du fait du grand déséquilibre de liaison qui existe dans une descendance F2.

Les 18 arbres qui composent notre échantillon proviennent de divers endroits de la forêt des Landes et n'ont aucun apparemment connu. Il n'est pas possible de raisonner comme si l'on travaillait avec une descendance F2 : deux allèles identiques observés chez les mégagamétophytes de deux arbres différents à un locus marqueur n'ont pas été hérités selon le même "schéma". De plus, parmi les locus étudiés certains ont plus de deux allèles (3 allèles au locus n°3207), et sont de plus les plus polymorphes de tous les locus (5 allèles au locus n°4209 et 6 au locus n°2323). Les allèles présents à un QTL proche n'ont donc pas de raison d'être semblables d'un arbre à l'autre. Pourtant, nous avons observé des effets très significatifs et seule l'existence d'une liaison forte ou totale entre les marqueurs et les QTL, pour chaque arbre, peut permettre de rendre compte de ces effets.

Chez le Pin maritime, l'endosperme représente 37% du poids de la graine, l'embryon 7%, et le tégument 56%. Les effets maternels, évalués dans un croisement diallèle, sont apparus très importants sur le poids de ces trois composantes de la graine (Guignard 1983). Les effets paternels n'étaient significatifs que sur le poids de l'embryon, d'ailleurs seule partie de la graine des gymnospermes contenant le génotype apporté par le pollen. Deux types d'endospermes de pigmentation différente parmi les graines d'un même pin maritime d'origine italienne possédaient également des poids des composantes de la graine différents (Petit 1988).

Les plants les plus vigoureux ne sont pas forcément associés aux plus grosses graines (Perry 1976). Comme nous l'avons signalé en introduction, l'effet maternel mesuré par le poids de la graine n'est pas systématique chez les conifères et s'il existe sur les jeunes plants, il disparaît rapidement. D'un point de vue évolutif, l'augmentation de la taille des graines représente un coût métabolique pour la mère, alors que le père profite de l'augmentation de cette taille sans en subir un coût direct : les intérêts divergent. La sélection des gènes déterminant la taille de la graine doit donc établir un équilibre entre une taille qui confère un avantage individuel aux plantules et le coût reproductif lié à une production d'un nombre plus faible de graines (Haig et Westoby 1991). L'absence d'une sélection forte sur la taille optimale des graines expliquerait alors la variabilité observée (Temme 1986). Cette variabilité est effective chez les conifères (Sorensen et Campbell 1985). La valeur adaptative du poids de la graine n'est donc pas simple à définir. L'hypothèse d'une sélection sur ce caractère dans la population de Pin maritime qui maintiendrait un déséquilibre de liaison stable entre les QTL et des locus marqueurs génétiquement proches, par un effet de "hitchhiking" (Thomson 1977 cité par Muona 1982), semble donc peu vraisemblable. Une hypothèse de cette ordre n'est pourtant pas à exclure et est intéressante : la sélection se traduirait au niveau haploïde.

L'existence d'une liaison totale entre les marqueurs et les QTL est une hypothèse forte mais constitue la plus simple interprétation de nos observations. Si un locus pléiotrope avait une action sur les protéines correspondant aux marqueurs mais également sur un processus intervenant dans le poids de la graine, l'effet observé serait expliqué. Encore plus simplement, les protéines en cause pourraient être impliquées dans les processus d'élaboration du poids de la graine, elles en sont d'ailleurs un constituant. A cet égard, il serait intéressant de connaître la nature de ces protéines. Deux des trois locus se distinguent notamment par leur grand polymorphisme (locus n°4209 et 2323, à quatre et six allèles respectivement). A l'image de ce qui a été réalisé chez le soja (Diers et al. 1992), des QTL pour la teneur en protéine ou en lipides, ou pour tout autre caractère concernant la graine pourraient être facilement mis en évidence chez les gymnospermes.

## 2. Relations entre le génotype des arbres et leurs performances

### 2.1. Liaison entre génotype et classes de performances adultes

Nous nous sommes intéressés dans un premier temps aux classes de performances adultes, en testant l'indépendance entre l'appartenance à une classe (vigueur, polycyclisme) et le génotype des arbres locus par locus.

Tableau 9 : Répartition des arbres selon leur génotype et leur classe de polycyclisme ou de vigueur.

		Génotypes		
		AA	aA	aa
Polycyclisme	+	n <sub>1</sub>	n <sub>2</sub>	n <sub>3</sub>
	-	N <sub>1</sub> -n <sub>1</sub>	N <sub>2</sub> -n <sub>2</sub>	N <sub>3</sub> -n <sub>3</sub>
	Total	N <sub>1</sub>	N <sub>2</sub>	N <sub>3</sub>
		AA	aA	aa
Vigueur	+	n <sub>11</sub>	n <sub>12</sub>	n <sub>13</sub>
	0	n <sub>21</sub>	n <sub>22</sub>	n <sub>23</sub>
	-	n <sub>31</sub>	n <sub>32</sub>	n <sub>33</sub>
	Total	N <sub>1</sub>	N <sub>2</sub>	N <sub>3</sub>

### 2.1.1. Méthode

Prenons l'exemple d'un locus biallélique (A,a) (le raisonnement est le même lorsque plus de deux allèles sont présents au locus). Connaissant le génotype à ce locus et la classe d'appartenance (polycyclisme ou vigueur) de chacun des 18 arbres, des tables de contingence classe/génotype sont construites (tableau 9).

Pour le polycyclisme, en se plaçant sous l'hypothèse de l'indépendance des deux facteurs, génotype et classe, chaque génotype  $i$  a autant de chances de correspondre à un arbre polycyclique qu'à un arbre monocyclique. Le nombre  $n_i$  de génotypes dans une classe (ou symétriquement le nombre  $N_i - n_i$  dans l'autre classe) suit donc une loi binomiale de paramètres  $N_i$  et  $1/2$ . Si  $g$  génotypes différents sont observés au locus étudié, la statistique  $S$  :

$$S = \sum_{i=1}^g \frac{(2n_i - N_i)^2}{N_i}$$

suit alors une loi du  $\chi^2$  à  $g$  degrés de liberté.

Pour les classes de vigueur, un test du  $\chi^2$  classique, à  $2(g-1)$  degrés de liberté, qui compare la distribution observée à une distribution théorique uniforme, peut être calculé :

$$S = \sum_{i=1}^3 \sum_{j=1}^g \frac{(n_{ij} - N_j/3)^2}{N_j/3}$$

Ces tests permettent un premier tri parmi l'ensemble des locus. Les locus correspondant aux tests les plus significatifs sont retenus. Compte tenu des faibles effectifs étudiés, la statistique utilisée ne converge pas de façon satisfaisante vers une loi du  $\chi^2$ . Pour connaître le niveau de signification de la liaison détectée, on procède donc à un test de permutations. Celui-ci permet d'obtenir une représentation de la "vraie" loi, inconnue, suivie par la statistique, sous l'hypothèse d'indépendance entre le génotype et la classe et conditionnellement aux marginales. Le test consiste, en fixant le nombre d'individus observés par colonne (génotype) à redistribuer au hasard dans les cases les individus correspondant aux différentes classes. La statistique  $S$  du test est recalculée à chaque permutation, et ceci est réalisé un grand nombre de fois. Le nombre de permutations ayant donné lieu à une valeur de  $S$  supérieure ou égale à la valeur observée est comptabilisé. Le niveau de signification correspondant à l'échantillon est alors égal au rapport de ce nombre sur le nombre total de permutations. L'ensemble de ces calculs a été réalisé avec le logiciel Splus (Becker et al. 1988).

Tableau 10 : Liaison entre le génotype des arbres à trois locus et les classes de polycyclisme.

Table de contingence		Test initial	Test de permutations					
			nombre	niveau $\alpha$				
Locus n°1301								
	v <sub>1</sub> q <sub>2</sub>	q <sub>1</sub> q <sub>2</sub>	q <sub>2</sub> q <sub>2</sub>					
+	4	4	1	S = 8.1 (4%)				
-	1	1	7					
			1100	2.2%				
			2100	2.5%				
Locus n°4404								
	aa	aq <sub>1</sub>	aq <sub>2</sub>	q <sub>1</sub> q <sub>1</sub>	q <sub>1</sub> q <sub>2</sub>	q <sub>2</sub> q <sub>2</sub>		
+	1	4	0	1	1	2	S = 10.5 (6%)	
-	0	0	3	2	4	0		
							1100	3.6%
							2100	3.7%
Locus n°2320								
	q <sub>1</sub> q <sub>1</sub>	q <sub>1</sub> q <sub>2</sub>	q <sub>2</sub> q <sub>2</sub>					
+	1	2	4	S = 7 (7%)				
-	5	1	0					
				1100	5.0%			
				2100	4.8%			

Tableau 11 : Liaison entre le génotype des arbres à quatre locus et les classes de vigueur.

Table de contingence		Test initial	Test de permutations		
			nombre	niveau $\alpha$	
Locus n°3206					
	q <sub>1</sub> q <sub>2</sub>	q <sub>2</sub> q <sub>2</sub>			
+	5	1			
0	3	3	S = 8.5 (1%)	1100 2.0%	
-	0	6		2100 1.7%	
Locus n°4305					
	aa	pa	pp		
+	0	5	1		
0	3	0	3	S = 10.1 (4%)	
-	1	2	3	1000 4.3%	
				2000 3.8%	
Locus n°1401					
	v <sub>1</sub> v <sub>1</sub>	v <sub>1</sub> v <sub>2</sub>			
+	4	2			
0	3	3	S = 6.1 (5%)	1000 4.3%	
-	0	6		2000 3.9%	
Locus n°3409					
	v <sub>1</sub> v <sub>1</sub>	v <sub>1</sub> v <sub>2</sub>	v <sub>1</sub> v <sub>3</sub>	v <sub>2</sub> v <sub>3</sub>	v <sub>3</sub> v <sub>3</sub>
+	0	2	3	0	1
0	0	0	1	1	4
-	1	0	4	1	0
			S = 13.9 (8%)	1000 5.1%	
				2000 4.6%	

### 2.1.2. Résultats

En choisissant un niveau de test initial peu sévère (10%), trois locus ont été conservés pour le polycyclisme (tableau 10) et quatre pour la vigueur (tableau 11). Ces 7 locus ont un effet significatif à 5% d'après les tests de permutations. Compte tenu du nombre de tests réalisés (168 au total), il est très vraisemblable d'avoir observé ces liaisons du seul fait du hasard. Notons que les génotypes aux locus n°1301 et 4404 se retrouvent parmi les génotype/locus aux contributions les plus élevées dans la discrimination observée entre polycycliques et monocycliques de l'analyse factorielle des correspondances précédemment décrite. Remarquons également que les locus n°1401, et 2320, en relation avec la vigueur et le polycyclisme respectivement, sont liés génétiquement (groupe de liaison n°4, figure 6).

### 2.2 Effet du génotype sur les caractères quantitatifs

Les arbres de l'échantillon ont été évalués pour différents caractères liés à la croissance. Leur valeurs propres en parc à clones et leurs aptitudes générales à la combinaison estimées à partir de familles de demi-frères sont disponibles. L'effet individuel des marqueurs sur chacun de ces caractères a été testé.

#### 2.2.1. Méthode

Des analyses de variance à un facteur ont été réalisées pour chaque caractère. La valeur du j<sup>ème</sup> individu présentant le génotype i à un locus donné se décompose de la façon suivante :

$$Y_{ij} = \mu + g_i + e_{ij}$$

$\mu$  est la moyenne du caractère

$g_i$  représente l'effet du i<sup>ème</sup> génotype au locus étudié

$e_{ij}$  est la résiduelle.

La procédure GLM (sommes des carrés de type III) du logiciel SAS (SAS Institute 1990) a été utilisée pour ces analyses. Un test F permet de déclarer si l'effet du locus est significatif sur le caractère. Les moyennes correspondant aux différents génotypes du locus sont alors comparées avec un test de Tukey, préconisé par le logiciel SAS.



## 2.2.2. Résultats

### 2.2.2.1. Choix d'un niveau de signification

Nous disposions de 49 caractères relatifs à tout ou partie des 18 arbres. Pour chaque caractère, l'effet de chacun des 84 locus a été testé. Donc  $49 \times 84 = 4116$  tests ont été réalisés. Les tests dont le niveau de signification était inférieur à 0.4%, ont été retenus. Nous avons calculé, à l'aide d'un programme réalisé avec Splus (Becker et al. 1988), la probabilité (voir C.1.2.) d'avoir observé le nombre (ou un nombre supérieur) de tests trouvés significatifs aux différents niveaux inférieurs à 0.4%, du seul fait du hasard. Les résultats suivants ont été obtenus :

Niveau	Nombre de tests significatifs	Probabilité
0.4%	29	0.3%
0.3%	24	0.2%
0.2%	17	0.5%
0.1%	8	5.8%

Sous l'hypothèse d'absence d'effet des locus, les probabilités d'avoir observé de tels nombres de tests significatifs sont faibles, et particulièrement pour des niveaux supérieurs à 0.1%. Pourtant, parmi ces tests, il est clair qu'une part non négligeable correspond à des erreurs de type I, d'autant plus que le niveau de signification retenu est élevé. Nous avons choisi de conserver les tests significatifs à 0.4% et avons cherché à repérer parmi les 29 cas lesquels correspondaient aux situations les plus remarquables, du point de vue du classement des arbres selon leur performances et selon leur génotype. Avant de réaliser ce tri, nous avons éliminé les cas extrêmes où seul un représentant d'un génotype était trouvé significativement différent des autres génotypes du locus. Les 22 tests qui ont été conservés sont présentés en annexe 3. Le tableau 12 propose un récapitulatif de ces tests. Treize caractères et 17 locus sont concernés.

### 2.2.2.2. Répartition des types de locus

Nous avons voulu voir si la répartition des 17 locus selon leur action sur les protéines (variation de position, de quantité ou présence/absence) était comparable à la répartition globale observée pour les 84 locus disponibles. Nous avons considéré les locus associant différents allèles, dont quantitatifs, comme faisant partie de la classe des variations de quantité. La répartition observée dans l'échantillon des 84 locus nous permet de calculer la répartition attendue dans le sous-échantillon, sous l'hypothèse que ce sous-échantillon provient d'un tirage aléatoire dans le grand échantillon :

Tableau 12 : Récapitulatif des liaisons trouvées entre caractères et génotype aux locus.

Origine des mesures	Caractère* (nombre de familles mesurées)	Locus	Classes génotypiques différentes		
			Groupe de liaison		
Test de descendance	H83 (15)	4406	7	6	$q_2q_2+q_2a < q_1q_2+q_1q_1$
	H87 (15)	4406	7	6	$q_2q_2+q_2a < q_1q_2+q_1q_1$
		4305		15	aa+pp < pa
		4202		15	$q_2q_2+q_1q_1 < q_1q_2$
	ECV83 (15)	3206		3	$q_1q_2 < q_2q_2$
Parc à clones	SPII (16)	3111		10	pa < pp
Stress hydrique	HST (16)	2428		nc**	$v_1v_2 < v_2v_2$
Phytotron :					
Première période de croissance	NB (9)	2322	7	* 12	$q_2q_2 < q_1q_2$
		3227	7	nc	aa < pa
	PB (9)	2421	7	nc	$q_2q_2 < q_1q_2$
		3418	7	6	pa < pp+aa
Deuxième période de croissance	NB1 (11)	1215	7	nc	$q_2q_2+q_1q_1 < q_1q_2$
	NAII2 (11)	4101	7	nc	$v_1v_2 < v_1v_1$
	MSUL3 (10)	1410	7	nc	$q_1q_1 < q_1q_2$
		2210	7	11	$q_1q_1 < q_1q_2$
		4211	7	5	pp < pa
	PAIII (11)	2322	7	12	$q_2q_2 < q_1q_2$
		3227	7	nc	aa < pa
	2421	7	nc	$q_2q_2 < q_1q_2$	
	3418	7	6	pa+pp < aa	
	PAI2 (11)	3309		17	pp < pa
	PB3 (11)	2215		nc	pp < aa < pa

\* voir l'annexe 3 pour la signification des abréviations

\*\* locus non cartographié

Nombre de locus selon le type de variations :

	Position	Quantité	Présence/absence
84 locus	30	22	32
17 locus	2	8	7
Attendu	6.07	4.45	6.48

Un test G (à 2 degrés de liberté, Sokal et Rohlf 1981), nous permet de comparer les deux répartitions. Ce test correspond à un niveau de signification de 5%. Dans l'échantillon des 17 locus il semble donc y avoir plus de locus responsables de variations de quantité et moins de locus responsables de variations de position que ce qui aurait été attendu si les 17 locus représentaient un tirage aléatoire parmi les 84 locus.

### 2.2.2.3. Tri des tests

Pour réaliser un tri parmi les 22 associations retenues entre locus et caractères quantitatifs, des tests de Fischer exact pouvaient être réalisés.

Quinze arbres de notre échantillon sont représentés dans un test de descendance<sup>1</sup> et leurs aptitudes générales à la combinaison pour la hauteur totale ont été mesurées à 9 et 13 ans. Lorsque les 15 arbres sont classés dans l'ordre croissant de leur valeurs pour les deux caractères, les quatre arbres possédant les génotypes ( $q_1q_2$ ) ou ( $q_1q_1$ ) au locus n°4406 sont les quatre derniers (voir annexe 3). Cette situation peut être représentée de la façon suivante :

Génotype :	Performances :	
	+	-
$q_2q_2$ ou $q_2a$	0	11
$q_1q_2$ ou $q_1q_1$	4	0

Un test du Fisher exact nous permet de constater que, sous l'hypothèse de l'indépendance entre les performances et le génotype, une telle répartition interviendrait dans moins de 5 cas sur 1000 (Siegel 1956) : le classement observé est remarquable. Les deux hauteurs étudiées sont très corrélées : la première est incluse dans la seconde. Pourtant, les locus n°4305 et 4202 n'ont un effet significatif que sur la seconde hauteur, et d'après un test du Fisher exact, le classement observé n'est significatif qu'à 2.5%.

Le classement observé pour l'écart à la verticalité, mesuré à 9 ans dans le même test, donne parallèlement un classement des génotypes au locus n°3206 significatif à 0.5% d'après le test de Fisher

<sup>1</sup> test n°2.44.17

exact (5 arbres  $q_1q_2$  sont en tête, l'un étant mal classé, voir annexe 3). Les arbres ont été également évalués sur leurs greffes dans un parc à clones. Le classement pour la somme des pousses secondaires fait apparaître un classement pour le génotype au locus n°3111 significatif à 2.5% d'après le test du Fisher exact.

L'étude de l'influence d'un stress hydrique sur de jeunes plants, bien que réalisée sur un nombre réduit de plants par famille, concerne 16 des 18 arbres de l'échantillon. Le classement du génotype des arbres au locus n°2428 est significatif à 0.5%.

Les caractères mesurés en phytotron sur un plus grand nombre d'individus par famille ne sont malheureusement disponibles que sur une dizaine de familles de l'échantillon. Avec un sous-échantillon de cette taille, des corrélations fortuites sont très vraisemblables. Le test du Fisher exact permet de repérer les classements les plus remarquables. En première période de croissance, seules 9 familles sont représentées. L'effet des locus n°2322 et 3227 sur le nombre de branches et des locus n°2421 et 3418 sur le poids des branches correspond à des classements significatifs à 5 et 2.5% respectivement. En deuxième période de croissance, les locus n°4101, 1410, 2210, 2322, et 3227, présentent des classements significatifs à 2.5%, les locus n°2421 et 3418 des classements significatifs à 1% et les locus n°1215, 4211, 3309 et 2215 des classements significatifs à 0.5%.

Nous faisons l'hypothèse que les classements les plus remarquables ont moins de chance d'être le fruit d'associations fortuites entre marqueurs et caractères que les autres. En choisissant de conserver les classements correspondant aux tests de Fisher exact les plus significatifs (0.5%), nous espérons donc réduire le nombre d'erreurs de type I parmi eux. Ces classements ont deux caractéristiques ; les génotypes identiques sont regroupés et rarement mélangés, et chaque génotype est représenté par 4 arbres au minimum (sauf dans le cas du locus n°1215, voir annexe 3).

### 2.2.2.4. Bilan

Les associations locus/caractère retenues à la suite de notre procédure de tri sont présentées tableau 13. Sept locus différents, situés d'après la carte génétique sur des groupes de liaison distincts, ou non cartographiés sont liés à 7 caractères. Six locus correspondent à des variations de quantité des protéines ou de présence/absence, un seul est lié à une variation de position. La hauteur du cycle réalisé à la suite d'un stress hydrique, qui montrait une corrélation négative avec le nombre total de locus hétérozygotes, est corrélée avec le génotype des arbres à ce dernier locus (n°2428). Ici encore, les familles issues d'arbres hétérozygotes ( $v_1v_2$ ) réalisent un cycle plus court que celles issues d'arbres homozygotes.

Tableau 13 : Associations remarquables trouvées entre 7 locus et 7 caractères.

Caractère	Locus	Groupe de liaison	Génotypes différents
H83, H87 <sup>1</sup>	4406	6	q <sub>2</sub> q <sub>2</sub> +q <sub>2</sub> a<q <sub>1</sub> q <sub>2</sub> +q <sub>1</sub> q <sub>1</sub>
ECV83 <sup>2</sup>	3206	3	q <sub>1</sub> q <sub>2</sub> <q <sub>2</sub> q <sub>2</sub>
HST <sup>3</sup>	2428	nc	v <sub>1</sub> v <sub>2</sub> <v <sub>2</sub> v <sub>2</sub>
NB1 <sup>4</sup>	1215	nc	q <sub>2</sub> q <sub>2</sub> +q <sub>1</sub> q <sub>1</sub> <q <sub>1</sub> q <sub>2</sub>
PAI2 <sup>5</sup>	3309	17	pp<pa
MSUL3 <sup>6</sup>	4211	5	pp<pa
PB3 <sup>7</sup>	2215	nc	pp<aa<pa

<sup>1</sup> Hauteurs à 9 et 13 ans

<sup>2</sup> Ecart à la verticalité à 9 ans

<sup>3</sup> Hauteur du cycle réalisé après un stress hydrique

<sup>4</sup> Nombre de branches du premier cycle en deuxième période de croissance

<sup>5</sup> Poids des aiguilles primaires du deuxième cycle en deuxième période de croissance

<sup>6</sup> Longueur moyenne d'un entre-noeud du troisième cycle en deuxième période de croissance

<sup>7</sup> Poids des branches du troisième cycle en deuxième période de croissance.

Avec le dispositif utilisé (petit échantillon d'arbres non apparentés), si l'effet d'un locus est détecté, on peut supposer que cet effet est très fort dans la population. Il semble que ce genre d'effet existe, puisque, parmi les associations détectées, certaines ne sont pas fortuites. En supposant que notre procédure de tri a été efficace, des effets s'observent pour des locus différents à chaque fois, sur des caractères de croissance mesurés à différents âges et dans trois conditions différentes (test de descendance, stress hydrique et conditions de croissance accélérée).

### 2.2.3 Signification possible des associations

Nous avons caractérisé le génotype des arbres mères des familles testées pour la croissance en utilisant les protéines exprimées dans les endospermes de ces arbres.

Chez les gymnospermes, le développement de l'endosperme est achevé avant la fertilisation, cet organe a donc une triple fonction de réserve : pendant le développement de l'embryon, pendant la germination et au cours de l'émergence de la plantule (Misra et Green 1991). De nombreuses protéines sont présentes aussi bien dans l'embryon que dans le mégagamétophyte d'après des comparaisons d'électrophorèses monodimensionnelles (Misra et Green 1990, \*Hakman et al. 1990, Gifford et al. 1991). La comparaison des électrophorèses bidimensionnelles de ces deux composantes de la graine chez le Pin maritime conduit à la même observation (Bahrman, résultats non publiés). Donc à ce stade, ces deux organes ont subi une synthèse protéique qualitativement proche, le génome contenu dans l'endosperme étant d'ailleurs parfaitement identique à la moitié du génome qui compose l'embryon de la même graine. L'endosperme comprend, en pourcentage de son poids sec 12% de protéines, l'embryon en contient 15% (Groome et al. 1991). Pendant les premiers jours de la germination, l'embryon est indépendant et utilise ses propres réserves, il fait appel aux sucres et aux acides aminés de l'endosperme ensuite (Groome et al. 1991).

Les protéines de l'endosperme ont donc une influence directe sur le développement des plantules. Si nous avons pu suggérer une relation de cause à effet entre la variabilité des protéines et le poids de l'endosperme qui les contenait, le passage du génotype diploïde de l'arbre mère, qui peut dans notre cas être vu comme un génotype "moyen" de ses endospermes, aux caractéristiques de croissance des familles de demi-frères issus des graines de cet arbre est moins direct.

Les caractères étudiés ici, des aptitudes générales à la combinaison évaluées sur des familles de demi-frères, ne concernent que l'effet additif des gènes de l'arbre mère. Lorsqu'un arbre est de génotype (Aa) à un locus, les individus composant la famille de demi-frères issue de cet arbre reçoivent soit l'allèle A, soit l'allèle a. Prenons l'exemple du locus n°3206, comparé à l'écart à la verticalité (tableau 13). La valeur moyenne des familles issues des arbres de génotype q<sub>2</sub>q<sub>2</sub> est plus élevée que celle des familles issues des arbres de génotype q<sub>1</sub>q<sub>2</sub>. Les individus des premières familles possèdent tous l'allèle q<sub>2</sub>,

alors que les individus des deuxièmes familles possèdent l'allèle  $q_1$  avec une probabilité de  $1/2$ . Il faut donc supposer que la présence de l'allèle  $q_1$  est liée à une valeur moyenne supérieure pour le caractère. Ce raisonnement peut être tenu pour tous les exemples du tableau 13 sauf dans le cas des locus n°1215 et 2215. Les arbres hétérozygotes  $q_1q_2$  au locus n°1215 donnent des familles de performances supérieures à celles des familles issues des deux homozygotes. Pourtant, les individus des deux groupes de familles reçoivent tous soit l'allèle  $q_1$ , soit l'allèle  $q_2$  : où est la différence? Au locus n°2215, les arbres homozygotes pp donnent lieu à des poids de branches faibles, les homozygotes aa à des poids plus élevés. Les individus d'une famille issue d'un arbre hétérozygote bénéficient soit de l'allèle a, soit de l'allèle b : leurs performances devraient être intermédiaires, et elles sont pourtant supérieures à celles des familles des deux homozygotes.

Il faudrait dans ces deux exemples faire l'hypothèse que l'hétérozygotie confère un avantage au niveau diploïde à la mère des familles, qui est transmis par effet maternel sur les plantules. Cette hypothèse peut être valable dans la mesure où les caractères en cause, dans les deux cas, ont été mesurés sur des plants très jeunes, en phytotron, en deuxième période de croissance. Il se trouve que l'effet du poids des graines sur les caractères mesurés n'a été observé qu'au cours de la première période de croissance en phytotron (Lascoux 1992). Dans cet effet on confond les interventions probables de l'endosperme haploïde, mais aussi ceux de l'environnement diploïde de la graine (arbre, cône, tégument). On peut supposer que la description du génotype des endospermes, et par conséquent de la mère, peut mettre en évidence des effets maternels plus fins. Avec notre approche, nous ne pouvons pourtant pas distinguer les effets maternels dus à l'endosperme haploïde, des effets génétiques stables liés à la transmission d'un allèle de la mère chez les descendants. On peut simplement supposer que les premiers effets ne se détectent que sur les jeunes plants et disparaissent ensuite.

Dans tous les cas, il serait intéressant de comprendre pourquoi la présence d'un allèle particulier entraîne une expression différente des caractères quantitatifs. Les 18 arbres sans apparemment de notre échantillon sont issus d'une population allogame dans laquelle les déséquilibres de liaison sont faibles (Muona 1982). La sélection sur la croissance est-elle capable de maintenir des déséquilibres de liaison suffisamment forts pour être détectés dans un échantillon comme le nôtre? On peut imaginer l'existence de linkats au niveau du génome, qui maintiendraient des relations entre un marqueur et un caractère quantitatif, sachant que ces relations ne peuvent être maintenues dans une population allogame qu'en présence d'une forte épistasie ou d'un linkage important (Hastings 1989).

Sans exclure cette éventualité, on peut aussi suggérer un effet propre des locus étudiés. Nous avons vu que parmi l'ensemble des 17 locus impliqués dans des tests significatifs au départ, la proportion de locus responsables de variations de quantité de protéines était plus élevée que prévu. Parmi les 7 locus finalement retenus, en excluant les locus n°1215 et 2215, précédemment discutés, un seul est responsable de modification de positions (n°2428), deux sont liés à des variations de quantités de

protéines (n°4406 et 3206) et deux sont liés à des présences/absences de spots (n°3309 et 4211, tableau 13). L'ambiguïté de ce dernier type de variation, qui peut correspondre à des variations de quantité, a déjà été signalée.

Si l'on accepte l'hypothèse d'un effet propre de ces locus, il faut d'abord supposer que la variabilité observée dans l'endosperme s'exprime aussi par exemple dans le méristème qui est le siège des processus de croissance. Le facteur génétique à l'origine des différences de quantité d'une protéine observées dans l'endosperme sera supposé agir de façon parallèle lors de l'élaboration du caractère quantitatif. La différence allélique à ce facteur se traduira par une différence au niveau du caractère. Si tel est le cas, on suppose aussi que le facteur agit de la même façon dans des contextes haploïdes (endosperme) et diploïdes (arbre) : son effet est additif.

Chez le maïs, les cas de non-additivité pour les quantités de protéines, observés en comparant les profils protéiques d'hybrides et de leur lignées parentales, visualisés par électrophorèse bidimensionnelle, sont assez peu nombreux. Ces cas semblent liés à des régulateurs agissant non pas aux premières étapes de la synthèse des protéines mais à partir de la traduction. Mais une comparaison des profils de différents tissus, de même ploïdie, révélait d'une part que 70% des variations quantitatives étaient spécifiques à l'organe considéré, et que, d'autre part, les cas de non-additivité étaient tout autant spécifiques (Leonardi 1989). Il existe également des exemples où la régulation de protéines, en l'occurrence d'enzymes, est spécifique de l'allèle observé au locus isoenzymatique (Zitnik et Martin 1987, Bush et Paigen 1992).

La comparaison des profils en électrophorèse bidimensionnelle des bourgeons végétatifs et des aiguilles des 18 pins de notre échantillon révèle 88% de spots communs. Mais les variations de quantité de ces spots d'un organe à l'autre sont importantes. De plus, comme chez le maïs (Leonardi et al. 1988) ou chez la souris (Klose 1982) les protéines spécifiques aux organes semblent plus variables que les protéines non spécifiques (Bahrman, résultats non publiés). Des variations saisonnières de la composition protéique pourraient aussi intervenir (Roberts et al. 1991). Néanmoins les phénomènes de régulation jouent vraisemblablement un rôle important lors de la croissance. L'application de cytokinine, qui stimule la formation de bourgeon, sur des cals de *Petunia* (Renaudin et al. 1990) ou sur des embryons d'épicéa (Stabel et al. 1990) modifie les profils protéiques en électrophorèse bidimensionnelle. Dans les deux cas les modifications de niveaux d'expression sont importantes. L'étude des phénomènes de maturation chez le mélèze suggère également que les taux d'expression de certaines protéines changent entre les stades jeunes et adultes (Greenwood et al. 1989, Hutchison et al. 1990). Des réponses différentes à ces modifications selon le génotype de la plante pourraient être envisagées.

La comparaison des protéines de l'endosperme et des organes végétatifs des pins n'est pas très facile, les protocoles d'extraction protéiques sont en effet très différents (Bahrman, non publié). Si des relations entre des variations observées dans l'endosperme et des caractères continus étaient confirmées, il serait pourtant intéressant de tenter d'établir des liens entre l'expression des protéines dans ces différents organes, notamment pour ce qui concerne la régulation.

D'après nos résultats, obtenus à partir d'un nombre d'arbres très réduit, il semble donc qu'il soit possible de mettre en évidence l'effet de locus particuliers, traduisant la variabilité des protéines exprimées dans les endospermes haploïdes des graines, sur des caractères concernant la croissance. Il apparaît également que les variations induites par des régulateurs soient plus souvent impliquées dans ces effets. Si des effets de ce type pouvaient être confirmés, on pourrait alors supposer que les locus en cause, révélés dans un organe particulier, haploïde qui plus est, correspondent à des gènes de régulation non spécifiques, à effet hiérarchique ou pléiotrope (de Vienne et al. 1990).

#### D. DISCUSSION

##### 1. Les QTL chez des plantes annuelles : intérêts et limites

La recherche de locus à effets quantitatifs, si elle n'est pas une idée nouvelle, a pris un essor particulier chez certaines plantes cultivées, d'abord avec les marqueurs isoenzymatiques puis lorsque des cartes génétiques importantes, basées sur des marqueurs de l'ADN, ont pu être réalisées.

Avec les marqueurs isoenzymatiques, des effets ont été mis en évidence par exemple sur des dimensions, des poids d'organes, des nombres d'éléments chez le maïs (Edwards et al. 1987), sur des données morphologiques similaires et sur des caractères de qualité du fruit chez la tomate (Tanksley et al. 1982, Weller 1987, Weller et al. 1988). Le même type de caractères a pu être marqué, chez le soja, avec des locus RFLP (Keim et al. 1990). Toujours chez la tomate, une carte génétique RFLP complète du génome a été utilisée pour l'étude de caractères quantitatifs. Des techniques biométriques particulières, basées sur le maximum de vraisemblance (Lander et Botstein 1989), et en particulier sur le lod score précédemment discuté, ont été utilisées. Paterson et al. (1988) localisent ainsi des zones chromosomiques impliquées dans la variabilité de la masse, du taux de sucre et du pH du fruit sur la carte génétique.

Les cartes plus denses fournies par les marqueurs RFLP permettent de se "rapprocher" des locus à effet quantitatifs : poursuivant leur étude sur la même population de maïs Edwards et al. (1992) détectent, autour des locus isoenzymatiques qui suggéraient la présence d'un QTL, des RFLP associés à des effets plus forts sur le même caractère. Ces études rapprochées permettent par exemple de s'intéresser aux bases génétiques de corrélations fortes entre caractères. Si une région génomique possède une action sur plus d'un caractère, ce qui a été souvent trouvé, on peut supposer que cette région contient soit plusieurs QTL très proches, soit un locus pléiotrope. Grâce à une cartographie très fine, une corrélation négative existant entre le pH de fruit et le taux de sucre chez la tomate a pu être attribuée à deux QTL très liés (Paterson et al. 1990). Ces auteurs pensent que les corrélations indésirables de ce type sont plus souvent dues à des liaisons génétiques qu'à de la pléiotropie. Pour rompre la corrélation, on peut alors espérer repérer, dans de grandes ségrégations, l'un des rares recombinants favorables.

Lorsque la précision des cartes est augmentée, on tend donc à se rapprocher des caractères responsables des variations. Mais cette précision demande des moyens conséquents, qui ne peuvent être généralisés. On compte donc généralement sur un déséquilibre qui maintient la liaison entre le marqueur et le QTL. Le matériel végétal utilisé dans les études citées ci-dessus favorise cette situation. Chez le maïs il s'agit de populations F<sub>2</sub>, et chez la tomate ou le soja, de descendances F<sub>2</sub> ou de backcross obtenus après un croisement interspécifique initial. Les chances de détecter des associations entre marqueurs et caractères quantitatifs sont maximisées dans ce type de populations où le déséquilibre de liaison est très important. Mais dans l'optique d'une utilisation des QTL en sélection, après leur détection dans ce type de populations, un certain nombre de problèmes sont posés.

Il paraît tout d'abord important de savoir dans quelle mesure les résultats obtenus dans une population peuvent être appliqués à une autre population. Nous avons vu précédemment que les cartes génétiques étaient variables d'un individu à l'autre, sans que l'importance de cette variabilité puisse être plus précisément envisagée. Il apparaît de même assez souvent que les QTL dépendent du contexte génétique dans lequel ils sont recherchés, chez la tomate (Tanksley et Hewitt 1988), chez le soja (Graef et al. 1989), ou chez le maïs (Stuber 1989b) par exemple. Ce résultat est, entre autres, lié au fait que seule la variabilité peut être détectée : deux populations peuvent très bien ne pas être polymorphes aux mêmes QTL (de Vienne 1990). De plus, le nombre d'allèles qui sont susceptibles d'exister à un QTL et le degré de polymorphisme de ce QTL au sein des populations sont inconnus. Il n'est donc pas possible de savoir a priori si un QTL détecté dans une population particulière correspond à un événement rare ou pourra être retrouvé dans d'autres contextes (Edwards 1992).

L'environnement pourrait aussi avoir une influence sur l'expression des QTL. Chez la tomate, sur un total de 29 QTL détectés pour la masse, le pH et le taux de sucre des fruits cultivés dans trois endroits différents, 14% étaient communs aux trois lieux, 34% étaient trouvés dans deux lieux et 52% étaient spécifiques à un lieu (Paterson et al. 1991). Si l'importance de l'interaction génotype-milieu sur l'expression des QTL était confirmée, elle obligerait à une certaine prudence dans leur utilisation et notamment pour l'extrapolation de résultats obtenus en un lieu.

Si un déséquilibre de liaison initial ( $D_0$ ) existe dans une population entre deux locus distants de  $r$ , après  $n$  générations de croisements, le déséquilibre de liaison  $D_n$  s'écrira :

$$D_n = (1-r)^n D_0$$

Celui-ci va donc diminuer progressivement, d'autant plus vite que les deux locus seront éloignés. L'utilisation d'un marqueur lié à un QTL dans un programme de sélection est donc problématique, l'érosion de la liaison initiale qui existe entre eux va réduire les gains génétiques à chaque nouvelle génération (Zhang et Smith 1992). Pour que la baisse des gains génétiques soit minimale, on peut envisager d'utiliser l'information de marqueurs flanquant le QTL, ou de marqueurs ayant une distance réduite au QTL (Soller et Plotzkin-Hazan 1977). Malgré ces limites, la sélection assistée par marqueur a donné des résultats intéressants, par exemple chez le maïs en première génération, où la sélection sur un sous-ensemble de marqueurs isoenzymatiques liés au rendement était aussi efficace qu'une sélection massale, qui doit pourtant impliquer le génome en entier (Stuber et al. 1982, Stuber 1989b).

D'après Lande et Thomson (1990), les marqueurs devraient être intégrés dans les schémas de sélection, en ajoutant dans les index une "performance moléculaire", pour utiliser l'effet additif des QTL. Cette procédure est intéressante si, pour un caractère influencé par le milieu, de faible héritabilité, une fraction substantielle de la variance génétique additive est associée aux marqueurs. Le gain est moins net pour des caractères de bonne héritabilité pour lesquels, par définition, la valeur du génotype est bien prédite par le phénotype : les marqueurs apportent peu. La sélection sur marqueur pourrait être très utile lorsque le caractère à sélectionner ne peut être évalué que sur des individus particuliers de la population (Lande et Thomson 1990). L'évaluation est faite sur un seul sexe dans le cas des bovins pour la production de lait (Soller et Beckman 1983). De la même façon, dans le cas des espèces forestières par exemple, la sélection doit être réalisée de façon obligatoire sur des individus jeunes, et les marqueurs amélioreraient certainement son efficacité, ne serait-ce que du fait du manque de corrélation entre caractères juvéniles et adultes. Une sélection précoce sur marqueurs pourrait aussi permettre des économies de moyens dans le cas d'espèces annuelles (Tanksley et al. 1981).

Dans tous les cas, l'une des solutions aux problèmes de la sélection sur marqueur, pour pouvoir envisager son utilisation à long terme, serait de travailler de plus en plus près des QTL, et donc d'aboutir à une sélection quasiment directe sur ces locus (Zhang et Smith 1992).

## 2. Cas particulier des populations panmictiques

Si la sélection assistée par des marqueurs promet d'être particulièrement intéressante pour l'amélioration des animaux domestiques ou des arbres forestiers, la recherche de QTL et leur utilisation chez ces espèces rencontrent des limitations identiques à celles que nous avons déjà citées mais de façon plus marquée encore.

En comparant des espèces autogames et allogames, Price et al. (1984) détectent des associations significatives entre caractères quantitatifs et génotypes isoenzymatiques chez les premières espèces mais pas chez les secondes. Cette observation s'explique par l'existence d'un déséquilibre de liaison, maintenu par autofécondation, chez les espèces autogames seulement. La dérive, la sélection, la migration, le mélange de populations, ou une réduction de la taille d'une population peuvent également générer des déséquilibres de liaison. Mais un déséquilibre stable ne peut être obtenu entre deux locus que s'ils sont liés génétiquement, particulièrement dans de grandes populations (Avery et Hill 1979, Muona 1982). Dans une population en panmixie de taille effective  $N_e$ , la dérive pourra créer des associations substantielles entre locus si une distance  $r$ , inférieure à  $1/(4N_e)$ , les sépare (Hill et Robertson 1968, cités par Lande et Thomson 1990). Chez les pins, les tailles de populations sont supposées supérieures à 1000 individus, et correspondent donc à des valeurs de  $r$  très faibles. Si des déséquilibres de liaison s'expliquent par un effet de la sélection, ceux-ci auront le même sens et la même importance dans différentes populations, ce qui ne sera pas le cas si la dérive est en cause (Lewontin 1974, cité par Avery et Hill 1979). Mais si un déséquilibre de liaison est détecté, les causes de ce déséquilibre restent difficiles à établir (Hastings 1989). Muona (1982) suppose qu'une sélection épistatique n'interviendrait, dans de grandes populations, qu'entre des locus impliqués dans la même fonction, et non pas entre des locus au hasard. Selon cet auteur, il est donc improbable de trouver des relations intéressantes entre quelques locus isoenzymatiques et des caractères comme la croissance dans les populations d'arbres forestiers. Une comparaison du génotype de l'arbre mère de 41 familles de demi-frères pour 8 locus, et de caractères mesurés sur les plantules chez le sapin Douglas en est un exemple (El-Kassaby 1982).

Malgré ces limitations, l'étude conjointe du génotype et des performances chez les arbres forestiers pourrait s'inspirer des travaux déjà réalisés chez les animaux domestiques : le terme de QTL a été défini lors d'une étude pour ces espèces (Gelderman 1975).

### 3. Recherche de QTL chez les animaux domestiques

La recherche de QTL chez les animaux domestiques repose sur l'étude des descendants d'un individu connu pour être hétérozygote à un locus marqueur. Il s'agit alors de tester si le fait d'avoir hérité l'un ou l'autre allèle à ce locus, et par conséquent un segment chromosomique autour de ce marqueur, a une influence sur un caractère quantitatif, dans le cas de familles de pleins-frères ou de demi-frères (Gelderman 1975, Soller et Genizi 1978).

Ce type d'approche a permis à Kluge et Gelderman (1982) de détecter des segments chromosomiques impliqués dans les variations de la taille et du poids dans un échantillon de 2300 souris issues d'un backcross. En étudiant les groupes sanguins, ainsi que le polymorphisme de protéines et d'enzymes de trois taureaux et de plus de 300 de leurs filles, le génotype de leur mère étant connu ou inconnu, Gelderman et al. (1985) mettent en évidence des effets de "transfert d'allèle" du père sur la production en lait de ses filles. Des différences entre les descendances des trois pères apparaissent, que l'on peut expliquer par les arguments cités plus haut, les pères n'étant pas hétérozygotes pour les mêmes marqueurs ni pour les mêmes QTL. Cette situation oblige donc, si plusieurs descendances sont étudiées simultanément, à avoir une approche qui s'apparente à une analyse hiérarchique, où les sous-groupes d'individus d'un génotype donné sont considérés relativement à chaque père (Weller et al. 1990). La recherche d'associations entre les allèles aux différents locus et le poids des endospermes que nous avons réalisée procède d'une démarche similaire.

Lorsque des QTL sont recherchés dans des familles de demi-frères chez les bovins, le père de la famille est équivalent au parent F1 d'un backcross, et la variation du caractère apportée par les mères fait partie de la variance d'erreur (Goddard 1992). Pour détecter des QTL avec ce schéma, Soller et Genizi (1978) préconisent des effectifs trop grands pour être envisageables, Weller et al. (1990) pensent au contraire qu'avec des cartes type RFLP, des effectifs de 5 à 20 taureaux et 200 à 2000 filles par taureau peuvent donner des résultats valables. Ces effectifs restent considérables. Pour réduire le nombre d'individus dont le génotype doit être établi, il est possible de ne travailler, après évaluation, que sur les individus dont les performances dévient substantiellement de la moyenne (Soller et Beckman 1983, Lander et Botstein 1989).

La nécessité de grands effectifs est liée en partie aux situations ambiguës où la transmission d'un allèle ne peut être suivie. Si, par exemple, les deux parents sont hétérozygotes à un locus biallélique et l'un de leurs descendants est également hétérozygote, il n'est pas possible de savoir d'où provient chacun des allèles de ce descendant. Les différentes situations où le transfert d'un allèle est ou n'est pas détectable peuvent être envisagées, connaissant le génotype du descendant et sachant celui du père et éventuellement celui de la mère (Gelderman et al. 1985). Il apparaît d'ailleurs que la prise en compte des informations issues de situations incertaines augmente l'efficacité de la détection d'une substitution chromosomique (Dentine et Cowan 1990). Mais l'idéal serait de travailler sur des marqueurs très

polymorphes, de type mini- ou microsattellites, pour lesquels les individus diffèrent tous. Le problème, avec ces marqueurs, reste la réalisation de leur cartographie, les notions de locus et d'allèles associées aux bandes visualisées étant loin d'être immédiates (Beckman et Soller 1988, Soller 1990).

Des études sont pourtant en cours pour cartographier ce type de marqueurs chez des espèces animales et ils sont utilisés pour localiser des gènes majeurs (voir Haley 1991). Indépendamment du marquage, des techniques biométriques ont en effet été développées et utilisées en génétique animale pour détecter l'effet de gènes majeurs sur des caractères quantitatifs intéressants (Hoeschele 1988, Elsen et al. 1991). L'utilisation d'une telle approche devrait être réfléchie dans le cas des arbres forestiers. Les modèles qui permettent d'étudier la ségrégation de gènes de maladies humaines à l'aide de marqueurs peuvent alors s'appliquer dans le cas de ces gènes majeurs (Elston 1992).

Lorsqu'une substitution chromosomique est détectée au cours d'une génération, elle devra être réévaluée à la génération suivante, à moins que le taux de recombinaison entre le marqueur et le QTL ne soit extrêmement faible (Dekkers et Dentine 1991). Si le marqueur n'est pas lui-même un QTL, ces auteurs pensent pourtant que des effets forts associés à des substitutions chromosomiques ne seraient pas rares dans des populations en ségrégation. Néanmoins, la sélection assistée par des marqueurs nécessite des connaissances plus poussées pour qu'elle puisse être appliquée à des coûts raisonnables et de façon efficace chez ces espèces.

### 4. Projets chez les conifères

Si la recherche de QTL a donné lieu à un certain nombre de travaux en génétique animale, les espèces forestières sont moins avancées. Les problèmes apparaissent assez semblables dans les deux cas, une famille de demi-frères de père commun chez les bovins serait à mettre en parallèle, par exemple dans le cas des pins, à une famille issue d'un arbre mère.

Malgré l'équilibre de liaison, il est clair que si des cartes denses sont disponibles, des QTL pourront être détectés dans ces populations. Quelques projets dans ce sens sont en cours, et utilisent essentiellement le polymorphisme de fragments d'ADN amplifiés (RAPD).

Pour réduire la taille des effectifs nécessaires, ces projets ne s'intéressent qu'aux phénotypes extrêmes. Ils reprennent pour cela l'idée d'une technique proposée par Michelmore et al. (1991) pour détecter rapidement des marqueurs liés à un caractère. Dans une population F2 de laitue ségrégeant pour une résistance au mildiou, ces auteurs mélangent l'ADN des individus homozygotes au locus de résistance et comparent le polymorphisme de marqueurs RAPD des deux mélanges. Si un produit d'amplification est présent dans l'un des mélanges, mais absent dans l'autre, les parents du croisement sont testés pour l'amorce d'ADN correspondante, pour vérifier leur différence. Ayant réalisé une carte génétique sur la

même population, sur laquelle le gène de résistance était localisé, la liaison de trois locus, détectés par cette technique, avec ce gène a pu être confirmée. Michelmore et al. (1991) suggèrent d'appliquer cette démarche dans le cas d'espèces allogames.

Sur *Pinus taeda*, une première équipe a pour objectif de trouver des QTL pour l'infradensité du bois, en utilisant un pedigree à trois générations, et en comparant des descendances de la troisième génération aux performances contrastées avec des marqueurs isoenzymatiques, RFLP et RAPD (Groover et al. 1992). Cette même équipe, sur *Pinus lambertiana*, compare des individus issus du croisement d'un arbre résistant à une rouille, de génotype hétérozygote Rr au locus conférant la résistance, avec un arbre sensible de génotype rr. Ils mélangent pour cela l'ADN des mégagamétophytes des individus résistants d'une part et sensibles d'autre part, et comparent les profils des mélanges avec des marqueurs RAPD (Devey et al. 1992). Parallèlement, une autre équipe propose une méthode d'étude un peu différente (Wilcox et al. 1992). Ils projettent de travailler sur des familles de demi-frères, donc uniquement sur l'effet additif des gènes, en comparant les mélanges d'ADN des mégagamétophytes d'individus aux performances contrastées. Les résultats de leurs simulations indiquent que 1000 individus par familles sont nécessaires pour détecter, avec un niveau de signification de 0.001%, des QTL ayant une contribution de 5% sur la variance génétique. Pratiquement, la technique consiste à faire germer les graines et à récupérer l'endosperme haploïde après germination. L'ADN correspondant à chaque endosperme de chaque plantule peut alors être extrait, et une cartographie de marqueurs RAPD peut être réalisée. Les mesures sur les plantules seront réalisées dans des conditions qui devraient leur permettre d'exprimer leurs potentialités de croissance adulte (O'Malley, communications personnelles).

Si certains croient à l'utilisation possible de la sélection assistée par marqueur chez les arbres de façon relativement rapide, et pensent que l'obstacle majeur est l'absence de résultats (O'Malley 1992), d'autres sont plus réservés sur la question. On peut en effet penser que des résultats seront obtenus à la suite des études engagées mais les désavantages déjà cités pour leur utilisation ultérieure ne manqueront pas de réapparaître : Strauss et al. (1991) invoquent les problèmes de coût, d'équilibre de liaison, de spécificité des QTL au contexte génétique, et d'interaction génotype-milieu. Il semble en effet important, si cette sélection veut à terme être utilisée, d'essayer de mieux cerner, dans un premier temps, l'expression génétique des caractères intéressants comme la croissance, d'autant que ce caractère au stade adulte est difficile à prévoir précocement. Si les causes génétiques du manque de corrélation juvénile-adulte était mieux comprises, des stratégies de sélection précoce efficaces pourraient être envisagées, avec l'aide des marqueurs.

## E. CONCLUSION

### 1. Place de l'électrophorèse bidimensionnelle

L'électrophorèse bidimensionnelle des protéines a une place originale par rapport aux outils évoqués précédemment. Pour rechercher des QTL, les marqueurs de l'ADN apparaissent les plus adéquats puisqu'ils permettent d'explorer l'ensemble du génome. Mais s'il s'agit de se rapprocher de l'expression des caractères quantitatifs, non plus de façon statistique, grâce à des liaisons génétiques, mais dans le but de mieux comprendre les raisons de leur variabilité, l'étude des différents types de variations d'un grand nombre de protéines peut être un bon outil. Il apparaît en effet à travers les diverses études de QTL que ce rapprochement est de plus en plus souhaitable, dès lors que leur utilisation en sélection est envisagée. Ceci est particulièrement vrai chez les espèces à longue durée de génération où l'équilibre de liaison est la règle, mais devient également intéressant lorsque plus d'une population de plantes annuelles, pourtant plus facile à "manipuler", est prise en compte.

Les protéines interviennent directement dans les processus biochimiques, certes extrêmement complexes (Peterson 1992), qui sont à la base du fonctionnement de la plante. Il paraît donc naturel de faire l'hypothèse que celles-ci sont plus proches de la "réalisation" du phénotype (Gallais 1990) que l'ADN, support d'information en amont de tous les processus. La variabilité des protéines aurait donc plus de sens au regard des variations exprimées au niveau de l'organisme. De plus l'activité d'éléments régulateurs, révélée en particulier par les variations de quantités des protéines, serait elle-même plus pertinente pour expliquer des variations de caractères continus. Cette hypothèse est justifiée par Hedrick et McDonald (1980) dans un modèle théorique. Ils distinguent les locus "producteurs", des locus qui contrôlent le moment et la quantité d'expression des premiers locus. Les premiers locus sont donc hiérarchisés aux seconds, et si les locus de contrôle interviennent tôt dans les processus, ce qui est vraisemblable, une modification du contrôle qu'ils exercent aurait un effet important sur le caractère résultant.

Ces arguments font de l'électrophorèse bidimensionnelle un outil intéressant pour une autre approche des QTL. Elle permet en effet de révéler un grand nombre de protéines, d'étudier entre autres la variation de leur quantité, et, comme nous l'avons montré, de cartographier les locus en cause, répartis sur une bonne partie du génome. Les résultats obtenus sur le maïs avec cette technique ont montré que des différences de quantités de protéines étaient corrélées à des différences de performances (Damerval et al. 1987a, Leonardi et al. 1991). Les pins pourraient constituer un autre matériel pour ce type d'études.



## 2. Perspectives pour les pins

Notre approche, réalisée sur un échantillon de taille réduite, mais décrit par un grand nombre de locus, constituait un premier pas dans ce sens. Nos résultats suggèrent que certaines variabilités de protéines mises en évidence dans l'endosperme de l'arbre mère de familles de demi-frères pourraient être en relation avec les performances de ces familles à différents âges. De plus, ces variations seraient plutôt liées à l'action de régulateurs.

Cette étude devrait être poursuivie et concentrée sur un matériel mis en place dans ce but. L'utilisation de l'analyse automatique des gels d'électrophorèse permettra de répertorier les variations et de créer, comme nous l'avons fait (figure 4), une référence facilitant les comparaisons des données d'un arbre à l'autre. La détection des variations quantitatives sera affinée, et des locus exprimant plus de deux allèles pourront éventuellement être détectés. Les questions des caractères à étudier et du matériel à utiliser devront être approfondies, et nécessiteront la maîtrise de techniques un peu particulières.

Nous avons vu en effet que si l'on dispose de nombreux caractères et de nombreux locus, la probabilité d'observer des associations fortuites, quelle que soit la méthodologie de test utilisée, est très élevée. Il paraît donc important de cibler les caractères à étudier. Gottlieb (1984) pense que l'analyse du développement d'un caractère comme la hauteur d'une plante, peut permettre de trouver des composants discrets pertinents, contrôlés par quelques gènes ou un seul. La décomposition de la croissance, suivant la procédure engagée par Kremer (1985), devrait être une bonne façon de se rapprocher du fonctionnement de la plante et plus particulièrement du méristème, à différents âges. Le méristème apical contient en effet l'information qui permet la mise en place des différentes parties de la plante et l'étude des gènes exprimés dans cet organe se développe (Medford et al. 1991).

Ayant défini les caractères à retenir, il serait du plus grand intérêt de pouvoir associer à chaque plantule étudiée le génotype de l'endosperme présent dans la graine dont elle est issue. La ségrégation des marqueurs serait disponible en parallèle à la ségrégation des caractères mesurés. La procédure que nous avons utilisée pour étudier l'effet des marqueurs sur le poids des endospermes pourrait être généralisée à des caractères plus nombreux. Cela nécessiterait de faire germer l'embryon seul, débarrassé du reste de la graine ou accompagné d'une moitié d'endosperme par exemple, en conditions contrôlées. La possibilité d'utiliser une telle technique devrait être étudiée. Les graines issues d'individus dont les deux parents auraient des types de croissance contrastés, comme par exemple les arbres n°1 et 18 de notre échantillon, pourraient être utilisées. Des arbres provenant de ce type de croisement, actuellement en âge de fructifier, doivent être disponibles. En étudiant le génotype de l'endosperme seulement, la contribution paternelle ne pourrait pas être connue plus précisément. On pourrait donc choisir de prendre un père unique pour tous les individus, ou, à l'exemple des bovins, d'envisager plusieurs familles de pleins-frères de mère identique et de pères différents. Le principal obstacle à la réalisation de croisement chez les pins est un problème de temps : en conditions

naturelles, 3 années sont nécessaires pour la mise en place d'une graine. Les cônes sont initiés la première année, leur pollinisation intervient la deuxième année, et la fécondation et la maturation ont lieu la troisième année. Les graines issues d'un croisement contrôlé ne sont donc à maturité que deux ans après le croisement.

Les plantules obtenues à partir des graines pourraient être bouturées. Ceci est réalisable chez le Pin maritime. Sur des plantules âgées d'un mois, la pousse terminale est sectionnée ce qui permet à plusieurs pousses axillaires de se développer. Ces pousses sont alors prélevées et enracinées grâce à un traitement hormonal de leur base. Cette technique peut procurer 10 boutures pour un même plant. Le plus intéressant serait de diviser ces boutures en deux lots, le premier étant installé en conditions de croissance normale, le second en conditions de croissance accélérées, pour faire exprimer aux plants un comportement adulte. Le taux de multiplication semble néanmoins un peu faible pour qu'un tel dispositif donne lieu à des résultats valables. Une seule condition de croissance pourrait donc être testée pour commencer, pour étudier l'évolution des caractères au cours du temps.

Chaque endosperme prélevé ayant été étudié par électrophorèse bidimensionnelle, une carte génétique des variations observées serait disponible, et à chaque gel pourrait être attribué des caractères quantitatifs mesurés sur les plantules. Des approches multilocus, du type de celles qui ont été réalisées chez le maïs (Damerval et al. 1987a, Leonardi et al. 1991) en distinguant les sous-ensembles de locus selon leur action sur les protéines, pourront être réalisées. L'effet individuel des locus sur des caractères cibles de la croissance pourra être également très précisément étudié, et notamment en ce qui concerne les locus responsables de variations de quantité des protéines.

## CONCLUSION GENERALE

L'existence d'un organe haploïde de grande taille chez les pins permet de réaliser des études génétiques généralement difficiles à aborder chez les espèces pérennes.

L'électrophorèse bidimensionnelle des protéines des endospermes des graines de 18 arbres nous a ainsi permis de réaliser une carte génétique du Pin maritime. A la suite d'une analyse systématique des variations observées, 84 locus ont été décrits, impliqués dans des variations de position, de quantité ou de présence/absence des protéines. Parmi ces locus, 65 ont pu être cartographiés, en utilisant des techniques de génétique humaine. Dix-sept groupés de liaison sont représentés pour un total de 530 centimorgans. D'après nos estimations, cette carte doit couvrir plus d'un quart du génome de cet arbre.

Le poids des graines a été utilisé pour mettre en évidence ce que l'on considère aujourd'hui comme l'un des premiers locus à effet quantitatif. Sax (1923) détecte une liaison entre trois locus indépendants déterminant la coloration des graines et le poids de ces graines chez des haricots (voir l'illustration de Soller (1990)). Quoique dans un contexte très différent, nous avons pu montrer que le poids des endospermes des graines des arbres de notre échantillon était influencé par la présence de différents allèles à certains locus.

Le génotype diploïde des pins aux 84 locus a été confronté à leurs performances de croissance. Nos résultats suggèrent que la variabilité des quantités de protéines pourrait être liée à la variation de caractères quantitatifs. Nous suggérons des expériences qui devraient permettre de tester cette hypothèse plus finement, sur des échantillons de plus grandes tailles et sur des caractères ciblés.

La conjonction de l'étude du génome et de l'analyse des bases génétiques des caractères quantitatifs peut donc être réalisée, chez le Pin maritime, en utilisant l'électrophorèse bidimensionnelle.

Pour une meilleure connaissance du génome, l'expression de la recombinaison mériterait d'être approfondie. La variabilité des taux de recombinaison selon les individus, ou selon le sexe serait alors appréhendée sur une bonne partie du génome. La relativité des cartes génétiques pourrait être mesurée. Utilisées en complément à l'électrophorèse bidimensionnelle, les techniques de biologie moléculaire appliquées aux génomes eucaryotes devraient permettre de mieux connaître leur organisation (Lapitan 1992). Dans le cas particulier des pins, la possibilité de cartographier en parallèle des marqueurs de variations protéiques et des marqueurs de l'ADN devrait être étudiée. Ces approches apporteront des informations quant à la signification de la taille du génome.

Les variations observées au niveau protéique reflètent des modifications de gènes de structure, de gènes de régulation ou de gènes modificateurs (Klose 1992). Avec le matériel haploïde utilisé chez les pins, les liaisons génétiques entre des modificateurs différents affectant une même protéine peuvent être étudiées dans quelques cas particuliers. L'ensemble des variations d'une protéine peut alors être considéré comme la plus petite unité polygénique d'un caractère polygénique complexe (Klose 1991). De plus les variations de quantité de protéines ne seraient pas neutres par rapport aux caractères quantitatifs. Il devient alors intéressant, comme de Vienne et al. (1991) le suggèrent, de situer différents types de locus sur la carte génétique. Les variations de quantités de protéines peuvent faire l'objet de recherche de locus à effet quantitatif (QTL). En effet, grâce à l'analyse automatique des gels d'électrophorèse, ces variations sont traitées comme des caractères continus. Les gènes de structure des protéines dont les quantités varient peuvent également être localisés. Enfin, il est également possible de repérer les QTL "classiques" sur les mêmes cartes. Les situations relatives des différents locus seront donc connues. Des coïncidences, sur la carte génétique, entre QTL et locus contrôlant la quantité de protéine pourraient conduire à une analyse physiologique et moléculaire des caractères quantitatifs (de Vienne et al. 1991)

L'étude des bases génétiques des caractères quantitatifs chez le pin pourrait s'inspirer de ces approches. Cette étude pourrait être axée sur la biologie du développement, et sur une étude fine des processus de croissance. La détermination des protéines et de leurs éventuels modificateurs serait concentrée sur des protéines particulières, aux variations liées aux caractéristiques de la plante entière. La cartographie du génome sera associée à une connaissance plus précise de l'expression des protéines, à une approche physiologique des caractères, et à leur étude biométrique. Une compréhension plus précise de la biologie de cet arbre pourra alors être envisagée.

## REFERENCES BIBLIOGRAPHIQUES

- Adams W.T., Neale D.B., Doerksen A.H., Smith D.B., 1990. Inheritance and linkage of isozyme variants from seed and vegetative bud tissues in coastal Douglas-fir (*Pseudotsuga menziesii* var. *menziesii* (Mirb) Franco). *Silvae Genetica* 39 : 3-4.
- Allendorf F.W., Knudsen K.L., Blake G.M., 1982. Frequencies of null alleles at enzyme loci in natural populations of ponderosa and red pine. *Genetics* 100 : 497-504.
- Allona I., Casado R., Aragoncillo C., 1992. Seed storage proteins from *Pinus Pinaster* Ait. : homology of major components with 11S proteins from angiosperms. *Plant Science*, in press.
- Anderson N.G., Tollaksen S.L., Pascoe F.H., Anderson L., 1985. Two-dimensional electrophoretic analysis of wheat seed proteins. *Crop Science* 25 : 667-674.
- Arnheim N., Li H., Cui X., 1991. Genetic mapping by single sperm typing. *Animal Genetics* 22 : 105-115.
- Avery P.J., Hill W.G., 1979. Distribution of linkage disequilibrium with selection and finite population size. *Genet Res Camb* 33 : 29-48.
- Avner P., 1991. Sweet mice, sugar daddies. *Nature* 351 : 519-520.
- Ayala F.J., McDonald J.F., 1980. Continuous variation : possible role of regulatory genes. *Genetica* 52/53 : 1-15.
- Bahrman N., Cardin M-L., Seguin M., Zivy M., Thiellement H., 1988. Variability of 3 cytoplasmically encoded proteins in the *Triticum* genus. *Heredity* 60 : 87-90.
- Bahrman N., Damerval C., 1989. Linkage relationships of loci controlling protein amounts in maritime pine (*Pinus pinaster* Ait.). *Heredity* 63 : 267-274.
- Bailey N.T.J., 1961. Introduction to the mathematical theory of genetic linkage. Oxford at the Clarendon Press.
- Baradat Ph., 1980. OPEP, bibliothèque basic de calcul scientifique. INRA, Laboratoire de Génétique et Amélioration des Arbres Forestiers, Bordeaux, document interne.
- Barrett J.W., Cheliak W.M., Knowles P.H., 1987. Variation in the Pgi/Aat linkage group between populations of black spruce. *Can J For Res* 17 : 756-758.
- Beavis W.D., Grant D., 1991. A linkage map based on information from four F2 populations of maize (*Zea mays* L.). *Theor Appl Genet* 82 : 636-644.
- Becker R.A., Chambers J.M., Wilks A.R., 1988. The new S language. A programming environment for data analysis and graphics. Wadsworth & Brooks/Cole. California.
- Beckmann J.S., Soller M., 1988. Detection of linkage between marker loci and loci affecting quantitative traits in crosses between segregating populations. *Theor Appl Genet* 76 : 228-236.
- Beevers L., 1982. Post-translational modifications. in *Encyclopedia of Plant Physiology*, C.Boulter, B.Parthier eds, New series, vol 14A, Springer Verlag Berlin : 136-168.
- Begun D.J., Aquadro C.F., 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* 356 : 519-520.

- Belanger F.C., Kriz A.L., 1991. Molecular basis for allelic polymorphism of the maize globulin-1 gene. *Genetics* 129 : 863-872.
- Bennett M.D., Smith J.B., 1976. Nuclear DNA amount in angiosperms. *Phil Trans R Soc Lond B* 274 : 227-274.
- Bennett M.D., Smith J.B., 1991. Nuclear DNA amount in angiosperms. *Proc R Soc Lond B* 334 : 309-345.
- Bennett M.D., Smith J.B., Heslop-Harrison J.S., 1982. Nuclear DNA amount in angiosperms. *Proc R Soc Lond B* 216 : 179-199.
- Binelli G., Gianfranceschi L., Pe M.E., Taramino G., Busso C., Stenhouse J., Ottaviano E., 1992. Similarity of maize and sorghum genomes as revealed by maize RFLP probes. *Theor Appl Genet* 84 : 10-16.
- Bobola M.S., Eckert R.T., Klein A.S., 1992. Restriction fragment variation in the nuclear ribosomal DNA repeat unit within and between *Picea rubens* and *Picea mariana*. *Can J For Res* 22 : 255-263.
- Bonierbale M.W., Plaisted R.L., Tanksley S.D., 1988. RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* 120 : 1095-1103.
- Botstein D., White R.L., Skolnick M., Davis R.W., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32 : 314-331.
- Boyle T.J.B., Morgenstern E.K., 1985. Inheritance and linkage relationships of some isozymes of black spruce in New Brunswick. *Can J For Res* 15 : 992-996.
- Bridgwater F.E., 1990. Shoot elongation patterns of loblolly pine families selected for contrasting growth potential. *Forest Science* 36 : 641-656.
- Brooks L.D., Marks R.W., 1986. The organization of genetic variation for recombination in *Drosophila melanogaster*. *Genetics* 114 : 525-547.
- Brown G.R., Amarasinghe V., Carlson J.E., 1992. Localization of rDNA loci to metaphase chromosomes of white spruce by fluorescence in situ hybridization. Fifth workshop of the IUFRO working party S20406, Molecular Biology of Forest Trees, Carcans-Maubuisson.
- Brown J., Sundaresan V., 1991. A recombination hotspot in the maize A1 intragenic region. *Theor Appl Genet* 81 : 185-188.
- Brown J.W.S., Bliss F.A., Hall T.C., 1981. Linkage relationships between genes controlling seed proteins in french bean. *Theor Appl Genet* 60 : 251-259.
- Burr B., Burr F.A., Thompson K.H., Albertson M.C., Stuber C.W., 1988. Gene mapping with recombinant inbreds in maize. *Genetics* 118 : 519-526.
- Burt A., Bell G., 1987. Mammalian chiasma frequencies as a test of two theories of recombination. *Nature* 326(6115) : 803-805.
- Burt A., Bell G., Harvey P.H., 1991. Sex differences in recombination. *J evol Biol* 4 : 259-277.
- Bush R.M., Paigen K., 1992. Evolution of b-glucuronidase regulation in the genus mus. *Evolution* 46(1) : 1-15.
- Bush R.M., Smouse P.E., 1991. The impact of electrophoretic genotype on life history traits in *Pinus taeda*. *Evolution* 45(3) : 481-498.

- Bush R.M., Smouse P.E., Ledig F.T., 1987. The fitness consequences of multiple-locus heterozygosity : the relation ship between heterosigosity and growth rate in Pitch Pine (*Pinus Rigida* Mill.). *Evolution* 41(4) : 787-798.
- Caetano-Anolles G., Bassam B.J., Gresshoff P.M., 1991. DNA amplification fingerprinting : a strategy for genome analysis. *Plant Molecular Biology Reporter* 9(4) : 294-307.
- Carlson J.E., Tulsieram L.K., Glaubitz J.C., Luk V.W.K., Kauffeldt C., Rutledge R., 1991. Segregation of random amplified DNA markers in F1 progeny of conifers. *Theor Appl Genet* 83 : 194-200.
- Casey R., 1979. Genetic variability in the structure of the  $\alpha$ -subunit of legumin from *Pisum* - A two-dimensional gel electrophoresis study. *Heredity* 43(2) : 265-272.
- Chakraborty R., 1987. Biochemical heterozygosity and phenotypic variability of polygenic traits. *Heredity* 59 : 19-28.
- Chakravarti A., Lasher L.A., Reefer J.E., 1991. A maximum likelihood method for estimating genome length using genetic linkage data. *Genetics* 128 : 175-182.
- Chandley A.C., Mitchell A.R., 1988. Hypervariable minisatellite regions are sites for crossing-over at meiosis in man. *Cytogenet Cell Genet* 48 : 152-155.
- Chang C., Bowman J.L., Dejohn A.W., Lander E.S., Meyerowitz E.M., 1988. Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 85 : 6856-6860.
- Chappell J., Chrispeels M.J., 1986. Transcriptional and posttranscriptional control of phaseolin and phytohemagglutinin gene expression in developing cotyledons of *Phaseolus vulgaris*. *Plant Physiol* 81 : 50-54.
- Charlesworth B., Langley C.H., Stephan W., 1986. The evolution of restricted recombination and the accumulation of repeated DNA sequences. *Genetics* 112 : 947-962.
- Chee P.P., Jones J.M., Slightom J.L., 1991. Expression of bean storage protein minigene in tobacco seeds : introns are not required for seed specific expression. *J Plant Physiol* 137 : 402-408.
- Cheliak W.M., Pitel J.A., 1985. Inheritance and linkage of allozymes in *Larix laricina*. *Silvae Genetica* 34(4/5) : 142-148.
- Chotai J., 1984. On the lod score method in linkage analysis. *Ann Hum Genet* 48 : 359-378.
- Clerget-Darpoux F., Babron M-C., Bonaiti-Pellie C., 1990. Assessing the effect of multiple linkage tests in complex diseases. *Genetic Epidemiology* 7 : 245-253.
- Collada C., Caballero R.G., Casado R., Aragoncillo C., 1991a. Seed storage proteins in *Fagaceae* : similarity between *Castanea* globulins and *Quercus* glutelins. *Plant Science* 75 : 145-154.
- Collada C., Caballero R.G., Casado R., Salcedo G., Aragoncillo C., 1991b. Subunit structure of legumin-like globulins of *Fagus sylvatica* seeds. *Journal of Experimental Botany* 42(243) : 1305-1310.
- Conkle M.T., 1981. Isozyme variation and linkage in six conifer species. *Isozymes of North American forest trees and forest insects*. USDA Forest ser gen tech rep PSW-48 : 11-17.
- Cooke, 1988. Electrophoresis in plant testing and breeding. In *Advances in electrophoresis*. A. Chrambach, M.J. Dunn, B.J. Radola eds. Verlagsgesellschaft Publishers. Vol 2 : 174-261.

- Cornall R.J., Prins J-B., Todd J.A., Pressey A., de Larato N.H., Wicker L.S., Peterson L.B., 1991. Type I diabetes in mice is linked to the interleukin-1 receptor and Lsh/Ith/Bcg genes on chromosome 1. *Nature* 353 : 262-265.
- Cui X., Li H., Goradia T.M., Lange K., Kazazian H.H., Galas D., Arnheim N., 1989. Single-sperm typing : determination of genetic distance between the G<sub>γ</sub>-globin and parathyroid hormone loci by using the polymerase chain reaction and allele-specific oligomers. *Proc Natl Acad Sci USA* 86 : 9389-9393.
- Cullis C.A., Creissen G.P., Gorman S.W., Teasdale R.D., 1987. The 25S, 18S, and 5S ribosomal RNA genes from *Pinus radiata* D.Don. *Molecular genetics of forest trees*. W.M.Cheliak, A.C.Yapa eds. Proc. second IUFRO working parties in molecular genetics, Canada : 34-40.
- Dale G., Gates C., Vogel H., Daggard G., Teasdale B., 1992. Mapping of the interspecific hybrid between *Pinus caribaea* and *Pinus elliottii* using RAPD markers. Sixth International Conifer Biotechnology Working Group at Radisson Governors Inn, Raleigh, NC.
- Damerval C., le Guilloux M., Blaissonneau J., de Vienne D., 1987b. A simplification of Heukeshoven and Dernick's silver staining of proteins. *Electrophoresis* 8 : 158-159.
- Damerval C., Hebert Y., de Vienne D., 1987a. Is the polymorphism of protein amounts related to phenotypic variability? A comparison of two-dimensional electrophoresis data with morphological traits in maize. *Theor Appl Genet* 74 : 194-202.
- Damerval C., de Vienne D., Zivy M., Thiellement H., 1986. Technical improvement in two-dimensional electrophoresis increase the level of genetic variation detected in wheat-seedling proteins. *Electrophoresis* 7 : 52-54.
- Damerval C., Zivy M., Granier F., de Vienne D., 1988. Two-dimensional electrophoresis in plant biology. In *Advances in electrophoresis*. A. Chrambach, M.J. Dunn, B.J. Radola eds. Verlagsgesellschaft Publishers. Vol 2 : 263-340.
- Daussant J., Sadowski J., Rorat T., Mayer C., Lauriere C., 1991. Independant regulatory aspects and posttranslational modifications of two b-amylases in rye. *Plant Physiol* 96 : 84-90.
- Dekkers J.C.M., Dentine M.R., 1991. Quantitative genetic variance associated with chromosomal markers in segregating populations. *Theor Appl Genet* 81 : 212-220.
- Delseny M., Grellet F., Tremousaygue D., Raynal M., Panabières F., 1986. Structure, évolution et expression de l'ADN nucléaire des plantes supérieures. Colloque de la Société Botanique, Orsay.
- Dentine M.R., Cowan C.M., 1990. An analytical model for the estimation of chromosome substitution effects in the offspring of individuals heterozygous. *Theor Appl Genet* 79 : 775-780.
- Devey M.E., Delfino-Mix A.A., Kinloch B.B., Neale D.B., 1992. Identification of RAPD markers linked with a major gene for resistance to white pine blister rust in sugar pine. Sixth International Conifer Biotechnology Working Group at Radisson Governors Inn, Raleigh, NC.
- Devey M.E., Jermstad K.D., Tauer C.G., Neale D.B., 1991. Inheritance of RFLP loci in a loblolly pine three-generation pedigree. *Theor Appl Genet* 83 : 238-242.
- Devos K.M., Atkinson M.D., Chinoy C.N., Liu C.J., Gale M.D., 1992. RFLP-based genetic map of the homoeologous group 3 chromosomes of wheat and rye. *Theor Appl Genet* 83 : 931-939.
- Diehl W.J., 1988. Genetic of carbohydrate metabolism and growth in *Eisenia foetida* (*Oligochaeta* : *Lumbricidae*). *Heredity* 61 : 379-387.

- Diers B.W., Keim P., Fehr W.R., Shoemaker R.C., 1992. RFLP analysis of soybean seed protein and oil content. *Theor Appl Genet* 83 : 608-612.
- Donis-Keller H. et al., 1987. A genetic linkage map of the human genome. *Cell* 51 : 319-337.
- Doolittle W.F., Sapienza C., 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284 : 601-603.
- Dooner H.K., 1986. Genetic fine structure of the bronze locus in maize. *Genetics* 113 : 1021-1036.
- Dooner H.K., Kermicle J.L., 1986. The transposable element Ds affects the pattern of intragenic recombination at the bz and R loci in maize. *Genetics* 113 : 135-143.
- Douches D.S., Quiros C.F., 1988. Genetic recombination in a diploid synaptic mutant and a *Solanum tuberosum* \* *S.chacoense* diploid hybrid. *Heredity* 60 : 183-191.
- Drayna D., Davies K., Hartley D., Mandel J-L., Camerino G., Williamson R., White R., 1984. Genetic mapping of the human X chromosome by using restriction fragment length polymorphisms. *Proc Natl Acad Sci USA* 81 : 2836-2839.
- Durel C-E., 1990. Paramètres génétiques et sélection en deuxième génération d'amélioration du Pin maritime (*Pinus pinaster* Ait.). Thèse de l'Institut National Agronomique Paris-Grignon.
- Eckert R.T., Joly R.J., Neale D.B., 1981. Genetics of isozyme variants and linkage relationships among allozyme loci in 35 eastern white pine clones. *Can J For Res* 11 : 573-579.
- Edwards M., 1992. Use of molecular markers in the evaluation and introgression of genetic diversity for quantitative traits. *Field Crops Research* 29 : 241-260.
- Edwards M.D., Helentjaris T., Wright S., Stuber C.W., 1992. Molecular-marker-facilitated investigations of quantitative trait loci in maize. 4. Analysis based on genome saturation with isozyme and restriction fragment length polymorphism markers. *Theor Appl Genet* 83 : 765-774.
- Edwards M.D., Stuber C.W., Wendel J.F., 1987. Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and type of gene action. *Genetics* 116 : 113-125.
- El-Kassaby Y.A., 1982. Association between allozyme genotypes and quantitative traits in Douglas-fir. *Genetics* 101 : 103-115.
- El-Kassaby Y.A., Meagher M.D., Parkinson J., Portlock F.T., 1987. Allozyme inheritance, heterozygosity and outcrossing rate among *Pinus monticola* near Ladysmith, British Columbia. *Heredity* 58 : 173-181.
- El-Kassaby Y.A., Sziklai O., Yeh F.C., 1982. Linkage relationships among 19 polymorphic allozyme loci in coastal douglas-fir (*Pseudotsuga menziesii* var. *menziesii*). *Can J Genet Cytol* 24 : 101-108.
- Ellis T.H.N., Turner L., Hellens R.P., Lee D., Harker C.L., Enard C., Domoney C., Davies D.R., 1992. Linkage maps in pea. *Genetics* 130 : 649-663.
- Elsen J.M., Le Roy P., Goffinet B., 1991. Comparison of four statistical tests for genotype determination at a major locus of progeny tested sires. *J Anim Breed Genet* 108 : 167-173.
- Elston R.C., 1992. Segregation and linkage analysis. *Animal Genetics* 23 : 59-62.
- Elston R.C., Stewart J., 1971. A general model for the genetic analysis of pedigree data. *Human Heredity* 21 : 523-542.

- Fischer R.L., Goldberg R.B., 1982. Structure and flanking regions of soybean seed protein genes. *Cell* 29 : 651-660.
- Flavell R., 1980. The molecular characterization and organization of plant chromosomal DNA sequences. *Ann Rev Plant Physiol* 31 : 569-596.
- Flavell R.B., O'Dell M., Smith D.B., Thompson W.F., 1985. Chromosome architecture : the distribution of recombination sites, the structure of ribosomal DNA loci and the multiplicity of sequences containing inverted repeats. in *Molecular form and function of the plant genome*. L van Vloten-Doting, G.S.S.P. Groot, T.C. Hall eds. NATO ASI, 83, Plenum Press, New York : 1-14.
- Flinn B.S., Roberts D.R., Taylor I.E.P., 1991b. Evaluation of somatic embryos of interior spruce. Characterization and developmental regulation of storage proteins. *Physiologia plantarum* 82 : 624-632.
- Flinn B.S., Roberts D.R., Webb D.T., Sutton C.S., 1991a. Storage protein changes during zygotic embryogenesis in interior spruce. *Tree Physiology* 8 : 71-81.
- Freeling M., 1976. Intragenic recombination in maize : pollen analysis methods and the effect of parental *Adh1+* isoalleles. *Genetics* 83 : 701-717.
- Friedman W.E., 1990. Double fertilization in *Ephedra*, a nonflowering seed plant : its bearing on the origin of angiosperms. *Science* 247 : 951-954.
- Friedman W.E., 1991. Double fertilization in *Ephedra trifurca*, a non-flowering seed plant : the relationship between fertilization events and the cell cycle. *Protoplasma* 165 : 106-120.
- Furnier G.R., Knowles P., Alesiuk M.A., Dancik B.P., 1986. Inheritance and linkage of allozymes in seed tissue of whitebark pine. *Can J Genet Cytol* 28 : 601-604.
- Furnier G.R., Stine M., Mohn C.A., Clyde M.A., 1991. Geographic patterns of variation in allozymes and height growth in white spruce. *Can J For Res* 21 : 707-712.
- Gallais A., 1990. *Théorie de la sélection en amélioration des plantes*. Masson, Paris.
- Gebhardt C., Ritter E., Barone A., Debener T., Walkemeier B., Schachtschabel U., Kaufmann H., Thompson R.D., Bonierbale M.W., Ganai M.W., Tanksley S.D., Salamini F., 1991. RFLP maps of potato and their alignment with the homeologous tomato genome. *Theor Appl Genet* 83 : 49-57.
- Gebhardt C., Ritter E., Debener T., Schachtschabel U., Walkemeier B., Uhrig H., Salamini F., 1989. RFLP analysis and linkage mapping in *Solanum tuberosum*. *Theor Appl Genet* 78 : 65-75.
- Geburek T., Wang Q., 1990. Inheritance of isozyme variants and their linkage relationships in Chinese fir (*Cunninghamia lanceolata* Hook). *Euphytica* 49 : 193-201.
- Geburek Th., von Wuehlich G., 1989. Linkage analysis of isozyme gene loci in *Picea abies* (L.) Karst. *Heredity* 62 : 185-191.
- Geetha K.B., Lending C.R., Lopes M.A., Wallace J.C., Larkins B.A., 1991. Opaque-2 modifiers increase  $\gamma$ -zein synthesis and alter its spatial distribution in maize endosperm. *The Plant Cell* 3 : 1207-1219.
- Gelderman H., 1975. Investigation on inheritance of quantitative characters in animals by gene markers. I. Methods. *Theor Appl Genet* 46 : 319-330.

- Geldermann H., Pieper U., Roth B., 1985. Effects of marked chromosome sections on milk performance in cattle. *Theor Appl Genet* 70 : 138-146.
- Gepts P., 1989. Genetic diversity of seed storage proteins in plants. in *Plant population genetics, breeding and genetic resources*. Brown, Clegg, Kahler and Weir editors. Sinauer associates Inc. Sunderland, Massachusetts : 64-82.
- Gifford D.J., 1988. An electrophoretic analysis of the seed proteins from *Pinus monticola* and eight other species of pine. *Can J Bot* 66 : 1808-1812.
- Gifford D.J., Dale P.L., Wenzel K.A., 1991. Lodgepole pine seed germination. III. Patterns of protein and nucleic acid synthesis in the megagametophyte and embryo. *Can J Bot* 69 : 301-305.
- Gillois M., 1991. Gene mapping today : applications to farm animals. *Genet Sel Evol* 23(suppl 1) : 19s-48s.
- Goddard M.E., 1992. A mixed model for analyses of data on multiple genetic markers. *Theor Appl Genet* 83 : 878-886.
- Goldman D., Rathna Giri P., O'Brien S.J., 1987. A molecular phylogeny of the hominoid primates as indicated by two-dimensional protein electrophoresis. *Proc Natl Acad Sci USA* 84 : 3307-3311.
- Gottlieb L.D., 1984. Genetics and morphological evolution in plants. *Am Nat* 123(5) : 681-709.
- Gottlieb L.D., de Vienne D., 1988. Assessment of pleiotropic effects of a gene substitution in *Pea* by two-dimensional polyacrylamide gel electrophoresis. *Genetics* 119 : 705-710.
- Gould S.J., 1984. *Quand les poules auront des dents. - Réflexions sur l'histoire naturelle*. Editions du Seuil.
- Graef G.L., Fehr W.R., Cianzio S.R., 1989. Relation of isozyme genotype to quantitative characters in Soybean. *Crop Sci* 29 : 683-688.
- Graner A., Jahoor A., Schondelmaier J., Siedler H., Pillen K., Fischbeck G., Wenzel G., Herrmann R.G., 1991. Construction of an RFLP map of barley. *Theor Appl Genet* 83 : 250-256.
- Granier F., de Vienne D., 1986. Silver staining of proteins : standardized procedure for two-dimensional gels bound to polyester sheets. *Anal Biochem* 155 : 45-50.
- Grattapaglia D., Wilcox P., Chaparro J., O'Malley D., McCord S., Whetten R., McIntyre L., Sederoff R., Weir B., Doerge R., Bridgewater F., 1992. Genetic map of loblolly pine. Sixth International Conifer Biotechnology Working Group at Radisson Governors Inn, Raleigh, NC.
- Green M., McLeod J.K., Misra S., 1991. Characterization of Douglas fir protein body composition by SDS-PAGE and electron microscopy. *Plant Physiol Biochem* 29(1) : 49-55.
- Greenwood M.S., Hopper C.A., Hutchison K.W., 1989. Maturation in larch. I-Effect of age on shoot growth, foliar characteristics, and DNA methylation. *Plant Physiol* 90 : 406-412.
- Grime J.P., Mowforth M.A., 1982. Variation in genome size - An ecological interpretation. *Nature* 299 : 151-153.
- Groome M.C., Seymour R.A., Gifford D.J., 1991. Hydrolysis of lipid and protein reserves in loblolly pine seeds in relation to protein electrophoretic patterns following imbibition. *Physiologia Plantarum* 83 : 99-106.

- Groover A.T., Devey M.E., Fiddler T.A., Lee J.M., Megraw R.A., Vujcic S.S., Williams C.G., Neale D.B., 1992. Mapping of genetic loci influencing wood specific gravity in loblolly pine (*Pinus taeda*). Sixth International Conifer Biotechnology Working Group at Radisson Governors Inn, Raleigh, NC.
- Guignard P., 1983. Contrôle génétique du développement de semis de Pin maritime (*Pinus pinaster* Ait.). Mise en évidence des effets maternels sur la croissance juvénile. DEA, Biologie et physiologie végétale, Université Bordeaux II.
- Guihenneuc C., 1990. Remarques sur la méthode du lod score en génétique. Pub Inst Stat Univ Paris, XXXV(fasc. 2) : 19-37.
- Guries R.P., Friedman S.T., Ledig F.T., 1978. A megagametophyte analysis of genetic linkage in pitch pine (*Pinus rigida* Mill). Heredity 40(2) : 309-314.
- Guyon J.P., Kremer A., 1982. Stabilité phénotypique de la croissance en hauteur et cinétique journalière de la pression de sève et de la transpiration chez le Pin maritime (*Pinus pinaster* Ait.). Can J For Res 12(4) : 936-946.
- Hager K-P., Jensen U., Gilroy J., Richardson M., 1992. The N-terminal amino-acid sequence of the  $\beta$ -subunit of the legumin-like protein from seeds of *Ginkgo biloba*. Phytochemistry 31(2) : 523-525.
- Haig D., Westoby M., 1991. Genomic imprinting in endosperm : its effect on seed development in crosses between species, and between different ploidy levels of the same species, and its implications for the evolution of apomixis. Phil Trans R Soc Lond B 333 : 1-13.
- Hakman I., Stabel P., Engstrom P., Eriksson T., 1990. Storage protein accumulation during zygotic and somatic embryo development in *Picea abies* (Norway spruce). Physiologia Plantarum 80 : 441-445.
- Haley C.S., 1991. Use of DNA fingerprints for the detection of major genes for quantitative traits in domestic species. Animal Genetics 22 : 259-277.
- Harry D.E., 1986. Inheritance and linkage of isozyme variants in incense-cedar. The Journal of Heredity 77 : 261-266.
- Harry D.E., Mordecai K.S., Kinlaw C.S., Loopstra C.A., Sederoff R.R., 1989. DNA sequence diversity in alcohol dehydrogenase genes from pines. Proc. 20th southern forest tree improvement conference, Charleston, South Carolina, 373-380.
- Hart D.L., 1989. Evolving theories of enzyme evolution. Genetics 122 : 1-6.
- Hastings A., 1989. The interaction between selection and linkage in plant populations. in Plant population genetics, breeding and genetic resources. Brown, Clegg, Kahler and Weir eds. Sinauer associates Inc. Sunderland, Massachusetts : 163-180.
- Hedrick P.W., McDonald J.F., 1980. Regulatory gene adaptation : an evolutionary model. Heredity 45(1) : 83-97.
- Heidecker G., Chaudhuri S., Messing J., 1991. Highly clustered zein gene sequences reveal evolutionary history of the multigene family. Genomics 10 : 719-732.
- Heidecker G., Messing J., 1986. Structural analysis of plant genes. Ann Rev Plant Physiol 37 : 439-466.
- Helentjaris T., 1987. A genetic linkage map for maize based on RFLPs. Trends In Genetics 3(8) : 217-221.

- Helentjaris T., Slocum M., Wright S., Schaefer A., Nienhuis J., 1986. Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphism. Theor Appl Genet 72 : 761-769.
- Heun M., Kennedy A.E., Anderson J.A., Lapitan N.L.V., Sorrells M.E., Tanksley S.D., 1991. Construction of a restriction fragment length polymorphism map for barley (*Hordeum vulgare*) Genome 34 : 437-447.
- Higgins T.J.V., 1984. Synthesis and regulation of major proteins in seeds. Ann Rev Plant Physiol 35 : 191-221.
- Hipkins V.D., Tsai C-H., Strauss S.H., 1990. Sequence of the gene for the large subunit of ribulose 1,5-biphosphate carboxylase from a gymnosperm, Douglas fir. Plant Molecular Biology 15 : 505-507.
- Hoeschele I., 1988. Statistical techniques for detection of major genes in animal breeding data. Theor Appl Genet 76 : 311-319.
- Hong Y-P., Ponoy B., Carlson J.E., 1992. DNA fingerprinting of conifer genomes using RAPD markers. Sixth International Conifer Biotechnology Working Group at Radisson Governors Inn, Raleigh, NC.
- Hulbert S.H., Ilott T.W., Legg E.J., Lincoln S.E., Lander E.S., Michelmore R.W., 1988. Genetic analysis of the fungus, *Bremia lactucae*, using restriction fragment length polymorphism. Genetics 120 : 947-958.
- Hull G.A., Halford N.G., Kreis M., Shewry P.R., 1991. Isolation and characterisation of genes encoding rye prolamins containing a highly repetitive sequence motif. Plant Molecular Biology 17 : 1111-1115.
- Hutchinson J., Rees H., Seal A.G., 1979. An assay of the activity of supplementary DNA in *Lolium*. Heredity 43(3) : 411-421.
- Hutchison K., Harvie P.D., Singer P.B., Brunner A.F., Greenwood M.S., 1990. Nucleotide sequence of the small subunit of ribulose-1,5-biphosphate carboxylase from the conifer *Larix laricina*. Plant Molecular Biology 14 : 281-284.
- Hutchison K.W., Sherman C.D., Weber J., Smith S.S., Singer P.B., Greenwood M.S., 1990. Maturation in larch. II - Effect of age on photosynthesis and gene expression in developing foliage. Plant Physiol 94 : 1308-1315.
- Janczewski D.N., Goldman D., O'Brien J.O., 1990. Molecular genetic divergence of orang utan (*Pongo pygmaeus*) subspecies based on isozyme and two-dimensional gel electrophoresis. The Journal of Heredity 81 : 375-387.
- Jansson S., Gustafsson P., 1987. Cloning of a cDNA encoding the light-harvesting chlorophyll A/B-binding protein (CAB) from scots pine (*Pinus sylvestris* L.). Molecular genetics of forest trees. W.M.Cheliak, A.C.Yapa eds. Proc. of the second IUFRO working parties in molecular genetics, Canada : 19-25.
- Jansson S., Selstam E., Gustafsson P., 1990. The rapidly phosphorylated 25 kDa polypeptide of the light-harvesting complex of photosystem II is encoded by the type 2 cab-II genes. Biochimica et Biophysica Acta 1019 : 110-114.
- Jensen U., Berthold H., 1989. Legumin-like proteins in gymnosperms. Phytochemistry 28(5) : 1389-1394.



- Jiang I.B.-J., 1987. Early testing in forest tree breeding : a review. Forest Tree Improvement 20, proc. of the meeting "Early testing, juvenile-mature correlations, and accelerated generation turnover", Denmark : 45-78.
- de Jong W.W., Zweers A., Cohen L.H., 1978. Influence of single amino-acid substitution on electrophoretic mobility of sodium dodecyl sulfate-protein complexes. Biochemical and Biophysical Research Communications 82(2) : 532-539.
- Jordan B., 1989. Les cartes du génome humain. La Recherche 216 : 1486-1494.
- Juo P.-S., Stotzky G., 1970. Electrophoretic patterns of globulins and albumins in conifer seeds. Can J Bot 48 : 1911-1912.
- Kaback D.B., Guacci V., Barber D., Mahon J.W., 1992. Chromosome size-dependent control of meiotic recombination. Science 256 : 228-232.
- Kang H., 1991. Component of juvenile-mature correlations in forest trees. Theor Appl Genet 81 : 173-184.
- Kaul M.L.H., Murthy T.G.K., 1985. Mutant genes affecting higher plant meiosis. Theor Appl Genet 70 : 449-466.
- Keim P., Diers B.W., Olson T.C., Shoemaker R.C., 1990. RFLP mapping in Soybean : association between marker loci and variation in quantitative traits. Genetics 126 : 735-742.
- Khalil M.A.K., 1985. Heritability of juvenile characters of white spruce (*Picea glauca* (Moench.) Voss.) in central Newfoundland, Canada. Theor Appl Genet 69 : 247-251.
- King J.N., Burdon R.D., 1991. Time trends in inheritance and projected efficiencies of early selection in a large 17-year-old progeny test of *Pinus radiata*. Can J For Res 21 : 1200-1207.
- King J.N., Dancik B.P., 1983. Inheritance and linkage of isozymes in white spruce (*Picea glauca*). Can J Genet Cytol 25 : 430-436.
- Kinlaw C.S., Harry D.E., Sederoff R.R., 1990. Isolation and characterization of alcohol dehydrogenase cDNAs from *Pinus radiata*. Can J For Res 20 : 1343-1350.
- Kinlaw C.S., Harry D.E., Sleeter D.D., Sederoff R.R., 1987. Use of heterologous probes to isolate and characterize conifer genes. Molecular genetics of forest trees. W.M.Cheliak, A.C.Yapa eds. Proc. second IUFRO working parties in molecular genetics, Canada : 9-18.
- Klose J., 1982. Genetic variability of soluble proteins studied by two-dimensional electrophoresis on different inbred mouse strains and on different mouse organs. J Mol Evol 18 : 315-328.
- Klose J., Bach-Nadvornik D., 1991. Molecular genetic analysis of the mouse genome starting from two-dimensional electrophoretic patterns. in 2D-PAGE 91. M.J. Dunn ed. Proc. of the international meeting on two-dimensional electrophoresis, London : 136-144.
- Kluge R., Geldermann H., 1982. Effects of marked chromosome sections on quantitative traits in the mouse. Theor Appl Genet 62 : 1-4.
- Koehn R.K., 1991. The genetics and taxonomy of species in the genus *Mytilus*. Aquaculture 94 : 125-145.
- Kremer A., 1981a. Déterminisme génétique de la croissance en hauteur du Pin maritime (*Pinus pinaster* Ait.). I- Rôle du polycyclisme. Ann Sci Forest 38(2) : 199-222.

- Kremer A., 1981b. Déterminisme génétique de la croissance en hauteur du Pin maritime (*Pinus pinaster* Ait.) III- Evolution des composantes de la variance phénotypique et génotypique. Ann Sci For 38(3) : 355-375.
- Kremer A., 1984. Distribution of relative growth rates and variation of cytological zonation in apical meristems of seedlings of two contrasting open-pollinated jack pine (*Pinus banksiana*) families. Can J For Res 14(3) : 297-310.
- Kremer A., 1985. Component analysis of height growth, compensation between components and seasonal stability of shoot elongation in maritime pine (*Pinus pinaster* Ait.). in Crop Physiology of Forest Trees. P.M.A. Tigerstedt, P. Puttonen, V. Koski, eds. Helsinki University Press : 201-207.
- Kremer A., 1992. Décomposition morphologique de la croissance en hauteur du Pin maritime : architecture génétique et application à la sélection. Thèse de l'Université Paris XI.
- Kremer A., Larson P.R., 1983. Genetic control of height growth components in jack pine seedlings. Forest Science 29(3) : 451-464.
- Kremer A., Lascoux M., Nguyen A., 1991. Morphogenetic subdivision of height growth and early selection in maritime pine. in proc. of the 21st Southern Forest Tree Improvement Conference, Knoxville, TN.
- Kremer A., Roussel G., 1982. Composantes de la croissance en hauteur chez le Pin maritime (*Pinus pinaster* Ait.) Ann Sci Forest 39(1) : 77-98.
- Kriebel H.B., 1985. DNA sequence components of *Pinus strobus* nuclear genome. Can J For Res 15 : 1-4.
- Lande R., Thompson R., 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124 : 743-756.
- 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., 1987. Construction of multilocus genetic linkage maps in humans. Proc Natl Acad Sci USA 84 : 2363-2367.
- 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 for experimental and natural populations. Genomics 1, 174-181.
- Landry B.S., Hubert N., Etoh T., Harada J.J., Lincoln S.E., 1991. A genetic map for *Brassica napus* based on restriction fragment polymorphisms detected with expressed DNA sequences. Genome 34 : 543-552.
- Lapitan N.L.V., 1992. Organization and evolution of higher plant nuclear genomes. Genome 35 : 171-181.
- Lascoux D.M., 1992. Croissance et développement de semis de Pin maritime (*Pinus pinaster* Ait.) en lumière continue : perspectives pour la sélection précoce. Thèse de l'Université Paris XI.
- Lathrop G.M., Lalouel J.-M., 1988. Efficient computation in multilocus linkage analysis. Am J Hum Genet 42 : 498-505.
- Ledig F.T., Guries R.P., Bonefeld B.A., 1983. The relation of growth to heterozygosity in Pitch Pine. Evolution 37(6) : 1227-1238.

- Leigh Brown A.J., Langley C.H., 1979. Reevaluation of level of genic heterozygosity in natural population of *Drosophila melanogaster* by two-dimensional electrophoresis. Proc Natl Acad Sci USA 76 : 2381-2384.
- Leonardi A., 1989. Variabilité et hérédité des quantités de protéines chez le maïs : analyse par électrophorèse bidimensionnelle et relations avec la variabilité agromorphologique. Thèse de l'Université Paris VII.
- Leonardi A., Damerval C., Hebert Y., Gallais A., de Vienne D., 1991. Association of protein amount polymorphism (PAP) among maize lines with performances of their hybrids. Theor Appl Genet 82 : 552-560.
- Leonardi A., Damerval C., de Vienne D., 1987. Inheritance of protein amounts : comparison of two-dimensional electrophoresis patterns of leaf sheaths of two maize lines (*Zea mays* L.) and their hybrids. Genet Res 50 : 1-5.
- Leonardi A., Damerval C., de Vienne D., 1988. Organ specific variability and inheritance of maize proteins revealed by two-dimensional electrophoresis. Genet Res 52 : 97-103.
- Leunissen J.A.M., Van den Hooven H.W., de Jong W.W., 1990. Extreme differences in charge changes during protein evolution. J Mol Evol 31 : 33-39.
- Lewontin R.C., 1984. Detecting population differences in quantitative characters as opposed to gene frequencies. Am Nat 123 : 115-124.
- Lewontin R.C., 1991. Twenty-five years ago in GENETICS : electrophoresis in the development of evolutionary genetics : milestone or millstone? Genetics 128 : 657-662.
- Li H., Gyllensten U.B., Cui X., Saiki R., Erlich H.A., Arnheim N., 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. Nature 335 : 414-417.
- Lincoln S.E., Daly M.J., Lander E.S., 1990. Constructing genetic linkage maps with MAPMAKER : a tutorial and reference manual. Whitehead Institute for Biomedical Research technical report, second edition.
- Lyon M.F., 1990. L.C. Dunn and mouse genetic mapping. Genetics 125 : 231-236.
- MacIntyre R.J., 1982. Regulatory genes and adaptation. Evol Biol 15 : 247-285.
- MacPherson P., Filion W.G., 1981. Karyotype analysis and the distribution of constitutive heterochromatin in five species of *Pinus*. The Journal of Heredity 72 : 193-198.
- Maheswari P., Singh H., 1967. The female gametophyte of gymnosperms. Biol Rev 42 : 88-130.
- Marshall D.R., Brown A.H.D., 1975. The charge-state model of protein polymorphism in natural populations. J Mol Evol 6 : 149-163.
- Martin G.B., Williams J.G.K., Tanksley S.D., 1991. Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines. Proc Natl Acad Sci USA 88 : 2336-2340.
- Mather K., 1963. The measurement of linkage in heredity. London : Methuen & Co. Ltd, New-York : John Wiley & Sons, Inc.
- Matta N.K., Gatehouse J.A., 1982. Inheritance and mapping of storage protein genes in *Pisum sativum* L. Heredity 48(3) : 383-392.

- Medford J.I., Elmer J.S., Klee H.J., 1991. Molecular cloning and characterization of genes expressed in shoot apical meristems. The Plant Cell 3 : 359-370.
- McConkey E.H., Taylor B.J., Phan D., 1979. Human heterozygosity : a new estimate. Proc Natl Acad Sci USA 76(12) : 6500-6504.
- McCouch S.R., Kochert G., Yu Z.H., Wang Z.Y., Khush G.S., Coffman W.R., Tanksley S.D., 1988. Molecular mapping of rice chromosomes. Theor Appl Genet 76 : 815-829.
- Messeguer R., Ganal M., de Vicente M.C., Young N.D., Bolkan H., Tanksley S.D., 1991. High resolution RFLP map around the root knot nematode resistance gene (Mi) in tomato. Theor Appl Genet 82 : 529-536.
- Metakovsky E.V., Novoselskaya A.Yu., Kopus M.M., Sobko T.A., Sozinov A.A., 1984. Blocks of gliadin components in winter wheat detected by one-dimensional polyacrylamide gel electrophoresis. Theor Appl Genet 67 : 559-568.
- Michelmore R.W., Paran I., Kesseli 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 using segregating populations. Proc Natl Acad Sci USA 88(21) : 9828-9832.
- Misra S., Green M.J., 1990. Developmental gene expression in conifer embryogenesis and germination. I. Seed proteins and protein body composition of mature embryo and the megagametophyte of white spruce (*Picea glauca* (Moench) voss.). Plant Science 68 : 163-173.
- Misra S., Green M.J., 1991. Developmental gene expression in conifer embryogenesis and germination. II. Crystalloid protein synthesis in the developing embryo and megagametophyte of white spruce (*Picea glauca* (Moench) voss.). Plant Science, 78 : 61-71.
- Mitton J.B., 1983. Conifers Isozymes. in Plant Genetics and Breeding, part B. S.D. Tanksley, T.J. Orton eds, Elsevier Science Publishers, Amsterdam : 443-472.
- Moran G.F., Bell J.C., Hilliker A.J., 1983. Greater meiotic recombination in male vs. female gametes in *Pinus radiata*. The Journal of Heredity 74 : 62.
- Morizot D.C., 1990. Use of fish gene maps to predict ancestral vertebrate genome organization. Isozymes : structure, function, and use in biology and medicine, Z-I. Ogita and C.L. Markert eds. Liss-Wiley, New York : 207-234.
- Morizot D.C., Slangenaupt S.A., Kallman K.D., Chakravarti A., 1991. Genetic linkage map of fishes of the genus *Xiphophorus* (Teleostei : Poeciliidae). Genetics 127 : 399-410.
- Morris R.W., Spieth P.T., 1978. Sampling strategies for using female gametophytes to estimate heterozygosity in conifers. Theor Appl Genet 51, 217-222.
- Morton N.E., 1955. Sequential tests for the detection of linkage. Am J Hum Genet 7 : 277-318.
- Mukai Y., Yamamoto N., Odani K., Shinohara K., 1991. Structure and expression of a gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from pine. Plant Cell Physiol 32(2) : 273-282.
- Muona O., 1982. Population structure of forest trees. Silva Fennica 16(2) : 107-114.
- Nam H-G., Giraudat J., den Boer B., Moonan F., Loos W.D.B., Hauge B.M., Goodman H.M., 1989. Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana*. The Plant Cell 1 : 699-705.

- Namkoong G., Kang H.C., Brouard J.S., 1988. Tree Breeding : Principles and strategies. Springer Verlag, New York.
- Neale D.B., Adams W.T., 1981. Inheritance of isozyme variants in seed tissues of balsam fir (*Abies balsamea*). Can J Bot 59 : 1285-1291.
- Neale D.B., Devey M.E., Groover A.T., Fiddler T.A., Vujcic C.G., Williams C.G., Megraw R.A., 1992. Molecular marker and quantitative trait mapping in loblolly pine. Fifth workshop of the IUFRO working party S20406, Molecular Biology of Forest Trees, Carcans-Maubuisson.
- Neale D.B., Tauer C.G., Gorzo D.M., Jermstad K.D., 1989. Restriction Fragment length polymorphism mapping of loblolly Pine : methods, applications, and limitations. 20th Southern Forest tree improvement conference. Charleston, South Carolina.
- Neale D.B., Williams C.G., 1991. Restriction fragment length polymorphism mapping in conifers and applications to forest genetic and tree improvement. Can J For Res 21 : 545-554.
- Neel J.V., Baier L., Hanash S., Erickson R.P., 1985. Frequency of polymorphism for alleles encoding for liver proteins of domesticated mice. The Journal of Heredity 76 : 314-320.
- Nei M., 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89 : 583-590.
- Nelson O., 1962. The waxy locus in maize. I. Intralocus recombination frequency estimates by pollen and by conventional analyses. Genetics 47 : 737-742.
- Newton C.H., Flinn B.S., Marquez I., Sutton B.C.S., 1992. Seed storage protein genes of spruce. Fifth workshop of the IUFRO working party S20406, Molecular Biology of Forest Trees, Carcans-Maubuisson.
- Niebling C.R., Johnson K., Gerhold H.D., 1987. Electrophoretic analysis of genetic linkage in scots pine (*Pinus sylvestris* L.). Biochem genet 25 : 803-814.
- Nordheim E.V., O'Malley D.M., Chow S-C., 1984. On the performance of a likelihood ratio test for genetic linkage. Biometrics 40 : 785-790.
- Nordheim E.V., O'Malley D.M., Guries R.P., 1983. Estimation of recombination frequency in genetic linkage studies. Theor Appl Genet 66 : 313-321.
- O'Farrell P.H., 1975. High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250 : 4007-4021.
- O'Malley D.M., 1992. Potential for application of marker assisted selection in forest tree breeding. Sixth International Conifer Biotechnology Working Group at Radisson Governors Inn, Raleigh, NC.
- O'Malley D.M., Allendorf F.W., Blake G.M., 1979. Inheritance of isozyme variation and heterozygosity in *Pinus ponderosa*. Biochemical Genetics 17(3/4) : 233-250.
- O'Malley D.M., Guries R.P., Nordheim E.V., 1986. Linkage analysis for 18 enzyme loci in *Pinus rigida* Mill. Theor Appl Genet 72 : 530-535.
- Ohri D., Khoshoo T.N., 1986. Genome size in Gymnosperms. Pl Syst Evol 153(1-2) : 119-131.
- Ohta T., Dover G.A., 1984. The cohesive population genetics of molecular drive. Genetics 108 : 501-521.
- Oliver S.G. et al., 1992. The complete DNA sequence of yeast chromosome III. Nature 357 : 38-46.

- Orgel L.E., Crick F.H.C., 1980. Selfish DNA : the ultimate parasite. Nature 284 : 604-607.
- de Pace C., Delre V., Scarascia Mugnozza G.T., Maggini F., Cremonono R., Frediani M., Cionini P.G., 1991. Legumin of *Vicia faba major* : accumulation in developing cotyledons, purification, mRNA characterization and chromosomal location of coding genes. Theor Appl Genet 83 : 17-23.
- Paran I., Kesseli R., Michelmore R., 1991. Identification of restriction fragment length polymorphism and random amplified polymorphic DNA markers linked to downy mildew resistance genes in lettuce, using near-isogenic lines. Genome 34 : 1021-1027.
- Paterson A.H., Damon S., Hewitt J.D., Zamir D., Rabinovitch H.D., Lincoln S.E., Lander E.S., Tanksley S.D., 1991. Mendelian factors underlying quantitative traits in tomato : comparison across species, generation and environments. Genetics 127 : 181-197.
- Paterson A.H., DeVerna J.W., Lanini B., Tanksley S.D., 1990. Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of Tomato. Genetics 124 : 735-742.
- Paterson A.H., Lander E.S., Hewitt J.D., Peterson S., Lincoln S.E., Tanksley D.T., 1988. Resolution of quantitative traits into mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335 : 721-726.
- Peetz E.W., Thomson G., Hedrick P.W., 1986. Charge changes in protein evolution. Mol Biol Evol 3(1) : 84-94.
- Perry D.J., Knowles P., 1989. Inheritance and linkage relationships of allozymes of eastern white cedar (*Thuja occidentalis*) in northwestern Ontario. Genome 32 : 245-250.
- Perry T.O., 1976. Maternal effect on the early performance of tree progenies. in Tree Physiology and Yield Improvement. M.G.R. Cannell, F.T. Last eds, Academic Press : 473-481.
- Peterson P.A., 1992. Quantitative inheritance in the era of molecular biology. Maydica 37 : 7-18.
- Petit R., 1988. Ressources génétique du Pin maritime : apport des marqueurs enzymatiques. DEA, Ressources génétique et amélioration des plantes, Université Paris XI.
- Poethig R.S., 1990. Phase change and the regulation of shoot morphogenesis in plants. Science 250 : 923-930.
- Polans N.O., Weeden N.F., Thompson W.F., 1985. Inheritance, organisation, and mapping of *rbcS* and *cab* multigene families in pea. Proc Natl Acad Sci USA 82 : 5083-5087.
- Prager E.M., Fowler D.P., Wilson A.C., 1976. Rates of evolution in conifers. Evolution 30 : 637-649.
- Price S.C., Shumaker K.M., Kahler A.L., Allard R.W., Hill J.E., 1984. Estimates of population differentiation obtained from enzyme polymorphisms and quantitative characters. The Journal of Heredity 75 : 141-142.
- Prum B., 1991. Statistique : Vraisemblance, estimateurs et tests associés. Cours du DEA "Statistique et Santé", Université Paris V.
- Queller D.C., 1983. Kin selection and conflict in seed maturation. J Theor Biol 100 : 153-172.
- Queller D.C., 1984. Models of kin selection on seed provisioning. Heredity 53(1) : 151-165.
- Racine R.R., Langley C.H., 1980a. Genetic analysis of protein variations in *Mus musculus* using two-dimensional electrophoresis. Biochemical genetics 18(1/2) : 185-197.

- Racine R.R., Langley C.H., 1980b. Genetic heterozygosity in a natural population of *Mus musculus* assessed using two-dimensional electrophoresis. *Nature* 283 : 855-857.
- Rake A.V., Miksche J.P., Hall R.B., Hansen K.M., 1980. DNA reassociation kinetic of four conifers. *Can J Genet Cytol* 22 : 69-79.
- Ramagopal S., 1990. Protein polymorphism in sugarcane revealed by two-dimensional gel analysis. *Theor Appl Genet* 79 : 297-304.
- Rees H., Durrant A., 1986. Recombination and genome size. *Theor Appl Genet* 73 : 72-76.
- Reiter R.S., Williams J.G.K., Feldmann K.A., Rafalski J.A., Tingey S.V., Scolnik P.A., 1992. Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. *Proc Natl Acad Sci USA* 89 : 1477-1481.
- Renaudin J-P., Tournaire C., Teyssendier de la Serve B., 1991. Quantitative analysis of protein changes during meristem initiation and bud development in protoplast-derived *Petunia hybrida* callus. *Physiologia Plantarum* 82 : 48-56.
- Rick C.M., 1969. Controlled introgression of chromosomes of *Solanum pennellii* into *Lycopersicon esculentum* : segregation and recombination. *Genetics* 62 : 753-768.
- Ringius G.S., Innes D.J., 1990. Large disparity in recombination fractions reported in *Picea mariana* for the Aat-1/Pgi-2 linkage group. *Can J Bot* 68 : 518-520.
- Risch N., 1992. Genetic linkage : interpreting lod scores. *Science* 255 : 803-804.
- Roberts D.R., Toivonen P., McInnis S.M., 1991. Discrete proteins associated with overwintering of interior spruce and Douglas-fir seedlings. *Can J Bot* 69 : 437-441.
- Robertson D.S., 1984. Different frequency in the recovery of crossover products from male and female gametes of plants hypoploid for B-A translocation in maize. *Genetics* 107 : 117-130.
- Rosenblum B.B., Neel J.V., Hanash S.M., 1983. Two-dimensional electrophoresis of plasma polypeptides reveals "high" heterozygosity indices. *Proc Natl Acad Sci USA* 80 : 5002-5006.
- Rudin D., Ekberg I., 1978. Linkage studies in *Pinus sylvestris* L. - using macro gametophyte allozymes. *Silvae genetica* 27(1) : 1-12.
- SAS Institute Inc., 1990. SAS User's Guide. Version 6, Fourth Edition.
- Santos J.L., Cano M.I., 1990. Evidence of a joint control of chiasma formation in spermatocytes and oocytes of a grasshopper. *Heredity* 65 : 419-422.
- Sax K., 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8 : 552-560.
- Saylor L.C., 1972. Karyotype analysis of the genus *Pinus* - Subgenus *Pinus*. *Silvae Genetica* 21(5) : 155-163.
- Schmid B., 1992. Phenotypic variation in plants. *Evolutionary trends in plants* 6(1) : 45-60.
- Searle S.R., 1971. Linear models. Wiley, New-York.
- Sengupta C., Deluca V., Bailey D.S., Verma D.P.S., 1981. Post-translational processing of 7S and 11S components of soybean storage proteins. *Plant Molecular Biology* 1 : 19-34.

- Shiraishi S., 1988. Linkage relationships among allozyme loci in Japanese black pine, *Pinus thunbergii* Parl. *Silvae genetica* 37(2) : 60-66.
- Shrewry P.R., Parmar S., Fulrath N., Kasarda D.D., Miller T.E., 1986. Chromosomal location of the structural genes for secalin in wild perennial rye (*Secale montanum* Guss.) and cultivated rye (*S. cereale* L.) determined by two-dimensional electrophoresis. *Can J Genet Cytol* 28 : 76-83.
- Siegel S., 1956. Nonparametric statistics for the behavioral sciences. International student Ed.
- Slocum M.K., Figdore S.S., Kennard W.C., Suzuki J.Y., Osborn T.C., 1990. Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea*. *Theor Appl Genet* 80 : 57-64.
- Smith S.C., Racine R.R., Langley C.H., 1980. Lack of genic variation in the abundant proteins of human kidney. *Genetics* 96 : 967-974.
- Smouse P.E., 1986. The fitness consequences of multiple-locus heterozygosity under the multiplicative overdominance and inbreeding depression models. *Evolution* 40(5) : 946-957.
- Sokal R.R., Rohlf F.J., 1981. Biometry. Freeman and Company, New-York.
- Soller M., 1990. Genetic mapping of the bovine genome using deoxyribonucleic acid-level markers to identify loci affecting quantitative traits of economic importance. *J Dairy Sci* 73 : 2628-2646.
- Soller M., Beckmann J.S., 1983. Genetic polymorphism in varietal identification and genetic improvement. *Theor Appl Genet* 67 : 25-33.
- Soller M., Genizi A., 1978. The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. *Biometrics* 34 : 47-55.
- Soller M., Plotkin-Hazan J., 1977. The use marker alleles for the introgression of linked quantitative alleles. *Theor Appl Genet* 51 : 133-137.
- Sorensen F.C., Campbell R.K., 1985. Effect of seed weight on height growth of Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco var. *menziesii*) seedlings in a nursery. *Can J For Res* 15 : 1109-1115.
- Spena A., Viotti A., Pirrotta V., 1983. Two adjacent genomic zein sequences : structure, organization and tissue-specific restriction pattern. *J Mol Biol* 169 : 799-811.
- Spicer G.S., 1988. Molecular evolution among some *Drosophila* species groups as indicated by two-dimensional electrophoresis. *J Mol Evol* 27 : 250-260.
- Stabel P., Erikson T., Engstrom P., 1990. Changes in protein synthesis upon cytokinin-mediated adventitious bud induction and during seedling development in Norway spruce, *Picea abies*. *Plant Physiol* 92 : 1174-1183.
- Strauss S.H., 1986. Heterosis at allozyme loci under inbreeding and crossbreeding in *Pinus attenuata*. *Genetics* 113 : 115-134.
- Strauss S.H., Conkle M.T., 1986. Segregation, linkage, and diversity of allozymes in knobcone pine. *Theor Appl Genet* 72 : 483-493.

- Strauss S.H., Lande R., Namkoong G., 1991. Limitations of molecular-marker-aided selection in forest tree breeding. Marker Aided Selection Workshop, Gatlinburg Tennessee, June 13-14 - To be published in Tree Physiology.
- Strauss S.H., Libby W.J., 1987. Allozyme heterosis in radiata pine is poorly explained by overdominance. The American Naturalist 130(6) : 879-890.
- Stuber C.W., 1989a. Isozymes as markers for studying and manipulating quantitative traits. in Isozymes in plant biology. Soltis D.E. Soltis P.S. editors, Chapman and Hall, London : 206-220.
- Stuber C.W., 1989b. Marker-based selection for quantitative traits. in Science for plant breeding. Proc. XII Congress of EUCARPIA, Göttingen. Paul Parey, Berlin and Hamburg.
- Stuber C.W., Goodman M.M., Moll R.H., 1982. Improvement of yield and ear number resulting from selection at allozyme loci in a maize population. Crop Science 22 : 737-740.
- Szmidt A.E., Muona O., 1989. Linkage relationships of allozyme loci in *Pinus sylvestris*. Hereditas 11 : 91-97.
- Tadmor Y., Zamir D., Ladizinsky G., 1987. Genetic mapping of an ancient translocation in the genus *lens*. Theor Appl Genet 73 : 883-892.
- Tanksley D., Medina-Filho H., Rick C.M., 1981. The effect of isozyme selection on metric characters in an interspecific backcross of tomato - Basis of an early screening procedure. Theor Appl Genet 60 : 291-296.
- Tanksley S.D., 1983. Gene mapping. in Isozymes in plant genetics and breeding, part A. S.D. Tanksley, T.J. Orton eds. Elsevier Science Publishers, Amsterdam : 109-138.
- Tanksley S.D., Hewitt J., 1988. Use of molecular markers in breeding for soluble solids content in tomato - a re-examination. Theor Appl Genet 75 : 811-823.
- Tanksley S.D., Bernatzky R., Lapitan N.L., Prince J.P., 1988. Conservation of gene repertoire but not gene order in pepper and tomato. Proc Natl Acad Sci USA 85 : 6419-6423.
- Tanksley S.D., Rick C.M., 1980. Isozymic gene linkage map of the tomato : applications in genetics and breeding. Theor Appl Genet 57 : 161-170.
- Tanksley S.D., Medina-Filho H., Rick C.M., 1982. Use of naturally-occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. Heredity 49(1) : 11-25.
- Temme D.H., 1986. Seed size variability : a consequence of variable genetic quality among offspring? Evolution 40(2) : 414-417.
- Teoh S.B., Rees H., 1976. Nuclear DNA amounts in populations of *Picea* and *Pinus* species. Heredity 36(1) : 123-137.
- Thiellement H., Seguin M., Bahrman N., Zivy M., 1989. Homology and phylogeny of the A, S and D genomes of the *Triticinae*. J Mol Evol 29 : 89-94.
- Thoday J.M., 1961. Location of polygenes Nature 192 : 368-370.
- Thompson E.A., 1990. R.A. Fisher's contribution to genetical statistics. Biometrics 46 : 905-914.
- Thoraval D., Regnacq M., Neuville P., Boucherie H., 1990. Functional analysis of the yeast genome : use of two-dimensional gel electrophoresis to detect randomly cloned DNA sequences. Curr Genet 18 : 281-286.

- Thuriaux P., 1977. Is recombination confined to structural genes on the eucaryotic genome? Nature 268 : 460-462.
- Todd J.A., Aitman T.J., Cornall R.J., Ghosh S., Hall J.R.S., Hearne C.M., Knight A.M., Love J.M., McAleer M.A., Prins J-B., Rodrigues N., Lathrop M., Pressey A., de Larato N.H., Peterson L.B., Wicker L.S., 1991. Genetic analysis of autoimmune type 1 diabetes mellitus in mice. Nature 351 : 542-547.
- Tomassone R., 1986. Biométrie. Support de cours, Institut National Agronomique Paris Grignon.
- Toon P.G., Haines R.J., Dieters M.J., 1991. Relationship between seed weight, germination time and seedling height growth in *Pinus caribea* Morelet var. *hondurensis* Barrett and Golfari. Seed Sci & Technol 19 : 397-402.
- Tulsieram L., Compton W.A., Morris R., Thomas-Compton M., Eskridge K., 1992. Analysis of genetic recombination in maize populations using molecular markers. Theor Appl Genet 84 : 65-72.
- Turner S.R., Barratt D.H.P., Casey R., 1990. The effect of different alleles at the r locus on the synthesis of seed storage proteins in *Pisum sativum*. Plant Molecular Biology 14 : 793-803.
- Vallejos C.E., Chase C.D., 1991. Extended map for the phaseolin linkage group of *Phaseolus vulgaris* L. Theor Appl Genet 82 : 353-357.
- Vedel F., Delseny M., 1987. Repetitiveness and variability of higher plant genomes. Plant Physiol Biochem 25(2) : 191-210.
- Vershinin A.V., Salina E.A., Solovyov V.V., Timofeyeva L.L., 1990. Genomic organization, evolution, and structural peculiarities of highly repetitive DNA of *Hordeum vulgare*. Genome 33 : 441-449.
- Vicente M.C., Tanksley S.D., 1991. Genome-wide reduction in recombination of backcross progeny derived from male versus female gametes in an interspecific cross of tomato. Theor Appl Genet 83 : 173-178.
- de Vienne D., 1990. L'analyse du déterminisme génétique des caractères quantitatifs chez les végétaux. m/s 10(6) : 11-16.
- de Vienne D., Gottlieb L.D., 1990. Comparison of leaflets and tendrils in wild-type and homeotic mutants morphs of pea by two-dimensional electrophoresis of protein. Journal of Heredity 81 : 117-122.
- de Vienne D., Leonardi A., Damerval C., 1988. Genetic aspects of variation of protein amounts in maize and pea. Electrophoresis 9 : 742-750.
- de Vienne D., Leonardi A., Damerval C., 1991. Maize genetics using 2D-PAGE. in 2D-PAGE 91. M.J. Dunn ed. Proc. of the international meeting on two-dimensional electrophoresis, London : 188-192.
- de Vienne D., Leonardi A., Damerval C., Gallais A., 1990. Recherche des bases moléculaires de la variation des caractères quantitatifs chez le maïs. Communication au congrès de l'Union pour la Protection des Obtentions Végétales.
- Vizir I.Y., Korol A.B., 1990. Sex difference in recombination frequency in *Arabidopsis*. Heredity 65 : 379-383.
- Volle M., 1985. Analyse des données. Edition Economica, collection "Economie et statistique avancées".

- Wagner H., Weber W.E., Wricke G., 1992. Estimating linkage relationship of isozyme markers and morphological markers in sugar beet (*Beta vulgaris* L.) including families with distorted segregations. *Plant Breeding* 108 : 89-96.
- Walbot V., Goldberg R., 1979. Plant genome organization and its relationship to classical plant genetics. in *Nucleic Acids in Plants I*, T.C. Hall, J.W. Davies eds : 3-39.
- Walling L., Drews G.N., Goldberg R.B., 1986. Transcriptional and post-transcriptional regulation of soybean seed protein mRNA level *Proc Natl Acad Sci USA* : 2123-2127.
- Walton K.E., Styer D., Gruenstein E.I., 1979. Genetic polymorphism in normal human fibroblasts as analysed by two-dimensional polyacrylamide gel electrophoresis. *The Journal of Biological Chemistry* 254(16) : 7951-7960.
- Weeden N.F., Muehlbauer F.J., Ladizinsky G., 1992. Extensive conservation of linkage relationships between pea and lentil genetic maps. *The Journal of Heredity* 83(2) : 123-129.
- Weller J.I., 1987. Mapping and analysis of quantitative trait loci in *Lycopersicon* (tomato) with the aid of genetic markers using approximate maximum likelihood methods. *Heredity* 59 : 413-421.
- Weller J.I., Kashi Y., Soller M., 1990. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *J Dairy Sci* 73 : 2525-2537.
- Weller J.I., Soller M., Brody T., 1988. Linkage analysis of quantitative traits in an interspecific cross of tomato (*Lycopersicon esculentum* \* *Lycopersicon pimpinellifolium*) by means of genetic markers. *Genetics* 118 : 329-339.
- Westoby M., Rice B., 1982. Evolution of the seed plants and inclusive fitness of plant tissues. *Evolution* 36(4) : 713-724.
- White R., Leppert M., Bishop D.T., Barker D., Berkowitz J., Brown C., Callahan P., Holm T., Jerominski L., 1985. Construction of linkage maps with DNA markers for human chromosomes. *Nature* 313 : 101-105.
- Wilcox P.L., Sederoff R.R., Chaparro J., Grattapaglia D., Weir B.S., Zeng Z.B., O'Malley D., 1992. A proposed method for mapping quantitative trait loci in conifer clones using half sib families. Sixth International Conifer Biotechnology Working Group at Radisson Governors Inn, Raleigh, NC.
- Williams G.K., Kubelik A.R., Livak K.J., Rafalski J.A., Tingey S.V., 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* 18 : 6531-6535.
- Xie C.Y., Dancik B.P., Yeh F.C., 1991. Inheritance and linkage of isozymes in *Thuja orientalis*. *The Journal of Heredity* 82 : 329-334.
- Yamamoto N., Kano-Murakami Y., Matsuoka M., Ohashi Y., Tanaka Y., 1988. Nucleotide sequence of a full length cDNA clone of ribulose biphosphate carboxylase small subunit gene from green dark-grown pine (*Pinus thunbergii*) seedling. *Nucleic Acids Research* 16(24) : 11830.
- Zhang W., Smith C., 1992. Computer simulation of marker-assisted selection utilizing linkage disequilibrium. *Theor Appl Genet* 83 : 813-820.
- Zhuchenko A.A., Korol A.B., 1983. Ecological aspects of the recombination problem. *Theor Appl Genet* 64 : 177-185.

- Zitnik G.D., Martin G.M., 1987. Murine "housekeeping" enzyme (genetic locus : *Idh-1*) is regulated in an allele specific manner. *Developmental Genetics* 8 : 135-150.
- Zivy M., Devaux P., Blaisonneau J., Jean R., Thiellement H., 1992. Segregation and linkage studies in microspore-derived double haploid lines of *Hordeum vulgare* L. *Theor Appl Genet* 83 : 919-924.
- Zivy M., Thiellement H., de Vienne D., Hofmann J-P., 1983. Study on nuclear and cytoplasmic genome expression in wheat by two-dimensional gel electrophoresis. 1. First results on 18 alloplasmic lines. *Theor Appl Genet* 66 : 1-7.

## ANNEXE 1: Calcul de l'information de Fisher.

Calculs de  $E(L''(\theta))$  et  $E(L'(\theta))$ .

Soit  $m$  le nombre d'arbres doubles hétérozygotes pour une paire de locus,  $N_i$  gamètes de l'arbre  $i$  ont été étudiés,  $n_i$  gamètes parmi eux sont dans une phase donnée.  $\theta$  représente le taux de recombinaison entre les deux locus, et  $p$  la probabilité pour un arbre d'être dans la phase considérée. On a alors  $L$ , la log vraisemblance de notre échantillon:

$$L \approx \sum_{i=1}^m \ln [p\theta^{n_i}(1-\theta)^{N_i-n_i} + (1-p)(\theta^{N_i-n_i}(1-\theta)^{n_i})]$$

on pose:

$$\gamma_i(\theta) = \theta^{\frac{N_i}{2}} (1-\theta)^{\frac{N_i}{2}}$$

$$\alpha_i(\theta) = \left[ \frac{\theta}{1-\theta} \right]^{n_i - N_i/2}$$

On a alors:

$$\theta^{n_i} (1-\theta)^{N_i-n_i} = \gamma_i(\theta) \times \alpha_i(\theta)$$

$$\theta^{N_i-n_i} (1-\theta)^{n_i} = \gamma_i(\theta) \times \alpha_i^{-1}(\theta)$$

La log vraisemblance s'écrit donc:

$$L \approx \sum_{i=1}^m \ln [p\alpha_i(\theta) + (1-p)\alpha_i^{-1}(\theta)] + \frac{1}{2} N \ln(\theta(1-\theta))$$

avec  $N = \sum_{i=1}^m N_i$

Les dérivées partielles de la log vraisemblance sont alors :

$$\frac{\partial L}{\partial p}(p, \theta) = \sum_{i=1}^m \frac{\alpha_i(\theta) - \alpha_i^{-1}(\theta)}{p\alpha_i(\theta) + (1-p)\alpha_i^{-1}(\theta)}$$

$$\frac{\partial L}{\partial \theta}(p, \theta) = \frac{1}{\theta(1-\theta)} \left[ N \frac{1-2\theta}{2} + \sum_{i=1}^m \left( n_i - \frac{N_i}{2} \right) \frac{p\alpha_i(\theta) - (1-p)\alpha_i^{-1}(\theta)}{p\alpha_i(\theta) + (1-p)\alpha_i^{-1}(\theta)} \right]$$

$$\frac{\partial^2 L}{\partial \theta^2}(p, \theta) = \frac{1}{\theta(1-\theta)} \left[ (2\theta-1) \frac{\partial L}{\partial \theta} - N + \frac{4p(1-p)}{\theta(1-\theta)} \sum_{i=1}^m \frac{(n_i - N_i/2)^2}{(p\alpha_i(\theta) + (1-p)\alpha_i^{-1}(\theta))^2} \right]$$

On en déduit :

$$E_{p_0, \theta_0} \left[ \frac{\partial^2 L}{\partial \theta^2} \right] = \frac{1}{\theta_0(1-\theta_0)} \left[ -N + \frac{4p_0(1-p_0)}{\theta_0(1-\theta_0)} \sum_{i=1}^m E_{p_0, \theta_0} \frac{(n_i - N_i/2)^2}{(p_0\alpha_i(\theta_0) + (1-p_0)\alpha_i^{-1}(\theta_0))^2} \right]$$

avec :

$$E \left[ \left( n - \frac{N}{2} \right)^2 \left( \frac{p\alpha - (1-p)\alpha^{-1}}{p\alpha + (1-p)\alpha^{-1}} \right)^2 \right] = \gamma_i(\theta) \sum_{n < \frac{N}{2}} (n - N/2)^2 \left[ (\alpha_n + \alpha_n^{-1}) \frac{(1-2p)^2 + p(1-p)(\alpha_n - \alpha_n^{-1})^2}{p^2 + p(1-p)(\alpha_n^2 + \alpha_n^{-2}) + (1-p)^2} \right] \binom{N}{n}$$

On calcule alors la limite de  $I(\theta, p)$  (information de Fisher) quand  $\theta$  tend vers  $1/2$ . En  $\theta = 1/2$ ,  $L(p, 1/2)$  ne dépend pas de  $p$  et donc :

$$\frac{\partial^2 L}{\partial p \partial \theta}(p, \theta) \equiv 0$$

$$\frac{\partial^2 L}{\partial p^2}(p, \theta) \equiv 0$$

$\frac{\partial^2 L}{\partial p \partial \theta}(p, \theta)$  et  $\frac{\partial^2 L}{\partial p^2}(p, \theta)$  sont des fonctions continues en  $\theta$  et  $p$  donc :

$$\lim \left( \frac{\partial^2 L}{\partial p \partial \theta}(p, \theta) \right)_{\theta \rightarrow \frac{1}{2}} = 0$$

et

$$\lim \left( \frac{\partial^2 L}{\partial p^2}(p, \theta) \right)_{\theta \rightarrow \frac{1}{2}} = 0$$

donc leurs espérances sont nulles.

D'autre part, quel que soit  $n_i$  :

$$\lim (\alpha_i(\theta))_{\theta \rightarrow \frac{1}{2}} = 1$$

donc

$$\lim \left( E_{\theta, p} \frac{(n_i - N_i/2)^2}{(p\alpha_i(\theta) + (1-p)\alpha_i^{-1}(\theta))^2} \right)_{\theta \rightarrow \frac{1}{2}} = \lim \left( E_{\theta, p} (n_i - N_i/2)^2 \right)_{\theta \rightarrow \frac{1}{2}} = \frac{N_i}{4} \text{ (variance de la binomiale } (N_i, 1/2))$$

Finalement :

$$\lim (I(\theta, p))_{\theta \rightarrow \frac{1}{2}} = \begin{bmatrix} 0 & 0 \\ 0 & 4N - 16p(1-p)N \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 4N(1-2p)^2 \end{bmatrix}$$



ANNEXE 2: Calcul de la puissance du test du  $X^2$ .

Pour deux locus donnés, soit  $m$  le nombre d'arbres double-hétérozygotes,  $N_i$  gamètes ont été étudiés pour chacun d'entre eux, dont  $n_i$  sont dans une phase donnée. Sous l'hypothèse  $H_0$  de l'indépendance entre les deux locus ( $\theta=0.5$ ), la statistique  $S$  :

$$S = \sum_{i=1}^m \frac{(2n_i - N_i)^2}{N_i}$$

peut être approchée par une loi du  $\chi^2$  à  $m$  degrés de liberté (d'espérance  $m$  et de variance  $2m$ ).

Pour tester l'hypothèse  $H_1$  ( $\theta \neq 0.5$ ), on se place sous l'alternative la plus défavorable, qui correspond à  $p=0.5$ . En supposant tous les  $N_i$  égaux à  $N$ , le nombre  $n_i$  de gamètes suit alors une somme de lois binomiales :

$$n \rightarrow \frac{1}{2}B(N, \theta) + \frac{1}{2}B(N, 1-\theta)$$

On calcule l'espérance et la variance de la statistique  $S$  sous  $H_1$ . On pose  $x = (1/2 - \theta)$ , on a alors :

$$E[S] = m[1 + 4x^2(N-1)]$$

$$V[S] = \frac{m}{N^2} (E[2n_i - N_i]^4 - E^2[2n_i - N_i]^2)$$

Si  $m$  et  $N$  sont suffisamment grands :

$$(2n - N) \rightarrow N(0, N(1 + 4x^2(N-1)))$$

donc :

$$E[2n - N]^4 = 3N^2[1 + 4x^2(N-1)]^2$$

$$E^2[2n - N]^2 = N^2[1 + 4x^2(N-1)]^2$$

d'où :

$$V[S] = 2m(1 + 4x^2(N-1))^2$$

Sous les hypothèses  $H_0$  et  $H_1$ , la statistique  $S$  peut alors être approchée par deux lois normales :

Sous l'hypothèse  $H_0$  :  $S \rightarrow N(m, 2m)$

Sous l'hypothèse  $H_1$  :  $S \rightarrow N(m(1+f), 2m(1+f)^2)$

avec :  $f = 4x^2(N-1)$

L'hypothèse  $H_0$  sera rejetée avec un risque  $\alpha$  quand :

$$\frac{S - m}{\sqrt{2m}} > z_\alpha$$

Lorsque la statistique  $S$  dépasse  $z_\alpha$ , on a, sous  $H_1$  :

$$\frac{S - m(1+f)}{\sqrt{2m(1+f)}} > \frac{z_\alpha - mf/\sqrt{2m}}{(1+f)} = z_\beta$$

La probabilité correspondant au seuil  $z_\beta$  nous donne la puissance du test pour différentes valeurs du taux de recombinaison  $\theta$ .

Avec  $m=18$ , et  $N=12$  dans notre expérience, pour  $\alpha=5\%$  ( $z_\alpha=1.65$ ), les valeurs suivantes sont obtenues pour la puissance du test :

$\theta$	0.47	0.45	0.43	0.40	0.38	0.35	0.3
$z_\beta$	1.47	1.19	0.82	0.23	-0.15	-0.66	-1.31
Puissance	7%	12%	20%	41%	56%	75%	90%

**ANNEXE 3 : Relations entre le génotype des arbres aux marqueurs et quelques caractères quantitatifs.**

Le numéro des arbres, les valeurs du caractère pour chaque arbre, classées dans un ordre croissant, ainsi que le génotype des arbres au locus, identifié par son numéro en italique, sont donnés (les allèles p, a, q, v, étant liés respectivement à une présence, une absence, une variation de quantité ou de position). La probabilité d'obtenir une valeur du test F supérieure à la valeur observée est également notée. Les moyennes des caractères (nom du caractère précédé par 'm') relatives aux arbres de même génotype (en nombre na), identifiées par la même lettre, ne diffèrent pas significativement selon un test de Tukey à 5%.

**Caractères mesurés sur arbres âgés**

**Mesure en test de descendance**

**Hauteurs totales mesurées en 1983 et 1987**

Arbre	H83	4406	Arbre	H87	4406	4305	4202
17	552.7	q2q2	10	873.2	q2q2	aa	q2q2
4	554.3	q2q2	4	877.7	q2q2	aa	q2q2
18	564.8	q2q2	7	876.2	q2q2	pp	q1q1
8	565.0	q2a	17	885.3	q2q2	aa	q2q2
3	569.3	q2q2	8	887.4	q2a	aa	q2q2
10	574.7	q2q2	13	891.2	q2a	pp	q1q1
12	574.9	q2q2	11	896.2	q2q2	pp	q1q1
15	576.1	q2q2	15	904.1	q2q2	pa	q1q2
13	576.3	q2a	3	906.8	q2q2	pa	q1q2
7	577.8	q2q2	18	908.8	q2q2	pp	q1q1
11	590.7	q2q2	12	912.5	q2q2	pp	q1q1
14	591.8	q1q2	14	919.2	q1q2	pa	q1q2
5	614.0	q1q2	6	940.9	q1q1	pa	q1q2
6	617.6	q1q1	5	948.5	q1q2	pa	q1q2
2	624.3	q1q1	2	950.3	q1q1	pa	q1q2

Test F : P=0.0006 P=0.0017 P=0.0016

**Tests de Tukey (5%)**

4406	na	mH83	mH87	4305	4202	na	mH87
q2q2	9	570.61a	893.44a	aa	q2q2	4	880.54a
q2a	2	570.64a	889.35a	pp	q1q1	5	897.30a
q1q2	2	602.93b	933.87b	pa	q1q2	6	928.31b
q1q1	2	620.95b	945.58b				

**Ecart à la verticalité mesuré en 1983**

Arbre	ECV83	3206
4	7.7	q1q2
12	8.5	q1q2
2	8.6	q1q2
10	9.0	q1q2
5	9.2	q1q2
13	9.8	q2q2
15	9.9	q2q2
17	10.0	q2q2
3	10.3	q1q2
11	10.4	q2q2
8	10.6	q2q2
6	10.8	q2q2
14	10.8	q2q2
7	11.4	q2q2
18	12.8	q2q2

**Test de Tukey**

3206	na	mECV83
q1q2	6	8.9a
q2q2	9	10.7b

Test F : P=0.0022

## Mesures en parc à clone

## Somme des pousses secondaires

Arbre	SPII	3111	Test de Tukey		
6	-36.5	pa	3111	na	mSPII
11	-30.4	pa			
17	-29.1	pa	pa	5	-23.8a
5	-19.5	pp			
12	-19.0	pa	pp	11	10.1b
4	-5.8	pp			
14	-4.0	pa			
10	-1.8	pp			
16	4.2	pp			
13	5.9	pp			
18	5.9	pp			
15	12.2	pp			
8	13.2	pp			
1	23.9	pp			
7	36.1	pp			
3	37.2	pp			
Test F		P=0.0015			

## Caractères mesurés sur jeunes plants en conditions expérimentales

## Hauteur mesurée après un stress hydrique

Arbre	HST	2428	Test de Tukey		
6	144.2	v1v2	2428	na	mHST
15	155.6	v1v2			
1	171.4	v1v2	v1v2	4	162.93a
12	180.5	v1v2			
7	185.7	v1v1	v1v1	12	215.96b
16	194.4	v1v1			
9	199.4	v1v1			
11	204.2	v1v1			
14	205.0	v1v1			
8	211.4	v1v1			
13	216.1	v1v1			
10	221.2	v1v1			
5	230.0	v1v1			
4	232.0	v1v1			
3	243.3	v1v1			
2	248.8	v1v1			
Test F		P=0.0003			

## Mesures en phytotron

## Première période de croissance

## Mesures morphologiques

## Nombre de branches

Arbre	NB	3227	2322	Test de Tukey			
2	4.52	aa	q2q2	3227	2322	na	mNB
1	4.62	aa	q2q2				
12	4.77	aa	q2q2	aa	q2q2	7	4.85a
6	4.78	aa	q2q2				
4	4.87	aa	q2q2	pa	q1q2	2	6.01b
5	5.15	aa	q2q2				
11	5.21	aa	q2q2				
14	5.60	pa	q1q2				
9	6.42	pa	q1q2				
Test F		P=0.0028					

## Mesures de poids

## Poids des branches

Arbre	PB	2421	3418	Tests de Tukey			
1	73.52	q2q2	pa	2421	na	mPB	
4	78.32	q2q2	pa				
5	103.72	q2q2	pa	q2q2	6	104.14a	
6	107.67	q2q2	pa				
12	108.82	q2q2	pa	q1q2	3	196.90b	
2	152.77	q2q2	pp				
11	177.59	q1q2	aa	3418	pa	5	94.42a
14	199.85	q1q2	aa				
9	213.25	q1q2	aa	pp	1	152.78b	
Test F		P=0.0014					

## Deuxième période de croissance

## Mesures morphologiques

## Nombres de branches sur le premier cycle

Arbre	NB1	1215	Test de Tukey		
1	1.09	q2q2	1215	na	mNB1
12	1.33	q2q2			
14	1.59	q2q2			
2	1.63	q1q1			
9	1.77	q2q2	q2q2	5	1.5146a
7	1.79	q2q2	q1q1	2	1.7867a
6	1.95	q1q1	q1q2	4	2.7893b
5	2.35	q1q2			
4	2.67	q1q2			
11	2.69	q1q2			
8	3.44	q1q2			

Test F P=0.0023

## Nombre d'aiguilles secondaires sur le deuxième cycle

Arbre	NAII2	4101	Test de Tukey		
12	23.17	v1v2	4101	na	mNAII2
7	24.23	v1v2			
8	31.23	v1v1			
5	31.53	v1v1			
14	32.68	v1v1	v1v2	2	23.70a
4	34.33	v1v1	v1v1	9	35.01b
6	34.48	v1v1			
11	35.81	v1v1			
1	36.39	v1v1			
9	37.97	v1v1			
2	40.68	v1v1			

Test F P=0.0008

## Longueur moyenne d'un entre-noeud sur le troisième cycle

Arbre	MSUL3	1410	2210	4211	Tests de Tukey			
5	0.594	q1q1	q1q1	pp	1410	2210	na	mMSUL3
2	0.611	q1q1	q1q1	pp	q1q1	q1q1	8	0.660a
9	0.613	q1q1	q1q1	pp	q1q2	q1q2	2	0.879b
11	0.617	q1q1	q1q1	pp				
14	0.647	q1q1	q1q1	pp				
6	0.687	q1q1	q1q1	pa	4211			
12	0.706	q1q1	q1q1	pa	pp	5	0.616a	
4	0.808	q1q1	q1q1	pa	pa	5	0.792b	
8	0.877	q1q2	q1q2	pa				
1	0.881	q1q2	q1q2	pa				

Tests F P=0.0032 P=0.0031

## Mesures de poids

## Poids des aiguilles secondaires du premier cycle

Arbre	PAII1	2322	3227	2421	3418	Tests de Tukey			
6	326.14	q2q2	aa	q2q2	pa	2322	3227	na	mPAII1
12	367.63	q2q2	aa	q2q2	pa	q2q2	aa	9	402.19a
7	384.91	q2q2	aa	q2q2	pp	q1q2	pa	2	556.53b
5	395.56	q2q2	aa	q2q2	pa				
2	400.49	q2q2	aa	q2q2	pp	2421			
8	415.84	q2q2	aa	q2q2	pa	q2q2	8	392.44a	
1	421.15	q2q2	aa	q2q2	pa	q1q2	3	531.10b	
4	427.76	q2q2	aa	q2q2	pa				
11	480.23	q2q2	aa	q1q2	aa	3418			
14	528.28	q1q2	pa	q1q2	aa	pa	6	392.35a	
9	584.77	q1q2	pa	q1q2	aa	pp	2	392.7a	
						aa	3	531.10b	

Tests F P=0.0012 P=0.0005 P=0.0033

## Poids des aiguilles primaires du deuxième cycle

Arbre	PAI2	3309	Test de Tukey		
4	12.2247	pp			
7	12.7440	pp			
8	13.9514	pp	3309	na	mPAI2
14	14.2049	pp			
1	14.6116	pp	pp	7	14.140a
5	14.6161	pp	pa	4	18.881b
9	16.6295	pp			
11	17.1563	pa			
2	18.1081	pa			
12	18.4063	pa			
6	21.8527	pa			

Test F : P=0.0014

## Poids des branches du troisième cycle

Arbre	PB3	2215	Test de Tukey		
7	4.84	pp			
1	8.12	pp			
12	8.26	pp	2215	na	mPB3
14	11.94	aa			
11	13.61	aa	pp	3	7.07a
5	13.86	aa	aa	7	14.33b
8	13.94	aa	pa	1	21.77c
6	14.99	aa			
4	15.00	aa			
2	15.94	aa			
9	21.77	pa			

Test F : P=0.0001