HAL
open science

# Caractérisation et variation d'une famille multigénique, l'ADN ribosomique 5S nucléaire, chez quatre espèces forestières des genres larix M. (Pinaceae) et Quercus L. (Fagaceae)

Jean-François Trontin

**To cite this version:**

Jean-François Trontin. Caractérisation et variation d'une famille multigénique, l'ADN ribosomique 5S nucléaire, chez quatre espèces forestières des genres larix M. (Pinaceae) et Quercus L. (Fagaceae). Sylviculture, foresterie. Université Henri Poincaré - Nancy 1, 2000. Français. NNT : 2000NAN10186 . tel-01748195

**HAL Id: tel-01748195**

**https://hal.univ-lorraine.fr/tel-01748195**

Submitted on 29 Mar 2018

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

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

UNIVERSITÉ DE LORRAINE

AVERTISSEMENT

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

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

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

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

U.F.R. Sciences & Techniques Biologiques
Ecole Doctorale
"Ressources, Procédés, Produits, Environnement"

# Thèse

présentée pour l'obtention du titre de

## Docteur de l'Université Henri Poincaré, Nancy I en Biologie Forestière

par

**Jean-François TRONTIN**

# Caractérisation et variation d'une famille multigénique, l'ADN ribosomique 5S nucléaire, chez quatre espèces forestières des genres *Larix* M. (*Pinaceae*) et *Quercus* L. (*Fagaceae*).

Soutenue le 20 décembre 2000

Membres du jury :

| | | |
|---|---|---|
| Rapporteurs : | M. **André BERVILLE** | Directeur de Recherches, Centre INRA de Montpellier |
| | M. **André RODE** | Professeur, Université Pierre et Marie Curie, Paris VI |
| Examinateurs : | M. **Jean-Michel FAVRE** | Professeur, Université Henri Poincaré, Nancy I |
| | M. **Antoine KREMER** | Directeur de Recherches, Centre INRA de Bordeaux |
| | M. **Francis MARTIN** | Directeur de Recherches, Centre INRA de Nancy |
| | M. **Marc PÂQUES** | Directeur de Recherches, Centre AFOCEL de Nangis |
| Invité : | M. **Hubert DULIEU** | Professeur Honoraire, Université de Bourgogne |

**Laboratoire de Biologie Forestière, Unité Associée INRA 977**
**Faculté des Sciences - 54506 Vandoeuvre-lès-Nancy**

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

***Il y a deux choses que l'expérience doit apprendre :***
***la première, c'est qu'il faut beaucoup corriger,***
***la seconde, c'est qu'il ne faut pas trop corriger.***

***Eugène Delacroix***

***Remerciements***

*Ce travail a été réalisé avec le support du Ministère de l'Enseignement Supérieur et de la Recherche au sein du Laboratoire de Biologie Forestière de l'Université Henri Poincaré, Nancy I, sous la direction de Monsieur Jean-Michel Favre. Je tiens simplement à lui exprimer ici toute ma reconnaissance pour sa confiance, son soutien permanent et sa rigueur constructive.*

*Messieurs André Bervillé et André Rode ont accepté la lourde tâche d'être rapporteurs de ce mémoire. Je les remercie vivement de cette attention et de leur regard critique sur ces travaux.*

*Messieurs Hubert Dulieu, Antoine Kremer, Francis Martin et Marc Pâques ont bien voulu se joindre au jury de thèse malgré leurs nombreuses occupations. Je suis honoré de leur participation. Qu'ils reçoivent ici mes remerciements.*

*Monsieur Antoine Kremer a témoigné à plusieurs reprises de son intérêt pour nos recherches. Je lui fait part de toute ma gratitude pour ses conseils précieux au cours de discussions fructueuses et pour son appui.*

*Ma rencontre avec Monsieur Hubert Dulieu a été une étape décisive de mon parcours universitaire. Je le remercie de m'avoir communiqué son enthousiasme pour la recherche. Je le remercie également pour son soutien.*

*La patience de Monsieur Marc Pâques est sans égale. Qu'il soit remercié de la confiance qu'il m'a accordée durant la rédaction de la partie finale de ce mémoire.*

*Je souhaite remercier tout particulièrement et chaleureusement Madame Marie-France Arnould. Le travail quotidien à ses côtés a été source de rigueur, méthode, confiance, persévérance et finalement grande satisfaction. Merci mille fois Marie-France. Tes confitures sont un sacré carburant ! Le secret des "manips" réussies !*

*Je remercie également Damien Chanteranne, Catherine Grandemange, Valérie Légué et Jean-Pierre Leclerc pour avoir pris part activement à un moment ou à un autre à ces recherches sur l'ADNr 5S.*

*Merci à Rémy Petit, Catherine Bodénès et Valérie Le Corre pour leur accueil à la station de recherche de Pierroton et leurs conseil au tout début de ce travail.*

*Merci également à tous les permanents et stagiaires qui ont fait du laboratoire de Biologie Forestière un lieu de travail chaleureux. Je pense en particulier à Jean Dexheimer, Jean-Claude Pargney, Paul Barnola, Etienne Kiffer, Sylvain Jeandroz, Isabel Le Disquet, Carole Kaupp, Bachir Chaba, Patricia Genêt, Anas Mansour, Bénédicte Sirot, Badr Alaoui-Sossé, Virginie George, Axelle André, Delphine Grivet, Catherine Grandemange, Marie-Noelle Jordy, Virginie Acheré ...*

*Elizabeth De Faÿ a fixé la barre très très haute dans l'échelle de la persévérance, de l'investissement et de la passion pour la recherche. Merci Elizabeth de m'avoir donné cette référence indispensable.*

*Je mesure ici toute la portée de votre amitié sincère au cours de ces longues années communes de thèse. Simplement merci à toi Rédouane et merci à toi Jacques.*

*Un grand merci également à toute ma famille et en particulier à mes parents qui m'ont encouragé dans cette voie.*

*Enfin, ce travail est aussi celui d'Anne-marie. De la paillasse à la rédaction, du soir au matin, elle a participé activement à toutes les étapes de l'élaboration de ce document. Ces quelques mots ne s'écrivent pas mais je les pense très fort.*

# ABREVIATIONS

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

**ADN** : Acide DésoxyriboNucléique

**AFLP** : Amplified Fragment lengh polymorphism

**ARN** : Acide RiboNucléique

**cccDNA** : covalently closed circular DNA

**DGGE** : Denaturing Gradient Gel Electrophoresis

**EDTA** : Ethylen Diamine Tetra acetic Acid

**ICR** : Internal Control Region

**IGS** : InterGenic Spacer

**ITS** : Internal Transcribed Spacer

**kpb** : kilopaire de bases ($10^3$ pb)

**LINE** : Long Interspersed Nuclear Elements

**LTR** : Long Terminal Repeats

**Ma** : Million d'années

**MITE** : Miniature Inverted-repeat Transposable Elements

**Mpb** : Mégapaires de bases ($10^6$ pb)

**NOR :** Nucleolar Organizer Region

**nt** : nucléotide

**NTS** : Non-Transcribed Spacer

**ORF** : Open Reading Frame

**pb** : paire de bases

**PCR** : Polymerase Chain Reaction

**Pol** : Polymérase

**QTL** : Quantitative Trait Loci

**RAPD** : Random Amplified Polymorphic DNA

**RFLP** : Restriction Fragment Length Polymorphism

**SCAR** : Sequence Characterized Amplified Region

**SINE** : Short Interspersed Nuclear Elements

**SSCP** : Single Strand Conformation Polymorphism

**SSR** : Simple (Short) Sequence Repeats

**STR** : Short Tandem Repeats

**TGGE** : Temperature Gradient Gel Eclectophoresis

**VNTR** : Variable Number of Tandem Repeats

# ABREVIATION DES BASES NUCLEOTIDIQUES

**Code IUPAC (International Union of Pure and Applied Chemistry)**

**Bases nucléotidiques**

**A** : Adénine

**C** : Cytosine

**G** : Guanine

**T** : thymine

**U** : Uracile

**Bases nucléotidiques ambiguës (T et U sont équivalents)**

**R** : A ou G (purines)

**Y** : C ou T (pyrimidines)

**M** : A ou C

**W** : A ou T

**S** : C ou G

**K** : G ou T

**B** : C, G ou T

**D** : A, G ou T

**H** : A, C ou T

**V** : A, C ou G

**N** = X : A, C, G ou T

# SOMMAIRE

**INTRODUCTION** ..... **10**

***BIBLIOGRAPHIE*** ..... ***13***

***SYNTHESE BIBLIOGRAPHIQUE*** ..... ***5***

**CHAPITRE I : IMPORTANCE, ORGANISATION ET EVOLUTION DES SEQUENCES REPETEES.** ..... **15**

**I- Contexte général : plasticité du génome et séquences répétées** ..... **15**
1- Variations de la taille du génome ..... 15
2- Signification des variations de la taille du génome ..... 16
a- Valeur adaptative des variations ..... 17
b- Le paradoxe de la valeur C ..... 18
3- Importance des séquences répétées dans les génomes eucaryotes ..... 19
a- Relation entre taille du génome et proportion de séquences répétées ..... 19
b- Contribution de l'amplification de séquences à l'évolution de la taille du génome ..... 20
4- Fonction des familles de séquences répétées ..... 20
a- Familles de l'ADN non codant ..... 20
b- Familles multigéniques ..... 23
5- Conclusion ..... 25

**II- Architecture, formation et évolution de la fraction répétée des génomes nucléaires** ..... **26**
1- Les familles de séquences répétées et dispersées ..... 27
a- Familles de transposons ..... 29
b- Familles de rétroéléments ..... 30
α- Rétrotransposons ..... 30
- Rétrotransposons de la classe I (ou viraux) ..... 31
- Rétrotransposons de la classe II (rétroposons) ..... 32
β- Rétroséquences ..... 32
c- Autres familles de séquences répétées et dispersées ..... 34
2- Les familles de séquences répétées et groupées en tandem ..... 34
a- Séries de l'ADN non codant ou de fonction inconnue ..... 35
α- Microsatellites et minisatellites ..... 35
β- Satellites ..... 36
b- Séries multigéniques ..... 37
α- Gènes multimériques ..... 37
β- Familles multigéniques à répétition de variants ..... 38
χ- familles de gènes des ARNt et ARNsn ..... 38
δ- Famille de gènes des histones ..... 39
ε- Télomères ..... 40
ϕ- Familles de gènes des ARNr ..... 41
- Famille des gènes des ARNr 18S, 5,8S et 25-28S (ADNr) ..... 41
- Famille(s) de gènes des ARNr 5S (ADNr 5S) ..... 43
c- Mécanismes de formation et d'amplification des répétitions en tandem ..... 45
α- Glissement de brins d'ADN ..... 46
β- Crossing-over inégal ..... 48
χ- Amplification par réplication en cerceau ou par réinitiations multiples ..... 50
3- Evolution des familles de séquences répétées ..... 53
a- Mise en évidence du processus d'évolution concertée ..... 53
b- Importance et fonction éventuelle de l'évolution concertée ..... 54
c- Mécanismes de l'évolution concertée ..... 55
α- Homogénéisations par crossing-over inégal ..... 56
β- Homogénéisations par conversion génique ..... 57
χ- Modèle d'évolution concertée des séries multigéniques ..... 59
d- Conséquences de l'évolution concertée ..... 61
4- Conclusion ..... 62

***BIBLIOGRAPHIE*** ..... ***64***

**CHAPITRE II : L'ADNr 5S NUCLEAIRE CHEZ LES PLANTES SUPERIEURES** ..... **73**

**I- Fonction et expression de l'ADNr 5S** ..... **73**
1- Signaux de contrôle de la transcription par la Pol III ..... 74
a- Séquences d'initiation de la transcription ..... 74
α- La région de contrôle interne (ICR) ..... 75
β- Les séquences régulatrices de l'espaceur ..... 76
b- Séquences de terminaison de la transcription ..... 77
2- Modulation de l'activité de la Pol III par des facteurs de transcription ..... 79

3- Conséquence : conservation de la structure des ARNr 5S chez les plantes ..... 79

**II- Organisation génomique de l'ADNr 5S ..... 81**

1- Localisation génomique ..... 81

a- Localisation chromosomique ..... 81

b- Cartographie génétique ..... 85

2- Méthylation de l'ADNr 5S ..... 86

**III- Evolution de l'ADNr 5S ..... 88**

1- Variations du nombre d'unités ..... 89

2- Variations de longueur des unités ..... 90

a- Variations de longueur dans les gènes ..... 93

b- Variations de longueur dans les espaceurs ..... 93

α- Les bordures 5' et 3' ..... 94

β- La région centrale ..... 94

3- Variations de séquence ..... 97

a- Variation intragénomique ..... 95

α- Variation "interloci" VS variation "intralocus" ..... 96

β- Variation dans les gènes VS variation dans les espaceurs ..... 98

χ- Pseudogènes ..... 99

b- Variation intraspécifique (entre les individus, populations, *etc.*) ..... 99

c- Variations interspécifiques ..... 100

α- Variations du gène ..... 100

β- Variations de l'espaceur ..... 101

4- Evolution concertée et sélection ..... 104

a- Interactions entre les loci ..... 104

b- Homogénéisation intralocus ..... 105

c- Sélection naturelle ..... 105

**IV- Conclusion ..... 106**

***BIBLIOGRAPHIE ..... 107***

***PARTIE EXPERIMENTALE ..... 73***

**PRESENTATION GENERALE ..... 112**

**I- Espèces étudiées ..... 112**

**II- Objectifs ..... 113**

1- Structure des unités ..... 114

2- Organisation des unités ..... 114

3- Polymorphisme et évolution des unités ..... 114

**III- Stratégie d'accès à l'ADNr 5S nucléaire des plantes supérieures ..... 115**

***BIBLIOGRAPHIE ..... 116***

**CHAPITRE III : ETUDE DE L'ADNR 5S CHEZ LES CHENES PEDONCULE ET SESSILE ..... 118**

**I- Les chênes ..... 118**

1- Importance, origine et distribution ..... 118

2- Classification générale du genre *Quercus* ..... 119

3- Les chênes pédonculé et sessile ..... 121

**II- Caractérisation et variation des unités d'ADNr 5S chez les chênes pédonculé et sessile ..... 126**

CHARACTERIZATION AND VARIATION OF NUCLEAR 5S RIBOSOMAL DNA UNITS IN TWO CLOSELY RELATED EUROPEAN OAKS, *QUERCUS ROBUR* L. (PEDUNCULATE OAK) AND *QUERCUS PETRAEA* (MATT.) LIEBL. (SESSILE OAK) ..... 127

1- Abstract ..... 127

2- Introduction ..... 128

3- Materials and Methods ..... 130

a- Plant material and DNA isolation ..... 130

b- 5S rDNA amplification and electrophoresis ..... 132

c- Sample relative diversity ..... 133

d- Cloning and Sequencing ..... 133

e- Sequence analysis ..... 133

f- Southern hybridization and copy number estimation ..... 134

4- Results ..... 134

a- Characterization of 5S rDNA repeat in *Quercus robur* and *Quercus petraea* ..... 134

b- 5S rRNA gene variations ..... 136

c- Spacer variations ..... 137

α- Sample diversity ..... 137

β- Nucleotide variations ..... 140

χ- Phylogenetic comparison of major 5S rDNA spacer classes (*a–g*) ..... 143

5- Discussion ..... 149
a- *Quercus* 5S rDNA structure and organization ..... 149
b- 5S rDNA molecular diversity in pedunculate and sessile oak ..... 149
c- The low differentiation of pedunculate and sessile oak at the 5S rDNA locus ..... 151
6- Conclusion : rappel des principaux résultats obtenus ..... 153

**III- Héritabilité et localisation génétique des unités d'ADNr 5S nucléaire chez *Quercus robur* L. ..... 158**
1- Introduction ..... 158
2- Matériel et méthodes ..... 159
a- Matériel végétal ..... 159
b- Amplification des espaceurs de l'ADNr 5S ..... 160
c- Séparation et détection des produits de l'amplification ..... 160
3- Résultats ..... 162
a- Phénotypes des arbres 3P et A4 ..... 162
b- Phénotypes des hybrides F1 ..... 162
c- Analyse génétique des marqueurs à hérédité mendélienne **h, a, k, j** (3P) et *i, b, d* (A4) ..... 163
α- Génotype du parent femelle 3P ..... 163
β- Génotype du parent mâle A4 ..... 164
χ- Génotypes proposés pour les descendants ..... 164
δ- Cartographie génétique du locus 5SA ..... 165
d- Cas des marqueurs à hérédité non mendélienne **f** (3P) et *g* (A4) ..... 166
4- Discussion ..... 166

***BIBLIOGRAPHIE* ..... *173***

**CHAPITRE IV : ETUDE DE L'ADNR 5S CHEZ LES MELEZES D'EUROPE ET DU JAPON ..... 178**

**I- Les mélèzes ..... 178**
1- Importance, origine et distribution ..... 178
2- Classification du genre *Larix* ..... 181
3- Les mélèzes d'Europe et du Japon ..... 182

**II- Structure et organisation des unités d'ADNr 5S chez les mélèzes d'Europe et du Japon ..... 188**

A- TWO HIGHLY DIVERGENT 5S RDNA UNIT SIZE CLASSES OCCUR IN COMPOSITE TANDEM ARRAY IN EUROPEAN LARCH (*LARIX DECIDUA* MILL.) AND JAPANESE LARCH (*LARIX KAEMPFERI* (LAMB.) CARR.) ..... 190

1- Abstract ..... 190
2- Introduction ..... 191
3- Materials and methods ..... 193
a- Plant material and DNA isolation ..... 193
b- DNA amplification, electrophoresis and cloning ..... 193
c- Sequence alignment and analysis ..... 194
d- Hybridization and copy number estimation ..... 194
4- Results ..... 195
a- 5S gene ..... 195
b- 5S rDNA spacer ..... 197
α- Amplification ..... 197
β- Sequence analysis in *Larix decidua* ..... 197
χ- Sequence comparison with higher plant 5S rDNA spacers ..... 199
c- 5S rDNA units organization ..... 201
5- Discussion ..... 203
a- Functional organization ..... 203
b- The two highly divergent unit size classes in larch ..... 203
c- The composite tandem array ..... 204
6- Conclusion ..... 206
a- Rappel des principaux résultats obtenus ..... 206
b- Analyse comparative des données récemment publiées chez les gymnospermes ..... 209

B- ETUDE PRELIMINAIRE D'UN MOTIF CONSERVE (5SMSCR1) AU SEIN DE LA REGION CENTRALE DES ESPACEURS DE L'ADNR 5S DES MELEZES ET D'AUTRES GYMNOSPERMES ..... 214

1- Introduction ..... 214
2- Matériel et méthodes ..... 215
a- Matériel végétal et extraction de l'ADN génomique ..... 215
b- Amplification ..... 215
α- Sélection des amorces ..... 215
β- Conditions d'amplification ..... 216
c- Electrophorèse et estimation de la taille des fragments amplifiés ..... 217
d- Recherche de similarités ..... 217
3- Résultats et discussion ..... 218
a- Fonction du motif 5SmsCR1 ..... 218

b- Recherche du motif 5SmsCR1 chez divers gymnospermes ..... 221
α- Les mélèzes ..... 221
- Les mélèzes eurasiens ..... 221
- Les mélèzes américains ..... 223
β- Les épicéas ..... 225
χ- Les pins ..... 226
δ- *Abies alba* et *Ginkgo biloba* ..... 229
4- Conclusions ..... 229
a- Organisation des unités d'ADNr 5S ..... 229
b- Présence du motif 5SmsCR1 ..... 229

**III- Héritabilité, diversité et divergence nucléotidique des unités d'ADNr 5S chez les mélèzes d'Europe (*Larix decidua* Mill.) et du Japon [*Larix kaempferi* (Lamb.) Carr.] ..... 231**
1- Introduction ..... 231
2- Matériel et méthodes ..... 232
a- Matériel végétal ..... 232
b- Extraction d'ADN génomique, PCR et électrophorèse ..... 233
c- Clonage et séquençage ..... 233
d- Alignement et analyse des séquences ..... 234
3- Résultats ..... 235
a- Héritabilité des unités d'ADNr 5S ..... 235
α- phénotypes des arbres 107 et 3076 ..... 235
β- Phénotypes des hybrides F1 et ségrégation des fragments parentaux ..... 236
χ- Génotypes proposés pour les membres de la famille 179 ..... 238
b- Diversité nucléotidique et divergence au sein des unités d'ADNr 5S ..... 238
α- Variation au sein des gènes 5S ..... 239
β- Variation au sein des espaceurs ..... 240
- Alignements et analyses phylogénétiques ..... 240
- Diversité nucléotidique et différences nucléotidiques interspécifiques ..... 244
4- Discussion ..... 245
a- Le locus 5SA et la composition hétérogène des séries d'unités d'ADNr 5S ..... 247
b- La présence de pseudogènes au sein des séries de l'ADNr 5S ..... 250
c- La faible diversité des espaceurs longs par rapport à celle des espaceurs courts ..... 252
d- L'absence de différenciation entre mélèze d'Europe et mélèze du Japon au locus 5SA ..... 255
***BIBLIOGRAPHIE* ..... *256***

**BILAN GENERAL ET PERSPECTIVES ..... 261**

**I- Structure des unités d'ADNr 5S ..... 261**
1- Taille et séquence nucléotidique des gènes 5S et des espaceurs ..... 261
2- Nature et topographie des variations ..... 263
3- Mécanismes de l'évolution structurale ..... 264
4- Aspects fonctionnels ..... 266

**II- Organisation des unités d'ADNr 5S ..... 268**
1- Nombre d'unités ..... 268
2- Structure et composition des séries ..... 269
3- Héritabilité des variants alléliques, nombre de loci et cartographie génétique ..... 271

**III- Polymorphisme et évolution des unités d'ADNr 5S ..... 272**
1- Diversité moléculaire intraspécifique et différenciation interspécifique ..... 272
2- Intensité de l'évolution concertée et conséquences ..... 276
3- Phylogenèse des variants alléliques - Paralogie et orthologie ..... 277
4- Evaluation du niveau d'information taxinomique fourni par l'ADNr 5S ..... 278
***BIBLIOGRAPHIE* ..... *283***

***ANNEXE* ..... *286***

# INTRODUCTION

L'application des techniques de biologie moléculaire au génome des plantes offre de multiples possibilités pour améliorer la quantité et la qualité des productions alimentaires, agricoles et forestières. Ces techniques sont aussi très utiles pour la réalisation de projets d'évaluation, de conservation, de gestion et de diffusion des ressources génétiques (Henry 1997). Identifier une plante, estimer la variation génétique à différentes échelles taxinomiques, sélectionner, améliorer ou créer un génotype sont en effet autant d'activités qui peuvent être assistées par l'utilisation de marqueurs moléculaires. Déterminer ou confirmer l'identité d'une plante, d'une variété ou d'une espèce est bien souvent la première étape de son étude et de son intégration dans des programmes d'amélioration. Le développement d'outils d'identification fiables exige cependant de connaître au préalable les relations génétiques au sein des espèces (génétique des populations) ou entre les espèces (analyse phylogénétique), particulièrement dans les groupes méconnus ou complexes. L'objectif de ce type d'étude est de disposer de marqueurs moléculaires permettant d'évaluer le polymorphisme et donc les distances génétiques relatives ou la différenciation existant entre différents taxa. Les méthodes mises en oeuvre en génétique des populations s'appuient généralement sur la technique de PCR (Polymerase Chain Reaction, Saiki *et al.* 1988) qui permet théoriquement d'accéder à l'ensemble d'un génome par amplification sélective *in vitro*. Par rapport aux méthodes plus classiques basées généralement sur l'hybridation de sondes marquées comme l'analyse RFLP (Restriction Fragment Length Polymorphism) de l'ADN (Acide DésoxyriboNucléique) génomique, la rapidité, la simplicité et la sensibilité importante de la PCR ont permis de réduire considérablement le temps et les efforts nécessaires pour mettre en évidence des différences génétiques. De plus, le potentiel discriminant de la PCR peut être renforcé par l'utilisation conjointe des endonucléases de restriction (AFLP, PCR-RFLP) ou de méthodes d'électrophorèse très résolutives (gels d'acrylamide non dénaturants, SSCP, TGGE, DGGE, etc.). Il est donc rarement indispensable d'avoir massivement recours à l'analyse ultime des différences nucléotidiques par séquençage. Celle-ci n'est réellement justifiée que dans l'étude des relations phylogénétiques entre taxa éloignés qui nécessite la comparaison de segments d'ADN relativement conservés.

Le choix de la méthode et/ou de la cible d'étude dépend surtout du niveau d'information recherché (**Tableau 1**) et exige donc une certaine flexibilité. Le système sélectionné ne doit pas être trop sensible (hypervariable) car il est alors difficile d'identifier des relations (similarités) entre taxa et donc d'interpréter les résultats obtenus. Trop peu variable en revanche, il ne procure pas suffisamment de marqueurs informatifs et l'analyse reste peu résolutive. Ainsi, les techniques d'amplification aléatoire de fragments d'un génome comme la RAPD (Random Amplified Polymorphic DNA) sont très performantes pour caractériser des variétés ou même différents génotypes au sein d'un cultivar, tandis que l'analyse des séquences de gènes très conservés permettent seulement de retracer les grandes subdivisions du monde vivant (familles, ordres, classes, etc.).

**Tableau 1.** Quelques méthodes moléculaires utilisées pour différents niveaux de distinction génétique.

| | Génétique des populations → Analyse phylogénétique | | | | | |
|---|---|---|---|---|---|---|
| | Individus | Variétés | Espèces | Genres | Familles | Niveaux supérieurs |
| **PCR** | | | | | | |
| Microsatellites | ** | *** | * | | | |
| RAPD | ** | *** | ** | * | | |
| **RFLP** | | | | | | |
| Génome nucléaire | ** | *** | ** | * | | |
| Génome mitochondrial | | * | ** | ** | | |
| Génome chloroplastique | | ** | *** | ** | | |
| **Analyse de séquences** | | | | | | |
| ITS | | ** | *** | ** | * | |
| Gènes ribosomiques (18S et 26S) | | | * | ** | *** | *** |
| Gènes chloroplastiques | | | | * | ** | *** |
| Gènes de protéines peu conservées | * | ** | *** | ** | * | |
| Gènes de protéines très conservées | | | * | ** | *** | *** |

D'après Henry (1997).
* Le nombre d'astérisques indique une utilisation croissante de la méthode.

Le génome nucléaire des plantes ($1{,}6.10^8$-$6{,}2.10^{11}$ pb, Cavalier-Smith 1985) est beaucoup plus grand que les génomes chloroplastique ($1{,}2.10^5$-$2.10^5$ pb, Griffiths *et al.* 1996) et mitochondrial ($2.10^5$-$2{,}5.10^6$ pb, Lejeune *et al.* 1988) et il offre donc de multiples possibilités pour étudier les variations génétiques. Il comporte en particulier une part importante de séquences répétées (souvent plus de la moitié) qui évoluent plus rapidement que les séquences uniques (Flavell *et al.* 1983, Röder *et al.* 1992), probablement en raison de la présence d'un ou plusieurs niveaux de variation supplémentaires (variations interloci et/ou intrachromosomique, **Figure 1**). Parmi les familles de séquences répétées, les gènes codant pour les ARN (Acide riboNucléique) ribosomiques (ARNr) présentent diverses propriétés particulièrement attrayantes et constituent probablement la source de marqueurs nucléaires la plus utilisée à ce jour (Zimmer *et al.* 1989, Henry 1997). Les raisons de cette situation sont multiples. Tout d'abord, leur universalité (liée à leur fonction importante dans la biogenèse des ribosomes) autorise les comparaisons entre taxa, même très éloignés. Ensuite, le nombre élevé de copies et l'organisation en tandem de ces familles facilitent leur détection, aussi bien par les techniques de PCR que par les méthodes RFLP. Enfin, l'association dans les unités de répétition de séquences codantes conservées (les gènes) et de régions beaucoup plus variables (les espaceurs) permet d'envisager l'utilisation de ces loci dans des contextes très différents, de l'analyse phylogénétique à l'identification variétale (**Tableau 1**). Par exemple, l'unité de transcription des ARNr 18S, 5,8S et 26S comporte trois gènes séparés par deux espaceurs transcrits mais non codants (ITS : Internal Transcribed Spacer), ces unités de transcription alternant par ailleurs avec un espaceur intergénique non transcrit (IGS : InterGenic Spacer). L'analyse des séquences des gènes des ARNr 18S et 26S a permis d'étudier l'évolution des plantes au niveau des familles et des groupes taxinomiques supérieurs (Solignac *et al.* 1991, Hamby et Zimmer 1991). Le polymorphisme des ITS

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

est en revanche plutôt adapté à l'analyse des relations entre les espèces (Baldwin 1992). L'IGS est, quant à lui, surtout utile dans les études de microévolution car il est très variable au sein et entre les populations d'une espèce (Schaal et Learn 1988).

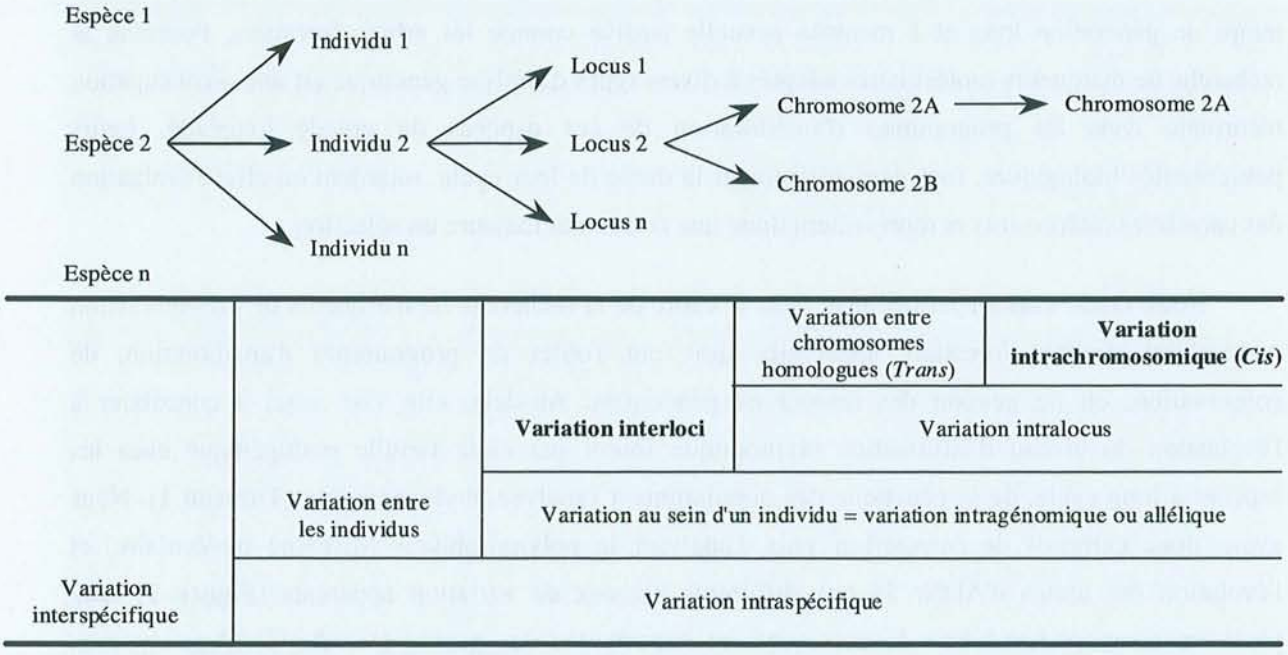



**Figure 1.** Les différents niveaux de variation envisageables pour une famille de séquences répétées. Les niveaux de variation supplémentaires par rapport aux séquences uniques sont indiqués en gras.

Les séquences répétées codant pour les ARNr 5S nucléaires sont physiquement indépendantes des unités d'ADNr mais elles présentent une organisation structurale et des caractéristiques d'évolution tout à fait similaires. Chaque unité est ainsi constituée d'un gène et d'un espaceur non transcrit (NTS : Non-Transcribed Spacer). La séquence des gènes, très conservée, a été exploitée pour étudier les relations phylogénétiques entre les grands groupes d'organismes photosynthétiques (Hori *et al.* 1985). Elle fournit aussi des amorces universelles chez les plantes qui permettent d'amplifier sélectivement la région intergénique. L'espaceur est beaucoup plus variable que le gène. Sa longueur et sa séquence sont généralement considérées comme d'excellents marqueurs d'espèces, même très proches (Kolchinsky *et al.* 1991, Ko et Henry 1996), ce qui permet par exemple leur utilisation pour l'analyse des hybrides obtenus au cours de programmes d'amélioration (Doyle 1988, Zanke *et al.* 1991, Henry 1997). De plus, le polymorphisme intraspécifique de cet espaceur est parfois suffisant pour permettre la distinction de variétés, lignées et écotypes (Kolchinsky *et al.* 1991, Röder *et al.* 1992, 1995, Campell *et al.* 1992). Sa taille, inférieure à 1000 pb, est tout à fait adaptée au séquençage et de nombreux travaux ont démontré son intérêt pour les études de systématique moléculaire au niveau du genre ou de la famille (Appels et Baum 1991, Udovicic *et al.* 1995). Ces gènes constituent

enfin un très bon modèle d'analyse des mécanismes de l'évolution des familles de séquences répétées en tandem (Baum et Appels 1992, Kellogg et Appels 1995, Cronn *et al.* 1996).

Malgré les multiples applications pratiques envisageables, peu de travaux ont été consacrés à l'étude de cette famille de gènes nucléaires codant pour les ARNr 5S (ADNr 5S) chez les plantes à temps de génération long et à maturité sexuelle tardive comme les arbres forestiers. Pourtant, la recherche de marqueurs moléculaires adaptés à divers types d'analyse génétique est une préoccupation récurrente dans les programmes d'amélioration de ces espèces de grande longévité. Leurs particularités biologiques, tout particulièrement la durée de leur cycle, retardent en effet l'évaluation des caractères intéressants et représentent donc une contrainte majeure en sélection.

Notre étude s'inscrit initialement dans le cadre de la recherche de marqueurs de différenciation moléculaire d'arbres forestiers apparentés qui font l'objet de programmes d'amélioration, de conservation, ou de gestion des ressources génétiques. Au-delà, elle vise aussi à contribuer à l'évaluation du niveau d'information taxinomique fourni par cette famille multigénique chez les espèces à long cycle, de la génétique des populations à l'analyse phylogénétique (**Tableau 1**). Nous avons donc entrepris de caractériser puis d'analyser le polymorphisme (diversité moléculaire) et l'évolution des unités d'ADNr 5S aux différents niveaux de variation apparents (**Figure 1**) chez plusieurs essences forestières économiquement importantes des genres *Quercus* L. (Angiosperme, *Fagaceae*, les chênes) et *Larix* M. (Gymnosperme, *Pinaceae*, les mélèzes). Notre choix s'est porté sur deux couples d'espèces proches constitués, d'une part, des chênes pédonculé (*Quercus robur* L.) et sessile [*Quercus petraea* (Matt.) Liebl.] et, d'autre part, des mélèzes d'Europe (*Larix decidua* Mill.) et du Japon [*Larix kaempferi* (Lamb.) Carr.].

Dans une première partie de ce mémoire (**synthèse bibliographique**), nous avons tout d'abord rappelé l'importance ainsi que la complexité d'organisation et d'évolution de la fraction répétée du génome nucléaire des eucaryotes (**Chapitre I**). Puis, nous avons fait le point des connaissances acquises sur la fonction, l'expression, l'organisation et l'évolution des unités d'ADNr 5S nucléaire chez les plantes supérieures (**Chapitre II**). Les résultats de nos travaux sont ensuite donnés dans une seconde **partie expérimentale**, traitant successivement, après une brève **présentation générale**, de l'ADNr 5S des chênes pédonculé et sessile (**Chapitre III**) et des mélèzes d'Europe et du Japon (**Chapitre IV**).

## Bibliographie

**Appels R, Baum B (1991)**. Evolution of the Nor and 5S DNA loci in the *Triticeae*. Dans *Molecular systematics of plants*, pp 92-116. Edité par Soltis PS, Soltis DE, Doyle JJ. Chapman & Hall, New York

**Baldwin BG (1992)**. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the *Compositae*. Mol Phyl Evol 1 : 3-16

**Baum BR, Appels R (1992)**. Evolutionary change at the 5S DNA loci of species in the *Triticeae*. Plant Syst Evol 183 : 195-208

**Campell BR, Song Y, Posch TE, Cullis CA, Town CD (1992).** Sequence and organization of 5S ribosomal RNA-encoding genes of *Arabidopsis thaliana.* Gene 112 : 225-228

**Cavalier-Smith T (1985).** Eukaryotic gene numbers, noncoding DNA and genome size. Dans *The evolution of genome size*, pp 69-103. Edité par Cavalier-Smith T. Wiley, Chichester, UK

**Cronn RC, Zhao X, Paterson AH, Wendel JF (1996).** Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. J Mol Evol 42 : 685-705

**Doyle JJ (1988).** 5S ribosomal gene variation in the soybean and its progenitor. Theor Appl Genet 75 : 621-624

**Flavell RB, Jones J, Lonsdale DM, O'Dell M (1983).** Higher plant genome structure and the dynamics of genome evolution. Dans *Advance in gene technology*, pp 47-59. Edité par Downey K, Voellmy RN, Ahmed F, Schultz J. Academic Press, New York

**Griffiths AJF, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM (1996).** *An introduction to genetic analysis.* Freeman WH et Cie, New York

**Hamby RK, Zimmer EA (1991).** Ribosomal RNA as a phylogenetic tool in plant systematics. Dans *Molecular systematics of plants,* pp 50-91. Edité par Soltis PS, Soltis DE, Doyle JJ. Chapman & Hall, New York

**Henry RJ (1997).** *Practical applications of plant molecular biology.* Chapman & Hall, Londre

**Hori H, Lim BL, Osawa S (1985).** Evolution of green plants as deduced from 5S rRNA sequences. Proc Nat Acad Sci USA 82 : 820-823

**Kellogg EA, Appels R (1995).** Intraspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. Genetics 140 : 325-343

**Ko HL, Henry RJ (1996).** Specific 5S ribosomal RNA primers for plant species identification in admixtures. Plant Mol Biol Rep 14 : 33-43

**Kolchinsky A, Kolesnikova M, Ananiev E (1991).** "Portraying" of plant genomes using polymerase chain reaction amplification of ribosomal 5S genes. Genome 34 : 1028-1031

**Lejeune B, Quétier F, Jubier MF, Falconet D, Rode A, Hartmann C (1988).** Le génome mitochondrial des plantes supérieures : organisation moléculaire et expression. Bull Soc Bot Fr, 135, Actual Bot (2) : 49-55

**Röder MS, Sorrells ME, Tanksley SD (1992).** 5S ribosomal gene clusters in wheat: pulsed field gel electrophoresis reveals a high degree of polymorphism. Mol Gen Genet 232 : 215-220

**Röder MS, Sorrells ME, Tanksley SD (1995).** Pulsed-field gel analysis of 5S and satellite DNA in Barley. Genome 38 : 153-157

**Saiki RK, Gelfand DH, Stoffel S, Starf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988).** Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239 : 487-491

**Schaal BA, Learn GHJr (1988).** Ribosomal DNA variation within and among plant populations. Ann Mo Bot Gard 75 : 1207-1216

**Solignac M, Pelandakis M, Rousset F, Chenvil A (1991).** Ribosomal RNA phylogenies. Dans *Molecular techniques in taxonomy*, pp 73-85. Edité par Hewitt GM, Johnston AWB, Young JPW. Springer-Verlag, Berlin

**Udovicic F, McFadden G, Ladiges PY (1995).** Phylogeny of *Eucalyptus* and *Angophora* based on 5S rDNA spacer sequence data. Mol Phyl Evol 4 : 247-256

**Zanke C, Borisjuk N, Ruoss B, Schilde-Rentschler L, Ninnemann H, Hemleben V (1995).** A specific oligonucleotide of the 5S rDNA spacer and species-specific elements identify symmetric somatic hybrids between *Solanum tuberosum* and *S. pinnatisectum.* Theor Appl Genet 90 : 720-726

**Zimmer EA, Hamby RK, Arnold ML, Leblanc DA, Theriot EC (1989).** Ribosomal RNA phylogenies and flowering plant evolution. Dans *The hierarchy of life*, pp 205-214. Edité par Fernholm B, Bremer K, Jornvall H. Elsevier Science Publishers, Amsterdam

# *SYNTHESE BIBLIOGRAPHIQUE*

# CHAPITRE I : Importance, organisation et évolution des séquences répétées.

## I- Contexte général : plasticité du génome et séquences répétées

### 1- Variations de la taille du génome

A l'exception de quelques virus, le génome de la plupart des êtres vivants est constitué d'une ou plusieurs molécules d'acide désoxyribonucléique (ADN) dont la séquence nucléotidique véhicule l'information génétique déterminant les caractères héréditaires d'un individu (génotype). La quantité d'ADN génomique est généralement considérée comme constante d'une cellule à l'autre, et entre les individus d'une même espèce (Dounce 1955), même si cette notion peut être discutée (Deumling et Clermont 1989, Greilhuber 1998). Afin d'éviter les confusions liées à la duplication de génomes entiers (polyploïdie) chez de nombreux organismes, Swift (1950) a introduit le terme de valeur C (ou valeur 1C) pour désigner la taille du génome ou quantité d'ADN caractéristique du génome haploïde et non répliqué d'une espèce.

Depuis les années cinquante, de nombreuses estimations de la valeur C ont été obtenues pour des milliers d'espèces appartenant aux principaux groupes taxinomiques (Cavalier-Smith 1985). La comparaison de ces données a révélé des différences considérables, aussi bien chez les procaryotes que chez les eucaryotes. Chez les procaryotes, la taille du génome (regroupant essentiellement les fractions dites "chromosomique" et plasmidique) peut ainsi varier d'un facteur 20, soit d'environ $6.10^5$ paires de bases (pb) chez certains parasites obligatoires et mycoplasmes, jusqu'à plus de $10^7$ pb chez plusieurs cyanobactéries (**Tableau 2**). Chez les eucaryotes, le facteur de variation est encore plus grand puisqu'il atteint 80000 fois (**Tableau 3**), la taille du génome passant de $8,8.10^6$ pb (certaines levures) à plus de $6,9.10^{11}$ pb (certaines amibes). Les variations de la valeur C les plus importantes sont observées chez les protistes ($2,3.10^7$–$6,9.10^{11}$ pb) et les plantes ($5,8.10^7$–$3,1.10^{11}$ pb), tandis que les écarts sont moindres chez les animaux ($4,9.10^7$–$1,4.10^{11}$ pb) et les champignons ($8,8.10^6$–$1,5.10^9$ pb). Ces chiffres cachent cependant d'énormes disparités au sein de ces grands ensembles. Chez les animaux par exemple, la taille du génome est relativement homogène au sein des trois groupes de vertébrés amniotes (mammifères, oiseaux et reptiles) mais très variable chez les insectes, les amphibiens et les poissons osseux. De même, la comparaison détaillée des deux principaux groupes de plantes spermaphytes révèle de fortes différences d'amplitude des variations et de distribution des tailles de génomes (**Tableaux 3** et **4**). Les conifères (gymnospermes) constituent en effet un groupe plus homogène que les plantes à fleurs (angiospermes) pour lesquelles les valeurs C sont plus faibles en moyenne (~ 2,5 fois) et surtout en mode[1] (plus de 20 fois). Il faut néanmoins souligner que les données disponibles ne concernent que 1 % des angiospermes et 16 % des gymnospermes (Leitch *et al.* 1998, Murray 1998).

---

[1] Valeur C correspondant à la classe dont l'effectif est le plus important

**Tableau 2.** Variation de la taille du génome dans quelques groupes de procaryotes.

| Taxon | Valeur C (pb) | R* |
|---|---|---|
| Eubactéries | $6{,}5.10^5$–$1{,}3.10^7$ | 20 |
| Gram positif | $6{,}5.10^5$–$7{,}8.10^6$ | 12 |
| Gram négatif | $1{,}6.10^6$–$1{,}2.10^7$ | 7 |
| Cyanobactéries | $3{,}1.10^6$–$1{,}3.10^7$ | 4 |
| Mycoplasmes | $6{,}5.10^5$–$1{,}8.10^6$ | 3 |
| Archéobactéries | $1{,}6.10^6$–$4{,}1.10^6$ | 3 |

D'après Cavalier-Smith (1985).
* Rapport entre les valeurs C maximale et minimale.

**Tableau 3.** Variation de la taille du génome dans quelques groupes d'organismes eucaryotes.

| Taxon | Valeur C (pb) | R* |
|---|---|---|
| Protistes | $2{,}3.10^7$–$6{,}9.10^{11}$ | 29191 |
| Euglénoïdes | $9{,}8.10^7$–$2{,}3.10^9$ | 24 |
| Ciliés | $2{,}3.10^7$–$8{,}6.10^9$ | 367 |
| Amibes | $3{,}5.10^7$–$6{,}9.10^{11}$ | 19433 |
| Champignons | $8{,}8.10^6$–$1{,}5.10^9$ | 167 |
| Animaux | $4{,}9.10^7$–$1{,}4.10^{11}$ | 2837 |
| Eponges | $4{,}9.10^7$–$5{,}4.10^7$ | 1 |
| Annélides | $8{,}8.10^8$–$5{,}2.10^9$ | 6 |
| Mollusques | $4{,}2.10^8$–$5{,}3.10^9$ | 13 |
| Crustacés | $6{,}9.10^8$–$2{,}2.10^{10}$ | 32 |
| Insectes | $9{,}8.10^7$–$9{,}3.10^9$ | 95 |
| Echinodermes | $5{,}3.10^8$–$3{,}2.10^9$ | 6 |
| Agnathes | $6{,}4.10^8$–$2{,}8.10^9$ | 4 |
| Raies et Requins | $1{,}5.10^9$–$1{,}6.10^{10}$ | 11 |
| Poissons osseux | $3{,}8.10^8$–$1{,}4.10^{11}$ | 364 |
| Amphibiens | $9{,}3.10^8$–$8{,}4.10^{10}$ | 91 |
| Reptiles | $1{,}2.10^9$–$5{,}3.10^9$ | 4 |
| Oiseaux | $1{,}2.10^9$–$2{,}2.10^9$ | 2 |
| Mammifères | $1{,}4.10^9$–$5{,}7.10^9$ | 4 |
| Plantes | $5{,}8.10^7$–$3{,}1.10^{11}$ | 3837 |
| Algues | $8{,}0.10^7$–$3{,}0.10^{10}$ | 375 |
| Ptéridophytes | $9{,}8.10^7$–$3{,}1.10^{11}$ | 3133 |
| Gymnospermes | $2{,}2.10^9$–$3{,}1.10^{10}$ | 14 |
| Angiospermes | $5{,}8.10^7$–$1{,}2.10^{11}$ | 2123 |

D'après Cavalier-Smith (1985), Li et Graur (1991), Bennett et Leitch (1995) et Murray (1998).
* Rapport entre les valeurs C maximale et minimale.

## 2- Signification des variations de la taille du génome

### a- Valeur adaptative des variations

L'ADN génomique influe sur les caractères observables d'un individu (phénotype), non seulement par l'expression de certaines de ses régions qui ont une fonction essentielle et spécifique (gènes), mais aussi par l'effet physique de sa masse et/ou de son volume (effet nucléotypique, Bennett 1982). Chez les eucaryotes, en particulier, la taille du génome nucléaire est positivement corrélée à la taille des chromosomes, à la taille des cellules et à la durée totale des cycles mitotique et méiotique (Van't Hof et Sparrow 1963, Bennett 1977). Une telle influence nucléotypique, indépendante de l'action des gènes, sur des caractères cellulaires aussi fondamentaux, indique que la taille du génome joue probablement un rôle primordial dans l'évolution et l'adaptation des organismes face aux contraintes de l'environnement. Cette évolution quantitative des génomes sous pression de la sélection naturelle est particulièrement difficile à analyser car, en dehors des changements du degré de ploïdie, il n'est pas possible de manipuler expérimentalement les valeurs C des organismes supérieurs (Narayan 1998). La valeur adaptative de la taille du génome apparaît cependant évidente, en particulier chez les angiospermes, en raison des très fortes variations de valeurs C enregistrées (Bennett *et al.* 1997). Par exemple, les valeurs C des herbacées annuelles sont en moyenne significativement plus faibles que celles des plantes pérennes (3,8 vs 6,5 pg, Bennett et Leitch 1995). De même, les espèces allogames ont un génome de taille généralement plus élevée que les espèces autogames (Price 1976, Govindaraju et Cullis 1991, Bharathan 1996). De nombreuses études tendent enfin à montrer que la taille du génome aurait une haute valeur prédictive de la réponse des plantes à divers changements climatiques ou écologiques tels que les radiations ionisantes (Sparrow et Miksche 1961, Grime 1986), l'intensité lumineuse (Price et Johnston 1996), le réchauffement climatique global (MacGillivray et Grime 1995, Grime 1996), la diminution de l'ozone (Bennett 1987), ou bien encore l'augmentation de la teneur en $CO_2$ de l'atmosphère (Jasienski et Bazzaz 1995).

**Tableau 4.** Comparaison de la taille du génome des angiospermes et des gymnospermes.

| | **Angiospermes** | **Gymnospermes** |
|---|---|---|
| Nombre approximatif d'espèces dans le monde | 250 000 | 730 |
| Nombre d'espèces dont la taille du génome a été estimée | 2802 (1,1 %) | 117 (16,0 %) |
| Moyenne (pg*) | 6,9 | 17,3 |
| Ecart type | 10,4 | 6,8 |
| Médiane (pg) | 3,3 | 17,4 |
| Mode (pg) | 0,7 | 15,8 |
| Minimum (pg) | 0,06 | 2,25 |
| Maximum (pg) | 127,4 | 31,75 |

D'après Bennett *et al.* (1997), Leitch *et al.* (1998), Murray (1998).
*1 pg = $10^{-12}$ g = $965.10^6$ pb (Bennett et Leitch 1995).

**Tableau 5.** Valeurs C chez différents organismes eucaryotes classés par taille croissante de génome.

| Espèce | Valeur C (pb) | Valeur C (pg*) |
|---|---|---|
| *Saccharomyces cerevisiae* (levure) | $8{,}8.10^{6}$ | 0,009 |
| *Navicola pelliculosa* (diatomée) | $3{,}5.10^{7}$ | 0,04 |
| *Caenorhabditis elegans* (nématode) | $8{,}0.10^{7}$ | 0,08 |
| *Drosophila melanogaster* (drosophile) | $1{,}8.10^{8}$ | 0,2 |
| *Arabidopsis thaliana* (cresson en oreille de souris) | $1{,}9.10^{8}$ | 0,2 |
| *Paramecium aurelia* (cilié) | $1{,}9.10^{8}$ | 0,2 |
| *Bombyx mori* (ver à soie) | $5{,}0.10^{8}$ | 0,5 |
| *Strongylocentrotus purpuratus* (oursin de mer) | $8{,}0.10^{8}$ | 0,8 |
| ***Quercus robur* (chêne pédonculé)** | **$8{,}9.10^{8}$** | **0,9** |
| *Gallus domesticus* (poulet) | $1{,}2.10^{9}$ | 1,2 |
| *Erysiphe cichoracearum* (champignon) | $1{,}5.10^{9}$ | 1,6 |
| *Cyprinus carpio* (carpe) | $1{,}7.10^{9}$ | 1,8 |
| *Lampreta planeri* (lamproie) | $1{,}9.10^{9}$ | 2,0 |
| *Boa constrictor* (serpent) | $2{,}1.10^{9}$ | 2,2 |
| *Ginkgo biloba* (ginkgo) | $2{,}2.10^{9}$ | 2,3 |
| *Parascaris equorum* (vers rond) | $2{,}5.10^{9}$ | 2,6 |
| *Carcarias obscurus* (requin) | $2{,}7.10^{9}$ | 2,8 |
| *Rattus norvegicus* (rat) | $2{,}9.10^{9}$ | 3,0 |
| *Mus musculus* (souris) | $3{,}0.10^{9}$ | 3,1 |
| *Xenopus laevis* (crapaud) | $3{,}1.10^{9}$ | 3,2 |
| *Bovis domesticus* (vache) | $3{,}1.10^{9}$ | 3,2 |
| *Homo sapiens* (humain) | $3{,}4.10^{9}$ | 3,5 |
| *Nicotiana tabacum* (tabac) | $3{,}8.10^{9}$ | 3,9 |
| *Zea mays* (maïs) | $5{,}0.10^{9}$ | 5,2 |
| *Paramecium caudatum* (cilié) | $8{,}6.10^{9}$ | 8,9 |
| *Schistocerca gregaria* (criquet) | $9{,}3.10^{9}$ | 9,6 |
| ***Larix decidua* (mélèze d'Europe)** | **$9{,}5.10^{9}$** | **9,8** |
| *Allium cepa* (oignon) | $1{,}8.10^{10}$ | 18,6 |
| *Coscinodiscus asteromphalus* (diatomée) | $2{,}5.10^{10}$ | 25,9 |
| *Pinus lambertiana* (pin) | $3{,}1.10^{10}$ | 31,8 |
| *Lilium formosanum* (lis) | $3{,}6.10^{10}$ | 37,3 |
| *Necturus maculosis* (salamandre) | $5{,}0.10^{10}$ | 51,8 |
| *Amphiuma means* (triton) | $8{,}4.10^{10}$ | 87,0 |
| *Fritillaria assyriaca* (fritillaire) | $1{,}2.10^{11}$ | 127,4 |
| *Protopterus aethiopicus* (dipneuste) | $1{,}4.10^{11}$ | 145,0 |
| *Ophioglossum petiolatum* (fougère) | $1{,}6.10^{11}$ | 165,8 |
| *Amoeba proteus* (amibe) | $2{,}9.10^{11}$ | 300,5 |
| *Amoeba dubia* (amibe) | $6{,}7.10^{11}$ | 694,3 |

D'après Cavalier-Smith (1985), Li et Graur (1991), Singer et Berg (1992), Bennett et Leitch (1995) et Murray (1998).
*1 pg = $10^{-12}$ g = $965.10^{6}$ pb (Bennett et Leitch 1995).

b- Le paradoxe de la valeur C

Il est clair que les variations de la valeur C reflètent aussi la complexification croissante (spécialisation) des organismes au cours des 3 à 4 milliards d'années d'évolution depuis les origines de la vie sur Terre. La taille du génome a en effet considérablement augmenté des procaryotes aux eucaryotes supérieurs tels que les mammifères (d'un facteur 1000, comparer les **Tableaux 2 et 3**).

Sachant que les formes de vie considérées comme les plus évoluées requièrent davantage d'information génétique pour subsister (Nei 1987), on pourrait aussi s'attendre à une correspondance étroite entre taille du génome, complexité d'un organisme et nombre de gènes à divers niveaux taxinomiques. C'est le cas chez les bactéries qui possèdent relativement peu d'ADN non codant ou sans fonction apparente (= ADN en excès selon Doolittle et Sapienza 1980). Les variations de la valeur C (d'un facteur 20) traduisent donc relativement bien le nombre de gènes qu'elles renferment, estimé à environ 400 chez les bactéries autonomes les plus simples (les mycoplasmes) à plus de 8000 chez des bactéries plus complexes telles que les cyanobactéries (Li et Graur 1991). Une telle relation entre taille du génome et complexité des organismes est pratiquement inexistante chez les eucaryotes. Par exemple, chez certains protozoaires (amibes), plusieurs amphibiens et une espèce de dipneuste, la taille du génome est beaucoup plus importante que chez les mammifères (**Tableau 5**). De même, dans le cas des plantes, certaines algues unicellulaires (diatomées) ou bien encore certaines fougères et de nombreuses angiospermes ont des génomes de tailles tout à fait semblables. De plus, les valeurs C sont souvent très différentes entre des espèces qui présentent un niveau d'organisation similaire (drosophile et criquet, oignon et lis, *Paramecium caudatum et aurelia, etc.*). Enfin, chez les eucaryotes, le nombre de gènes ne varie, au mieux, que d'un facteur 50 (les estimations les plus larges indiquent 3000 gènes environ chez certains champignons et 150000 chez les mammifères, Li et Graur 1991), ce qui est nettement inférieur aux différences précédemment décrites (80000 fois) au niveau de la taille des génomes. Cette absence de correspondance entre la valeur C et la complexité d'un organisme ou quantité présumée d'information génétique qu'un génome contient, constitue ce qui est mentionné dans la littérature comme étant le paradoxe de la valeur C (Thomas 1971). L'existence de ce paradoxe est généralement attribuée à la présence massive chez les eucaryotes d'ADN non codant ou sans fonction apparente, souvent répété et en quantité extrêmement variable selon les espèces ($3.10^6$ pb à plus de $10^{11}$ pb, Cavalier-Smith 1985).

### 3- Importance des séquences répétées dans les génomes eucaryotes

#### a- Relation entre taille du génome et proportion de séquences répétées

La découverte de l'existence d'ADN répété (= réitéré) dans la plupart des génomes eucaryotes (Britten et Kohne 1968, Flamm 1972) a été déterminante pour rendre compte, au moins en partie, des importantes fluctuations de quantités d'ADN. Il existe en effet une relation assez générale mais positive entre la taille du génome nucléaire et la proportion d'éléments répétés qu'il contient (Flavell *et al.* 1974). Les plus petits génomes d'eucaryotes tels que celui de la levure *Saccharomyces cerevisiae* (~ $10^7$ pb) ne renferment ainsi, comme les procaryotes, qu'une faible proportion de séquences répétées (moins de 10 %). Chez certains protozoaires ou la drosophile (~ $10^8$ pb), les réitérations représentent 10 à 20 % du génome, et chez les mammifères (~ $10^9$ pb), 20 à 50 %. Enfin, parmi les plus grands génomes, tels que ceux de certains amphibiens et de nombreuses plantes (~ $10^{10}$ pb), il n'est pas rare d'observer des proportions de séquences répétées dépassant 50 % et atteignant plus de

90 % (Flavell *et al.* 1974, Flavell 1986). A plus petite échelle taxinomique, cette relation est particulièrement évidente chez les plantes supérieures qui maintiennent de très fortes proportions de séquences répétées (Nei 1987, Cullis 1990, Kubis *et al.* 1998a) dans leur génome nucléaire (**Tableau 6**). Flavell et ses collaborateurs (1974) ont en effet démontré que les plus petits génomes d'angiospermes (moins de 2 pg, souvent des annuelles) contiennent en moyenne moins d'éléments répétés que les plus grands (62 % contre 80 %, **Tableau 6**).

De même, dans le genre *Lathyrus* (gesse), Narayan (1998) a montré que les accroissements discontinus des valeurs C qui sont intervenus au cours de l'évolution (5 classes majeures entre 5,4 et 12,3 pg) s'expliquent par l'augmentation de la proportion d'éléments répétés. C'est la divergence de taille (nombre de copies) et de séquence des différentes catégories (familles) existantes de réitérations qui serait la base moléculaire de l'évolution de la taille du génome des espèces de ce genre plutôt que l'émergence de nouvelles familles.

b- Contribution de l'amplification de séquences à l'évolution de la taille du génome

Il ne fait maintenant aucun doute (Singer et Berg 1992) que la multiplication de séquences ancestrales soit à l'origine de l'établissement dans les génomes contemporains de nombreuses familles de séquences répétées d'importance variable (jusqu'à plusieurs millions de copies) et dont l'unité élémentaire peut varier considérablement en taille (de 2 à plus de 10000 pb). Les mécanismes conduisant à l'amplification (ou la déplétion) locale de séquences d'ADN (**Partie II de ce Chapitre I**) contribuent donc très largement à l'évolution de la taille des génomes eucaryotes (Nei 1987, Li et Graur 1991), en particulier des plantes supérieures (Cullis 1990, Narayan 1998), même si d'autres mécanismes plus globaux (**Encadré 1**) tels que la duplication de génomes entiers (endopolyploïdisation), de chromosomes complets (aneuploïdie) ou de régions chromosomiques (aneuploïdie partielle associée à la présence de chromosomes surnuméraires ou chromosomes B) peuvent aussi intervenir de manière significative dans certains groupes d'organismes (Li et Graur 1991). L'impact considérable que peuvent avoir les mécanismes d'amplification locale d'ADN est illustrée de manière éloquente par l'observation chez le maïs d'une augmentation récente de la taille du génome liée à la multiplication massives de séquences répétées particulières (des éléments mobiles, San Miguel *et al.* 1996, SanMiguel et Bennetzen 1998).

**4- Fonction des familles de séquences répétées**

a- Familles de l'ADN non codant

Au-delà de leur impact sur la taille du génome, dont on a vu qu'elle avait une forte valeur adaptative, la fonction éventuelle de la plupart des familles de séquences répétées est encore

**Tableau 6.** Proportion de séquences répétées dans le génome de quelques plantes supérieures.

| Espèce | Catégorie | Valeur C (pg) | Proportion de séquences répétées (%) |
|---|---|---|---|
| Angiospermes (valeurs C < 2 pg) | | | |
| *Linum usitatissimum* | Annuelle | 0,7 | 59 |
| *Capsella bursa-pastoris* | Annuelle | 0,8 | 46 |
| *Veronica persica* | Annuelle | 0,9 | 63 |
| *Daucus carota* | Annuelle/Bisannuelle | 1,0 | 62 |
| *Stellaria media* | Annuelle | 1,2 | 69 |
| *Lamium purpureum* | Annuelle | 1,3 | 60 |
| *Beta vulgaris* | Annuelle/Bisannuelle | 1,3 | 63 |
| *Senecio vulgaris* | Annuelle | 1,7 | 74 |
| | | | Moyenne : 62 ± 8,1* |
| Angiospermes (valeurs C > 2 pg) | | | |
| *Poa trivialis* | Vivace | 3,4 | 82 |
| *Tropaeolum majus* | Annuelle | 3,6 | 70 |
| *Pisum sativum* | Vivace | 4,9 | 75 |
| *Helianthus annuus* | Annuelle | 5,3 | 69 |
| *Zea mays* | Annuelle | 5,5 | 78 |
| *Hordeum vulgare* | Annuelle/Bisannuelle | 6,7 | 76 |
| *Poa annua* | Annuelle/Bisannuelle | 6,9 | 87 |
| *Triticum monococcum* | Annuelle/Bisannuelle | 7,0 | 80 |
| *Secale cereale* | Annuelle/Bisannuelle | 9,4 | 92 |
| *Vicia faba* | Annuelle | 14,6 | 85 |
| *Allium cepa* | Vivace | 16,7 | 95 |
| *Triticum aestivum* | Annuelle/Bisannuelle | 18,1 | 83 |
| *Avena sativa* | Annuelle | 21,5 | 83 |
| *Tulipa kaufmanniana* | Vivace | 31,2 | 73 |
| *Hyacinthus orientalis* | Vivace | 49,0 | 75 |
| | | | Moyenne : 80 ± 7,6* |
| Gymnospermes | | | |
| *Picea glauca* | Vivace | 8,5 | 68 |
| *Pinus resinosa* | Vivace | 20,5 | 72 |
| *Pinus strobus* | Vivace | 19,4 | 76 |
| *Pinus banksiana* | Vivace | 14,9 | 78 |
| *Pinus lambertiana* | Vivace | 17,6 | 79 |
| | | | Moyenne : 75 ± 4,6* |

D'après Flavell *et al.* (1974) et Dhillon (1987).
* Ecart type.

largement méconnue car elles constituent surtout l'ADN non codant (Lewin 1975, Orgel et Crick 1980), qualifié d'ADN "rebut" (Ohno 1972), "non sens" (Wang et Fudenberg 1974), "ignorant" (Dover 1980), "égoïste" (Doolittle et Sapienza 1980, Orgel *et al.* 1980, Orgel et Crick 1980), ou "parasite" (Orgel et Crick 1980). Certains auteurs ont suggéré que les répétitions de l'ADN non codant pouvaient avoir un rôle structural et fonctionnel lié, par exemple, à leur intégration préférentielle dans les zones hétérochromatiques des chromosomes comme les centromères ou les télomères ou dans les



S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique

régions de régulations associées aux gènes (Brutlag 1980, Cavalier-Smith 1985, Britten 1994).

**Encadré 1.** Mécanismes globaux d'augmentation de la taille des génomes eucaryotes.

Duplication du génome (endopolyploïdisation).

La duplication d'un génome (polyploïdie) intervient à la suite de l'absence de disjonction entre les chromatides soeurs (mitose et méiose II) ou entre les chromosomes homologues (méiose I) après la réplication de l'ADN. Un tel événement ne contribue pas immédiatement à augmenter la taille du génome définie comme la quantité haploïde d'ADN génomique (valeur C). Cependant, les mutations, translocations et autres réarrangements entre les différents lots de chromosomes peuvent conduire à la création d'un nouveau génome de taille approximativement doublée (Li et Graur 1991). En d'autres termes, un génome polyploïde peut redevenir diploïde au cours de l'évolution (Cavalier-Smith 1985) avant de subir éventuellement de nouveaux cycles de duplication. Comme chez les procaryotes (Herdman 1985), ce mécanisme semble avoir significativement contribué à l'augmentation de la taille du génome de certains groupes d'eucaryotes comme les plantes angiospermes, les champignons, les échinodermes, les insectes et, dans une moindre mesure, les amphibiens et les poissons osseux (Rees et Jones 1972, Grime et Mowforth 1982). La distribution des tailles de génome est en effet largement polymodale dans ces groupes d'organismes. Par exemple, 5 classes principales de valeurs C, clairement multiples, sont observées chez les angiospermes monocotylédones (0.60, 1,18, 2,16, 4,51 et $8{,}53.10^9$ pb, Sparrow et Nauman 1976). Etrangement, alors que la contribution de la polyploïdie dans l'évolution de la taille du génome est particulièrement évidente chez les angiospermes (plus de 70 % des espèces actuelles sont polyploïdes, Goldblatt 1980, Lewis 1980), elle semble extrêmement limitée chez les gymnospermes (moins de 1 % des espèces sont polyploïdes, Koshoo 1959).

Duplication totale ou partielle de chromosomes (aneuploïdie).

La duplication de chromosomes entiers (aneuploïdie) ou en partie (aneuploïdie partielle) est le plus souvent délétère (trisomies, infertilité, etc.). Cependant, des chromosomes surnuméraires (chromosomes B) peuvent s'accumuler et se maintenir chez certains organismes, en particulier dans les populations de plantes angiospermes (Jones 1985, Beukeboom 1994, Burt et Trivers 1998). Ces chromosomes B sont le plus souvent riches en séquences répétées (ils sont fortement hétérochromatiques) et proviennent probablement de la duplication de régions chromosomiques plus ou moins vastes. L'aneuploïdie partielle peut donc contribuer à l'augmentation de la taille du génome.

D'autres ont indiqué que la mutabilité importante de ces familles de séquences répétées contribue au maintien à long terme d'un fort potentiel d'évolution au sein des populations naturelles (Nevers et

Saedler 1977, Syvanen 1984). Cependant, loin d'être toujours bénéfique ou sélectivement neutre pour l'organisme hôte, comme le suggèrent les concepts d'ADN rebut, non sens ou ignorant, l'accumulation incontrôlée des séquences répétées dans un génome (concepts d'ADN égoïste ou parasite) peut à partir d'un certain seuil devenir néfaste (maladie, coût énergétique pour la cellule) et conduire périodiquement à leur élimination par sélection phénotypique (Mackay 1986, Wallace *et al.* 1991, Lakich *et al.* 1993, Charlesworth *et al.* 1994). Ces familles d'ADN non codant ne sont donc pas universelles.

b- Familles multigéniques

Le rôle de la répétition de séquence est plus clair (non controversé) lorsqu'elle affecte la fraction réduite du génome qui contient les gènes fonctionnels codants (transcrits). Les copies surnuméraires d'un gène essentiel peuvent être quasiment identiques et permettre ainsi, par effet additif, une production plus grande de tel ARN ou de telle protéine. Elles peuvent aussi accumuler des mutations sans affecter les fonctions physiologiques cellulaires. Fortuitement, les copies altérées peuvent coder pour une nouvelle fonction ou répondre à des signaux de régulation différents. S'ils sont utiles, ces nouveaux gènes peuvent alors être conservés et éventuellement amplifiés par la pression de sélection. Dans le cas contraire les copies inactivées qui constituent des pseudogènes peuvent être éliminées ou servir de support à de nouvelles évolutions. La duplication des gènes existants peut donc être à la base de l'apparition de nouveaux gènes (Ohno 1970, Singer et Berg 1992).

Deux grandes classes de familles multigéniques peuvent donc être distinguées selon la structure et la fonction avérée de la répétition (non mutuellement exclusives). La première classe est constituée des familles multigéniques dites polymorphes (répétition de variants) qui regroupent plusieurs gènes (de deux à plusieurs dizaines) dont les séquences nucléotidiques fortement (sous-familles) ou plus faiblement apparentées (super-familles) résultent d'une évolution par duplication et divergence d'un gène ancestral commun (Long et Dawid 1980). Les gènes appartenant à de telles familles codent respectivement pour des polypeptides modérément divergents (plus de 50 % de similarité, comme les actines, les tubulines, les zéines, *etc.*) ou assez divergents (moins de 50 % de similarité, comme la prolactine et l'hormone de croissance, les α-et β-globines, la sérumalbumine et l'α-foetoprotéine, *etc.*). Ils sont souvent associés à différents signaux de régulation permettant leur expression différentielle dans les tissus ou à des moments différents du développement. L'existence de ces sous-familles et super-familles multigéniques confèrent donc des capacités physiologiques diversifiées et permettent un contrôle précis des processus cellulaires.

La deuxième classe regroupe les familles multigéniques à répétition dite "de dosage" (Long et Dawid 1980) qui doivent répondre aux demandes instantanées de la cellule en certaines molécules requises en grande quantité. Il s'agit essentiellement des gènes codant pour les acides ribonucléiques (ARN) ribosomiques (ARNr) et de transfert (ARNt) qui participent à la traduction des ARN

messagers (ARNm), et des gènes transcrits en petits ARN nucléaires (ARNsn) et cytoplasmiques (ARNsc) qui interviennent dans la maturation des ARN précurseurs et le transport intracellulaire des protéines. Contrairement aux gènes codant pour des protéines dont chaque transcription en ARNm fournit plusieurs molécules de produit par l'intermédiaire de traductions multiples, une transcription d'un gène d'ARN ne fournit qu'une seule molécule du produit final. Chaque famille de cette classe comporte donc généralement un grand nombre de gènes identiques ou très peu divergents (de quelques dizaines à plusieurs milliers de copies). Cependant, quelques familles multigéniques codant pour des protéines nécessaires en très grande quantité comme les histones (compactage de l'ADN) ou l'ubiquitine (impliquée dans des mécanismes protéolytiques sélectifs) présentent aussi une répétition de dosage chez certains organismes tels que la drosophile (Coen *et al.* 1982), les rongeurs et l'homme (Nenoi *et al.* 1998). Tout comme la taille du génome, il est intéressant de remarquer que la taille des familles multigéniques à répétition de dosage (particulièrement les familles des ARNr et ARNt) augmente au cours de l'évolution en réponse aux besoins accrus des organismes supérieurs en ARN pour la synthèse protéique (**Tableau 7**). Il existe donc une corrélation très générale entre les valeurs C et le nombre de copies de ces gènes (Long et Dawid 1980, Nei 1987, Li et Graur 1991). De plus, la grande taille de ces familles leur confère une plus grande tolérance à la mutation. Elles peuvent en effet accumuler des copies non-fonctionnelles (pseudogènes) sans être éliminées par la pression de sélection (jusqu'à un certain seuil). Ces familles au rôle bien défini constituent donc d'excellents modèles d'étude des mécanismes de l'évolution.

**Tableau 7.** Nombre approximatif de gènes codant pour les ARNr et les ARNt chez divers organismes classés par taille de génome.

| Source génomique | Gènes des ARNr[a] | Gènes des ARNr 5S | Gènes des ARNt[b] | Valeur C (pb) |
|---|---|---|---|---|
| Procaryotes | | | | |
| Mitochondrie humaine | 1 | 1 | 22 | $1,7.10^4$ |
| *Mycoplasma capricolum* | 2 | 2 | ? | $1,0.10^6$ |
| *Escherichia coli* | 7 | 7 | 100 | $4,0.10^6$ |
| Eucaryotes | | | | |
| *Saccharomyces cerevisiae* | 140 | 150 | 320–400 | $8,8.10^6$ |
| *Drosophila melanogaster* | 100–250 | 100–200 | 590–900 | $1,8.10^8$ |
| *Rattus norvegicus* | 150 | 830 | 6500 | $2,9.10^9$ |
| *Xenopus laevis* | 400–760 | 400 (cellules somatiques)<br>9000-24000 (ovocytes) | 6500–7800 | $3,1.10^9$ |
| *Homo sapiens* | 50–300 | 1700-2000 | 1310 | $3,4.10^9$ |
| *Pisum sativum* | 3900 | 3000–7000 | ? | $4,4.10^9$ |
| *Triticum aestivum* | 6350 | 7000–10000 | ? | $1,7.10^{10}$ |

D'après Long et Dawid (1980), Ellis *et al.* (1988), Li et Graur (1991), Sorensen et Frederiksen (1991), Singer et Berg (1992), Röder *et al.* (1992).

? : inconnu.

[a] ARNr 16S–23S (procaryotes) ou 18S–5,8S-28S (eucaryotes).

[b] Nombre total de gènes pour l'ensemble des classes d'ARNt.

**5- Conclusion**

L'étude de la dynamique évolutive des familles de séquences répétées revêt donc un intérêt considérable chez les eucaryotes (Charlesworth *et al.* 1994). L'amplification des séquences d'ADN fournit en effet le matériel de base nécessaire à l'évolution du génome en taille et en séquence. Elle lui octroie ainsi une certaine flexibilité et la possibilité d'évoluer sous l'effet de la sélection naturelle. L'analyse des différences intraspécifiques et interspécifiques en matière de séquence nucléotidique, d'organisation génomique et de polymorphisme des unités répétées constituent donc des apports importants pour les études plus classiques des relations entre organismes et des mécanismes de l'évolution. Dans la partie suivante, nous nous attacherons donc à décrire la structure de la fraction répétée des génomes nucléaires, les mécanismes particuliers qui lui donnent naissance et ceux qui gouvernent son évolution.

**Tableau 8.** Dispersion dans le génome des membres de quelques familles multigéniques codant pour des polypeptides chez les eucaryotes.

| Famille | Organisme | Taille[a] | Dispersion[b] |
|---|---|---|---|
| Phosphatase acide | Levure | ≥ 4 | G, d |
| Ovalbumine | Poule | 3 | G |
| α-amylase | Souris | ≥ 3 | G |
| Collagène (interstitiel) | Homme | 4 | d |
| Interférons | Homme | ≥ 23 (5) | G, d |
| Hormone de croissance | Homme | 6 | G, d |
| Petite sous-unité de la Rubisco[c] | Tomate | 5 | G, d |
| Zéine | Maïs | 100 | G, d |
| Immunoglobuline | Homme | ~ 220 (+) | G, d |
| Complexe majeur d'histocompatibilité | Homme | ≥ 27 (+) | G, d |
| | Souris | ≥ 31 (+) | G, d |
| Sérumalbumine | Souris | 2 | G |
| | Homme | 3 | G |
| Actines | *Dictyostelium* | 17 (+) | G, d |
| | Nématode | 4 | G, d |
| | Oursin | 8 (2) | G, d |
| | Drosophile | 6 | d |
| | Poule | 8-10 (+) | d |
| | Rat | ≥ 8 (+) | d |
| | Homme | 20-30 (15-25) | d |
| Tubulines | Levure | 3 | d |
| | *Chlamydomonas* | 4 | d |
| | Trypanosome | 30 | T |
| | Oursin | 15 (10) | G, d |
| | Drosophile | 8 | d |
| | Poule | 14 (5) | d |
| | Mammifères | 25 (15) | d |
| Globines | Xenope | 12 | G, d |
| | Poule | 7 | G, d |
| | Souris | 9 (2) | G, d |
| | Homme | 13 (4) | G, d |
| Leghémoglobine | Soja | 4 (1) | G |
| Histones | Levure | 4 | Td |
| | Oursin | 510 | T, d |
| | Triton | 700 | T |
| | Xenope | 25 | Td |
| | Drosophile | 100 | T |
| | Poule | 11 | Td |
| | Homme | 15-25 | Td |

D'après Singer et Berg (1992) et Nei *et al.* (1997).

[a] Nombre approximatif de gènes et éventuels pseudogènes membres de la famille. Lorsqu'il est connu le nombre de pseudogènes (maturés ou non) est indiqué entre parenthèse. (+) indique la présence probable de pseudogènes.

[b] G : gènes groupés en série imparfaite, T : gènes groupés en série parfaite, d : gènes dispersés, Td : séries parfaites dispersées.

[c] Ribulose 1,5-bisphosphate carboxylase.

## II- Architecture, formation et évolution de la fraction répétée des génomes nucléaires

L'architecture moléculaire de la fraction redondante du génome nucléaire eucaryote est extrêmement complexe et varie énormément d'un organisme à l'autre. Les différentes familles de séquences répétées qui la constituent peuvent néanmoins être schématiquement classées en deux grandes catégories sur la base de leur organisation et de leur localisation physique sur les chromosomes (Charlesworth *et al.* 1994). Il s'agit d'une part des familles dont les membres sont dispersés en de nombreux loci du génome, et d'autre part, des familles qui se présentent sous la forme d'alignements (séries ou "clusters" en anglais) ininterrompus d'unités répétées en tandem (séries parfaites d'unités adjacentes, "array" en anglais). La nature même des séquences répétées les rend particulièrement variables et donc, toutes les formes d'organisation intermédiaires peuvent se rencontrer. Ainsi, les membres d'une même famille peuvent être en partie groupés en tandem et en partie dispersés (comme les gènes de certains polypeptides et des ARNt par exemple). Parfois, les répétitions en tandem, directes ou inverses, sont interrompues par d'autres types de séquences non apparentées (séries imparfaites). Les séries parfaites ou imparfaites peuvent enfin être elles-mêmes distribuées sur plusieurs chromosomes (Singer et Berg 1992).

### 1- Les familles de séquences répétées et dispersées

L'analyse des profils de dispersion des séquences d'ADN obtenus par cinétiques de réassociation (Britten et Kohne 1968) a démontré qu'une part importante du génome des eucaryotes supérieurs était constituée de séquences répétées et dispersées qui représentent approximativement 9-15 % du génome de la drosophile (Manning *et al.* 1975, Pimpinelli *et al.* 1995), jusqu'à 20 % du génome humain (Schmid et Deininger 1975) et probablement plus de 50 % du génome des plantes (Kubis *et al.* 1998a). La dispersion de séquence affecte totalement ou en partie seulement les nombreux gènes codant pour les ARNt (**Tableau 7**) et certaines familles de gènes de polypeptides (**Tableau 8**). Par exemple, chez la levure, les gènes d'ARNt sont complètement dispersés tandis que chez la drosophile, ils sont distribués au niveau d'au moins une cinquantaine de loci (Clarkson 1983). De même, chez la plupart des espèces, les gènes d'actines et de tubulines sont distribués en plusieurs sites éloignés du génome. Ces familles multigéniques sont néanmoins trop peu nombreuses et de taille trop faible (généralement moins de 30 membres pour les familles de gènes de polypeptides) pour rendre compte des quantités énormes de répétitions dispersées dans les génomes eucaryotes. A côté de ces familles de faible importance, il existe aussi plusieurs centaines à plusieurs milliers de familles de séquences dispersées de taille parfois imposante (jusqu'à $10^6$ copies) dans l'ADN non codant ou sans fonction apparente (Singer et Berg 1992). Remarquablement diversifiées d'un organisme à l'autre, ces familles se révèlent pour la plupart composées d'éléments mobiles ou transposables qui engendrent la dispersion (Doolittle et Sapienza 1980, Kidwell et Lisch 1997, Kubis *et al.* 1998a).

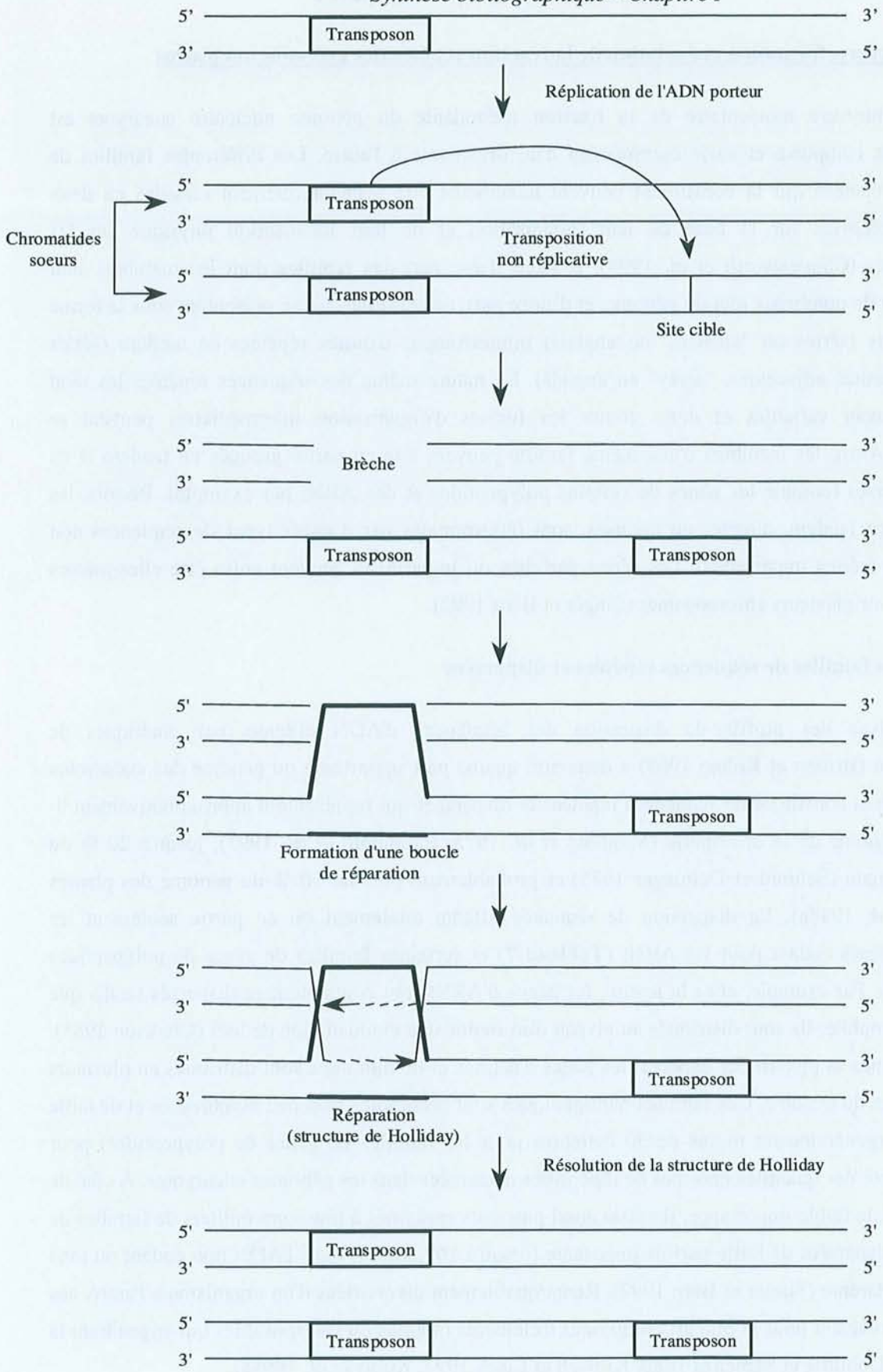



**Figure 2.** Mécanisme potentiel de formation et d'amplification des familles de transposons. Modèle proposé par Engels (1992) pour les éléments P de la drosophile.

La transposition apparaît comme le mécanisme majeur de formation des familles de répétitions dispersées (Li et Graur 1991, Kidwell et Lisch 1997). Le déplacement des éléments transposables s'accompagne en effet le plus souvent de leur duplication. Des cycles successifs de transposition peuvent donc conduire à la multiplication de ces éléments dans un génome. Deux groupes d'éléments transposables et répétés peuvent être distingués selon que le mécanisme de transposition implique ou non un ARN intermédiaire. Il s'agit respectivement des rétroéléments et des transposons. Les familles de rétroéléments sont généralement de taille très importante (des milliers de copies) et contribuent donc en majorité à la fraction répétée dispersée d'un génome. Les familles de transposons ont en revanche une taille beaucoup plus faible (quelques dizaines de copies seulement) en raison de phénomènes d'autorégulation. L'augmentation du nombre de copies de ces éléments induit en effet souvent la répression du mécanisme de transposition (Kidwell et Lisch 1997).

a- Familles de transposons

Les transposons codent directement pour une enzyme (la transposase) responsable de leur excision du site où ils se trouvent (site donneur) et sont caractérisés par des répétitions inversées terminales (approximativement 10-30 pb) nécessaires à leur intégration au site cible (par recombinaison spécifique). Celui-ci comporte généralement de 3 à 8 pb et se trouve dupliqué aux extrémités du transposon lors de son insertion. De nombreuses familles analogues de transposons ont été identifiées chez les animaux et les plantes (McClintock 1951, Bingham *et al.* 1982, Coen et Carpenter 1986, Gierl *et al.* 1988, Engels 1989, Fedoroff 1989, Kidwell et Lisch 1997). Par exemple, il existe au moins une douzaine de ces familles de transposons chez le maïs (Fedoroff 1989) et entre 50 et 100 chez la drosophile (Rubin 1983). Les mieux connues sont probablement la famille des éléments de contrôle "Ac-Ds" du maïs (McClintock 1951, Fedoroff 1989) et la famille des éléments "P" de la drosophile (Bingham *et al.* 1982, Engels 1989). Chez le maïs, la famille Ac-Ds est composée d'environ 35 membres/C (par valeur C) appartenant à deux catégories d'éléments bien définies : les éléments activateurs (Ac) autonomes et instables car capables de transposition (ils codent pour la transposase) et les éléments de dissociation (Ds, responsables en particulier de cassures chromosomiques) non autonomes et stables car ils ne peuvent se déplacer qu'en présence de l'élément Ac. Les éléments Ds sont de taille variable (400 à 4500 pb) et dérivent simplement des éléments Ac complets (d'environ 4600 pb) par délétion. La famille d'éléments P de la drosophile a une organisation similaire. Selon la souche et l'espèce, il est possible de recenser jusqu'à 50 éléments P/C de 500 à 2900 pb. Les fragments de 2900 pb codent pour la transposase et peuvent suppléer les éléments plus courts non autonomes. La fréquence de dispersion de ces éléments a été estimée à une insertion par chromosome et par génération.

La formation et l'amplification de la plupart de ces familles de transposons se réalisent probablement selon un mécanisme proche du modèle proposé par Engels (1992) pour les éléments P de la drosophile (**Figure 2**). Dans ce modèle, la duplication d'un transposon ne peut intervenir que

lorsque la transposition s'effectue après réplication de l'ADN porteur en prophase de division cellulaire. Après excision d'une des deux copies du transposon portées par les chromatides soeurs et insertion au niveau d'un site cible (transposition non réplicative similaire à celle de nombreux transposons procaryotes), le site donneur ouvert subit une réparation basée sur la résolution d'une structure de Holliday générée par l'intermédiaire de la copie soeur non transposée (processus analogue à la réparation double-brin de la levure, Griffiths *et al.* 1996). Le transposon est donc rétabli à sa position d'origine et se trouve dupliqué sur l'une des deux chromatides soeurs. Les éléments Ac de maïs et Tam3 d'*Antirrhinum majus* (Coen et Carpenter 1986) peuvent aussi se transposer au cours de la réplication et se trouver dupliqués si l'insertion a lieu sur une chromatide non encore répliquée (Berg et Howe 1989).

b- Familles de rétroéléments

Les rétroéléments des eucaryotes utilisent une enzyme particulière, la transcriptase inverse, qui catalyse la synthèse d'une copie ADN du rétroélément dite complémentaire (ADNc) à partir des transcrits primaires d'ARN (transcription inverse). Ces copies libres des rétroéléments peuvent alors s'accumuler en grand nombre et se disperser (sous leurs formes ARN ou ADN) pour finalement s'intégrer (sous leur forme ADN) en de multiples loci du génome en provoquant une petite duplication du site cible (jusqu'à 14 pb). La genèse et l'amplification des familles de rétroéléments se réalisent donc simplement par cycles successifs de transcription "normale" et transcription inverse. La multiplication des rétroéléments peut ainsi prendre des proportions énormes. Chez les plantes en particulier, ils peuvent représenter plus de 50 % du génome nucléaire (Bennetzen 1996, SanMiguel *et al.* 1996, Pearce *et al.* 1996, SanMiguel et Bennetzen 1998). On distingue généralement (Hull et Covey 1996) les rétroéléments autonomes (les rétrotransposons) qui codent directement pour la transcriptase inverse requise pour leur transposition et les rétroéléments passifs ou non autonomes (les rétroséquences) qui n'ont pas la capacité de produire cette enzyme mais peuvent l'utiliser si elle est fournie en *trans* (par exemple par un élément autonome).

α- Rétrotransposons

Ces éléments ont une taille de plusieurs milliers de pb et possèdent deux ou trois cadres de lecture ouverts (ou ORF : Open Reading Frame) qui codent généralement pour plusieurs protéines spécifiques (protéase, intégrase, ARNase H) en plus de la transcriptase inverse. Les extrémités des rétrotransposons de la classe I sont constituées de longues séquences répétées de même orientation (quelques centaines à un millier de pb) et pratiquement identiques au sein d'une même famille (LTRs : Long Terminal Repeats) qui sont absentes chez les rétrotransposons de la classe II (ou rétroposons). Les rétrotransposons de la classe I sont également appelés rétrotransposons viraux car ils ressemblent aux provirus des rétrovirus (fréquents chez les mammifères mais inconnus chez les plantes, Kubis *et al.* 1998a) par leur structure ainsi que par leur mode de multiplication et de transposition. Ils s'en distinguent cependant par l'absence de toute forme extracellulaire viable.

*- Rétrotransposons de la classe I (ou viraux)*

Deux grandes catégories de rétrotransposons de classe I peuvent être distinguées en fonction de la disposition des gènes dans le second ORF (Kubis *et al.* 1998a). Les rétrotransposons du groupe Ty1-copia possèdent un domaine intégrase situé en amont du gène codant pour la transcriptase inverse. Au contraire, chez les rétrotransposons du type Ty3-gypsy, le domaine intégrase est situé en aval. Dans ce groupe, un troisième ORF est parfois observé et donc ces rétroéléments sont en ce sens très similaires aux provirus. Les premières familles d'éléments de ces deux groupes ont été caractérisées chez la levure *Saccharomyces cerevisiae* (Ty1, Clare et Farabaugh 1985, Ty3, Hansen *et al.* 1988) et *Drosophila melanogaster* (copia, Mount et Rubin 1985, gypsy, Marlor *et al.* 1986). Les éléments répétés de la famille Ty1 de *S. cerevisiae* ont une taille d'environ 5900 pb et comportent des LTRs de 330 pb (nommés δ). Il en existe jusqu'à 35 copies/C selon la souche de levure, ce qui représente approximativement 2 % de leur génome. De plus, on trouve de nombreux fragments δ (au moins une centaine/C) qui ne sont pas rattachés à des éléments Ty complets. Chez la drosophile, il existerait au moins 20 familles distinctes d'éléments du type copia ou gypsy (Bingham et Zachar 1989) représentant plus de 10 % du génome (5 à 100 copies par famille). Comme pour les éléments Ty de la levure, les membres de ces familles ont une taille de plusieurs kpb (5100 à 8500 pb) et des LTRs de 300 à 500 pb. Des familles analogues ont été détectées chez l'amibe de terreau *Dictyostelium discoideum*, diverses espèces de souris et de nombreuses plantes. Le génome de *Dictyostelium discoideum* contient ainsi 40 copies complètes d'un motif DIRS-1 et 200 à 300 fragments de cet élément (Cappello *et al.* 1984). L'ADN IAP qui code pour les ARN des particules intracisternales A du réticulum endoplasmique de souris (particules ribonucléoprotéiques) est une famille typique de rétrotransposons de classe I du groupe Ty1-copia. Selon les espèces, 20 à 1000 unités de 7100 pb (il en existe aussi de plus courtes) bordées par des LTR de 338 pb sont dispersées sur tous les chromosomes (Kuff et Lueders 1988). Chez les plantes supérieures et les algues vertes, les familles de rétrotransposons viraux sont extrêmement fréquentes (Flavell *et al.* 1992a, Voytas *et al.* 1992, Lindauer *et al.* 1993) qu'elles soient du groupe Ty1-copia (Bennetzen 1996) ou Ty3-gypsy (Suoniemi *et al.* 1998). Les familles du groupe Ty1-copia sont actuellement les mieux décrites. La taille des unités varie de 4800 pb pour l'élément "Hopscotch" de maïs (White *et al.* 1994) à plus de 12000 pb pour l'élément BAR-1 de l'orge (Manninen et Schulmann 1993). Le nombre de ces éléments peut atteindre le million/C et représente 50 % du génome du maïs (SanMiguel *et al.* 1996), 10 à 40 % du génome de *Vicia faba* (Pearce *et al.* 1996) et au moins 5 % du génome du blé (Moore *et al.* 1991). Environ 25 % du génome du maïs est ainsi constitué de seulement 5 familles dominantes de plus de 30000 membres/C. Les copies sont généralement complètement dispersées, même si des regroupements sont parfois observés dans certaines régions du génome (Brandes *et al.* 1997, Heslop-Harrison *et al.* 1997). Enfin, les familles de rétrotransposons viraux de plantes sont souvent plus hétérogènes que celles des champignons ou des insectes et de nombreuses sous-familles d'éléments divergents sont donc généralement observées (Flavell *et al.* 1992b).

*- Rétrotransposons de la classe II (rétroposons)*

Plusieurs familles de rétrotransposons sans LTR ont été identifiées chez les insectes (Finnegan 1989), les mammifères (Hutchison *et al.* 1989) et les plantes (Kubis *et al.* 1998b, Noma *et al.* 1998). Leur importance varie selon les organismes. Les familles des éléments F (4700 pb ou moins) ou I (5400pb) de la drosophile comportent ainsi une cinquantaine de membres/C seulement. Les familles de rétroposons chez les mammifères sont en revanche pour la plupart constituées d'un grand nombre de longues séquences très dispersées (LINE : Long Interspersed Nuclear Elements). La famille LINE-1 est dominante chez la plupart des mammifères. Elle comporte $10^4$ à $10^5$ copies complètes (de 6000 à 7000 pb) ou tronquées à leur extrémité 5' (moins de 6000 pb, souvent les plus nombreuses). Chez les plantes, des familles de rétroposons dominantes analogues aux LINE ont été caractérisées chez *Zea mays* (Cin4, Schwarz-Sommer *et al.* 1987), *Lilium speciosum* (del2, Leeton et Smyth 1993), *Beta vulgaris* (BNR1, Schmidt *et al.* 1995), *Arabidopsis thaliana* (Ta11-1, Wright *et al.* 1996) et *Chlorella vulgaris* (Zepp, Higashiyama *et al.* 1997). Les 250000 copies de la famille del2 représentent ainsi environ 4 % du génome de *L. speciosum*. Il semble que ces familles imposantes sont très courantes chez les plantes (Kubis *et al.* 1998b, Noma *et al.* 1998).

β- Rétroséquences

La plupart des rétroséquences semblent être des copies non fonctionnelles (pseudogènes) transposées de gènes transcrits (rétropseudogènes) qui sont souvent bordées d'une petite duplication du site cible qui témoigne de leur mobilité. Elles dérivent probablement de la transcription inverse des ARN et portent donc les marques de la maturation post-transcriptionnelle. En particulier, certaines séquences adjacentes régulatrices de l'initiation de la transcription sont perdues au cours de la transposition. On parle dans ce cas de pseudogènes maturés par opposition aux copies inactives d'un gène non transposé (les pseudogènes "classique" non maturés). La transposition, et donc la multiplication de ces séquences, semble être passive : l'ARN pourrait être rétro-transcrit grâce à des enzymes fournies en *cis* par des rétrotransposons (ou des rétrovirus). Les rétropseudogènes correspondant à des ARNm (pseudogènes maturés de polypeptides) sont courants chez les mammifères mais rares chez les autres vertébrés, les invertébrés et les plantes. Avec le ou les gènes dont ils dérivent, les rétropseudogènes forment des familles de taille variable mais relativement modeste (1 à 200 copies, **Tableau 9**). En revanche, les familles de rétropseudogènes issus de la transcription inverse de certains autres ARN (ARNt, ARNsn, ARNsc) sont connues chez la plupart des animaux et peuvent représenter plus de 10 % de la taille de certains génomes (Singer et Berg 1992). Les courtes séquences très répétées et dispersées des mammifères, tout d'abord rassemblées sous l'appellation de SINE (Short Interspersed Nuclear Elements) sur la base de leur longueur (100-500 pb) et de leur abondance (jusqu'à $10^6$ copies), sont en fait des pseudogènes maturés (Deininger 1989, Smit 1996) généralement dérivés des gènes codant pour les ARNt ou les petits ARNsc tel que l'ARN 7SL (intégré dans la particule de reconnaissance du signal de translocation des polypeptides au

travers de la bicouche lipidique du reticulum endoplasmique). La famille de SINE *Alu* typique des primates (éléments d'environ 300 pb) est ainsi constituée de $9.10^5$ copies/C chez l'homme (environ 8 % de l'ADN humain) et de $3.10^5$ copies chez le chimpanzé.

**Tableau 9.** Quelques familles de séquences dispersées comportant des gènes et pseudogènes maturés (rétropseudogènes) de polypeptides chez les vertébrés.

| Espèce | Famille | Nombre de gènes fonctionnels | Nombre de rétropseudogènes |
|---|---|---|---|
| Homme | Arginosuccinate synthétase | 1 | 14 |
| | Actine | 6 | 15-25 |
| | β-tubuline | 2-3 | 15-20 |
| | Superoxide dismutase Cu/Zn | 1 | ≥ 4 |
| | Cytochrome c | 2 | 20-30 |
| | Dihydrofolate réductase | 1 | ~ 5 |
| | Tropomyosine non musculaire | 1 | ≥ 3 |
| | Glycéraldéhyde-3-phosphate déshydrogénase | 1 | 10-30 |
| | Phosphoglycérate kinase | 2* | 1 |
| | Protéine ribosomique L32 | 1 | ~ 20 |
| | Triosephosphate isomérase | 1 | 5-6 |
| | Interféron α | 15 | 5 |
| Souris | α-globine | 2 | 1 |
| | Cytokératine endo A | 1 | 1 |
| | Glycéraldéhyde-3-phosphate déshydrogénase | 1 | ≥ 200 |
| | Myosine (chaîne légère) | 1 | 1 |
| | Propiomélanocortine | 1 | 1 |
| | Protéine ribosomique L7 | 1-2 | ≥ 20 |
| | Protéine ribosomique L30 | 1 | ≥ 15 |
| | Protéine ribosomique L32 | 1 | 16-20 |
| | Antigène tumoral p53 | 1 | 1 |
| Rat | α-tubuline | 1-2 | 10-20 |
| | Cytochrome c | 1 | 20-30 |
| Poulet | Glycéraldéhyde-3-phosphate déshydrogénase | 1 | 1 |

D'après Li et Graur (1991) et Singer et Berg (1992).
*Une des deux copies est un rétrogène (copie transposée fonctionnelle).

Le motif répété est lui-même dimérique (clivé en 2 fragments répétés en tandem imparfait de 170 et 120 pb par l'endonucléase *Alu* I) et les deux monomères contiennent respectivement des séquences homologues des régions 5' et 3' de l'ARN 7SL (codé chez l'homme par une famille de 3 ou 4 gènes fonctionnels liés à une centaine de pseudogènes). La zone centrale de l'ARN 7SL (155 pb) est absente des SINE *Alu* mais elle est elle-même répétée de nombreuses fois dans le génome, indépendamment de ces dernières (Ullu et Tschudi 1984). Une famille de SINEs homologue est observée chez les rongeurs (familles de SINE de type I). Elle comporte $10^5$ copies/C d'un motif de 130 pb non dupliqué équivalent d'un des deux monomères typiques des familles *Alu* de primates. Certaines SINE de rongeurs, comme les familles de type II ($8.10^4$ copies/C) et ID ($10^5$ copies/C), dérivent quant à elles probablement des gènes d'ARNt (Daniels et Deininger 1985). Des familles équivalentes ont été

détectées dans les génomes de nombreux organismes y compris les plantes (Okada 1991, Deininger et Batzer 1993). Enfin, les familles de gènes codant pour les petits ARNsn (U1-U6 des métazoaires) incluent souvent un grand nombre de pseudogènes maturés. Par exemple, la famille des ARNsn U1 humains comporte 30 gènes disposés en tandem et environ $10^4$ rétropseudogènes plus ou moins dispersés dans le génome (Lindgren *et al.* 1985).

L'amplification importante des familles de pseudogènes maturés de certains petits ARN (ARNt et ARNsc en particulier) pourrait s'expliquer par un phénomène de transcription en cascade (Bains 1986). En effet, contrairement aux gènes de polypeptides transcrits par l'ARN polymérase II (Pol II) dont les séquences promotrices sont toutes situées dans les zones adjacentes non transcrites, le promoteur principal nécessaire à la transcription par la polymérase III (Pol III) des gènes codant pour les ARNt et les ARNsc est interne. Dans ce cas, le processus de transposition par transcription inverse peut s'effectuer en préservant l'intégrité de la région promotrice et donc permettre la transcription des rétropseudogènes des ARNt et ARNsc. Les éventuels rétropseudogènes transcrits constituent alors autant de nouvelles sources de transposition par transcription inverse. Britten *et al.* (1988) ont ainsi détecté au moins 4 sous-familles de séquences *Alu* dans le génome humain correspondant à 4 sources différentes dérivées initialement des séquences d'ARN 7SL.

c- Autres familles de séquences répétées et dispersées

Il existe probablement bien d'autres familles de séquences répétées et dispersées dans les génomes des eucaryotes. Récemment, une nouvelle catégorie d'éléments mobiles (MITE : Miniature Inverted-repeat Transposable Elements) dont le mécanisme de transposition n'est pas connu a été identifié chez les plantes (Wessler *et al.* 1995) et plusieurs groupes d'animaux (Kidwell et Lisch 1997). Ce sont des éléments de petite taille (100-400 pb) qui forment des familles de 3000 à 10000 membres. Différents autres éléments répétés *a priori* non mobiles et très dispersés ont aussi été signalés chez les plantes (Kubis *et al.* 1998a). Par ailleurs, il est probable que de nombreuses familles de séquences courtes (moins de 100 pb) à très courtes (moins de 10 pb) sont présentes dans la fraction d'ADN hautement répété des génomes eucaryotes (Li et Graur 1991). Enfin, les rétrovirus des mammifères et certains autres virus chez les plantes peuvent aussi occasionnellement se multiplier dans un génome (Charlesworth *et al.* 1994).

**2- Les familles de séquences répétées et groupées en tandem**

La répétition en tandem de séquences d'ADN est une caractéristique commune remarquable des génomes eucaryotes. Le phénomène affecte non seulement les zones non codantes ou de fonction encore inconnue comme les régions centromériques, mais aussi les zones fonctionnelles (gènes des histones et des ARNr, télomères, etc.). Il se traduit par la présence de séries parfaites ou imparfaites de taille parfois considérable (jusqu'à $10^8$ pb). Certains gènes sont eux-mêmes composés d'éléments répétés en tandem (gènes multimériques). Les séries dominantes (~ $10^6$ pb) sont toujours associées à

l'hétérochromatine nucléaire et occupent généralement des positions spécifiques sur les chromosomes telles que les régions péricentromériques ou centromériques, subtélomériques ou télomériques et les constrictions secondaires des régions intercalaires. Elles forment généralement des bandes bien distinctes ("satellites") après fractionnement et séparation de l'ADN génomique par centrifugation à l'équilibre sur gradient de densité de chlorure de césium (centrifugation isopycnique, Singer et Berg 1992) en raison de leur richesse en GC souvent supérieure à la moyenne de l'ADN génomique (40-45 % de GC). L'analyse du polymorphisme de longueur des fragments de restriction (RFLP) de l'ADN génomique permet souvent de déterminer la taille du motif répété et son abondance (Pech *et al.* 1979). Les séries de taille plus réduite (divers microsatellites et minisatellites de l'ADN non codant) sont en revanche plutôt observées dans les régions euchromatiques, parfois en amont des gènes, dans les introns ou espaceurs intergéniques (Singer et Berg 1992, Morgante et Olivieri 1993). Leur détection est plus difficile que celle des séries dominantes et requiert souvent un criblage de l'ADN génomique par hybridation Southern ou hybridation *in situ* à l'aide de sondes marquées.

a- Séries de l'ADN non codant ou de fonction inconnue

Les différentes séries de l'ADN non codant sont généralement classées en trois catégories selon leur taille (Tautz 1993, Charlesworth *et al.* 1994). Il s'agit des microsatellites (moins de 500 pb), des minisatellites (500 à 30000 pb) et des satellites (jusqu'à $10^8$ pb). Les unités répétées en tandem sont généralement très courtes pour les microsatellites (2-6 pb) et les minisatellites (10-40 pb) mais peuvent être beaucoup plus longues dans le cas des satellites (de 2 pb à plusieurs milliers de pb).

α- Microsatellites et minisatellites

Les microsatellites ou SSRs [Simple (Short) Sequence Repeats] ou STRs (Short Tandem Repeats) et dans une moindre mesure les minisatellites ou VNTRs (Variable Number of Tandem Repeats) sont très fréquents chez la plupart des eucaryotes (Tautz 1989, Lagercrantz *et al.* 1993, Wang *et al.* 1994) mais plus abondant chez les animaux (1 microsatellite d'au moins 20 pb tous les 6 kpb en moyenne) que chez les plantes (1 microsatellite tous les 29 à 50 kpb, Lagercrantz *et al.* 1993). Chez l'homme, il y aurait ainsi plus de 30000 loci microsatellites dispersés sur tous les chromosomes (Weber 1990, Dib *et al.* 1996). Chaque séquence microsatellite ou minisatellite a une distribution génomique spécifique qui peut comporter un ou plusieurs sites chromosomiques (séries dispersées). Chez *Beta vulgaris* par exemple, la séquence microsatellite GA est présente à de multiples loci dispersés tandis que la séquence microsatellite GATA n'occupe que 3 loci différents (Kubis *et al.* 1998a). Le nombre de copies cumulé d'une séquence microsatellite ou minisatellite organisée en séries dispersées peut donc représenter une part importante du génome. Par exemple, les séquences microsatellites AAG ($2{,}4.10^6$ copies) et TTAGGG ($2{,}2.10^6$ copies) et la séquence minisatellite ACACAGCGGG ($1{,}2.10^6$ copies) rendent compte à elles seules de plus de 50 % de la taille du génome du rat kangourou *Dipodomys ordii* (Widegren *et al.* 1985). Les plus petits motifs microsatellites (2 pb) sont ubiquistes chez la plupart des eucaryotes (Lagercrantz *et al.* 1993, Wang *et*

*al.* 1994). Le motif AC est particulièrement abondant chez les animaux tandis que chez les plantes les motifs GA et AT semblent prépondérants. Il y aurait au moins $3{,}6.10^4$ microsatellites GA par génome haploïde chez le blé (Röder *et al.* 1995), $4{,}5.10^4$ chez les mélèzes (Volkaert *et al.* 1992) et de l'ordre de $10^4$ à $10^5$ chez les chênes (Dow *et al.* 1995, Steinkellner *et al.* 1997, Isagi et Suhandono 1997). Les microsatellites et minisatellites sont particulièrement instables (hypervariables) et évoluent rapidement en taille au sein des populations (Jeffreys *et al.* 1985, Charlesworth *et al.* 1994). Un motif microsatellite de 14 pb situé 500 pb en amont du gène de l'insuline humaine est ainsi répété de 26 à près de 200 fois selon les individus (Singer et Berg 1992). De même, les différents loci microsatellites GA identifiés et caractérisés chez le chêne sessile sont tous hautement polymorphes (voir Steinkellner *et al.* 1997).

β- Satellites

Contrairement aux microsatellites et minisatellites, les familles d'ADN satellite ont une taille plutôt stable (Charlesworth *et al.* 1994). Elles sont relativement peu nombreuses dans les génomes eucaryotes et souvent associées aux régions centromériques, péricentromériques ou subtélomériques. Kubis *et al.* (1998a) n'ont ainsi détecté que 9 familles distinctes à partir de 5 espèces du genre *Beta*. Au moins 4 d'entre elles occupent des positions péricentromériques ou centromériques (exclusivement ou non). Il s'agit des satellites pBV1 de *B. vulgaris* (répétitions de 327-328 pb contenant un microsatellite AC), pRN1 de *B. nana* (répétitions de 209-233 pb), pTS5 et pTS4.1 de *B. procumbens* (unités de 153-160 pb et 312 pb, respectivement). Les satellites centromériques représentent souvent une part importante de l'ADN génomique et illustrent parfaitement la diversité d'organisation des séries dominantes de l'ADN non codant. L'unique satellite centromérique du crabe terrestre *Geocarcinus lateralis* est ainsi constitué d'une simple répétition homogène du motif dinucléotidique AT et il représente plus de 21 % du génome (Singer et Berg 1992). Les unités répétées sont cependant rarement aussi courtes et homogènes. Un satellite peut ainsi être composé de plusieurs sous-séries (ou domaines) chacune constituée d'un motif répété différent. En outre, il existe souvent plusieurs satellites centromériques variables en taille et en séquence au sein d'une même espèce et entre les espèces, même très proches. Chez *D. melanogaster* par exemple, il y a 4 satellites centromériques différents (1 par chromosome) qui représentent 16 % du génome au total (Singer et Berg 1992). Les satellites I (3 % du génome) et IV (5 % du génome) sont manifestement apparentés et contiennent deux domaines homogènes composés de deux motifs répétés simples et homologues de 5 et 7 pb. Le satellite II (3 % du génome) est une répétition presque parfaite d'un petit motif de 10 pb. Enfin, le motif répété du satellite III (5 % du génome) est beaucoup plus long (359 pb) et très polymorphe aussi bien en longueur (un petit domaine contient des unités de 254 pb seulement) qu'en séquence (polymorphisme de sites de restriction). Les 4 satellites centromériques de l'espèce proche *D. virilis* (environ 19 % du génome) ne comportent en revanche que des répétitions homogènes de 7 pb différentes de celles présentes dans les satellites I et IV de *D. melanogaster*. La situation est encore plus complexe chez les bovins (Pages et Roizes 1984). Sept satellites centromériques ont été isolés et

caractérisés chez le boeuf (23 % du génome). Deux présentent des répétitions non apparentées de 700 pb (0,5 % du génome) et 3800 pb (4,6 % du génome). Cinq autres ont clairement une origine commune car les motifs répétés qu'ils contiennent (46, 1399, 1413, 2348 et 2600 pb) possèdent une ou plusieurs régions homologues. Les motifs satellites de 46 pb (0,1 % du génome) et 1399 pb (5,1 % du génome) sont de simples arrangements en tandem de variants d'une séquence de 23 pb (2 et 61 unités, respectivement). L'insertion d'un fragment de 1200 pb dans la série de 1399 pb est probablement à l'origine du motif satellite de 2600 pb (7,1 % du génome). Le motif satellite de 2348 pb (4,2 % du génome) est quant à lui constitué d'une mosaïque de 4 éléments de 1200, 247, 650 et 251 pb associés dans cet ordre. Les éléments de 1200 et 650 pb contiennent des séquences de 23 pb tandis que les éléments de 247 et 251 pb sont formés d'une série de 8 unités de 31 pb. Enfin, le motif satellite de 1413 pb (1,7 % du génome) contient une série de 24 unités de 23 pb (550 pb) et une série de 8 unités de 31 pb (252 pb) séparées par une séquence de 611 pb (apparentée avec le fragment de 1200 pb inséré dans le motif satellite de 2600 pb). Cette périodicité de séquence à longue distance superposée à une périodicité à plus courte distance (répétitions multimériques) semble caractéristique des ADN satellites (centromériques ou non) de très grande taille (> 200 kpb, Charlesworth *et al.* 1994). Le motif répété de 234 pb du satellite centromérique majeur (8 % du génome, environ $10^6$ copies) de *Mus musculus* contient ainsi 4 répétitions imparfaites de 58-60 pb qui sont chacune composées de deux segments de 28 et 30 pb. Ces segments pourraient aussi avoir évolué à partir d'un motif de base de 9 pb. Chez les primates enfin, les répétitions de la famille des satellites α contiennent toutes un arrangement en tandem de 2 à 16 monomères de 170 pb environ (Warburton *et al.* 1993). Le satellite α présent sur le chromosome X humain porte ainsi environ 5000 unités de 2100 pb constituées de 12 monomères de 170 pb qui divergent entre eux de 10 à 20 %.

b- Séries multigéniques

La répétition en tandem de séquences d'ADN fonctionnelles est omniprésente chez les eucaryotes et constitue des séries imparfaites ou parfaites selon que les unités élémentaires sont ou non interrompues par des segments d'ADN non apparentés. De nombreuses familles multigéniques à répétition de variants, les gènes des petits ARNt et ARNsn, les différents gènes codant pour les histones, les séquences télomériques ou bien encore les gènes codant pour les ARNr (ADN ribosomique ou ADNr) sont concernés par une telle organisation (partiellement ou totalement). Certains gènes ont eux-même une structure interne remarquable constituée de répétitions (gènes multimériques).

α- Gènes multimériques

Les gènes multimériques sont nombreux chez les eucaryotes et constituent un bon exemple démontrant l'intervention des mécanismes d'amplification de séquences en tandem (**Partie II-2-c de ce Chapitre I**) dans la formation de nouveaux gènes au cours de l'évolution (Singer et Berg 1992). La structure de ces gènes peut être totalement ou en partie seulement multimérique et impliquer des

séries interrompues ou ininterrompues d'éléments. Une région du gène du polypeptide α2(I) de collagène de poule d'environ 28 kpb comporte ainsi 52 exons homologues séparés par des introns de 80 pb à plus de 2000 pb (Singer et Berg 1992). Les exons de 45 à 108 pb sont tous multiples d'une séquence de 9 pb et semblent construits par des répétitions en tandem de ce petit motif. De manière analogue, le gène unique du récepteur des lipoprotéines de faible densité (LDL : Low Density Lipoprotein) humaines (18 exons) contient deux séries imparfaites d'exons apparentés (exons 2-6 et exons 7, 8 et 14). Les exons 2 à 6 (environ 840 pb) sont en outre chacun constitués de 7 répétitions en tandem d'une séquence d'environ 120 pb (Brown et Goldstein 1986). Le gène Sgs4 (~ 900-1050 pb) de la protéine d'adhérence des chrysalides de drosophile (Muskavitch et Hogness 1982) comporte quant à lui , près de son extrémité 5', 19 à 27 variants d'une séquence de 21 pb organisés en série parfaite et codant pour 7 acides aminés différents par le jeu de leurs variations. On peut citer enfin le cas des deux gènes apparentés codant pour la sérumalbumine et l'α-foetoprotéine des mammifères (famille multigénique à répétition de variants, voir ci-dessous) qui sont chacun constitués d'une série parfaite de 3 répétitions divergentes (de 564, 576 et 551 pb pour l'α-foetoprotine de souris) qui forment les 3 domaines de la protéine de 580 acides aminés (Ohno 1981).

β- Familles multigéniques à répétition de variants

Les gènes métaboliquement apparentés des eucaryotes (sous-familles et super-familles multigéniques à répétition de variants) sont habituellement très dispersés. Selon les espèces, il arrive cependant qu'ils soient en partie ou totalement regroupés en séries imparfaites au niveau d'un ou plusieurs sites chromosomiques. C'est le cas de plusieurs familles multigéniques codant pour des polypeptides comme les sérumalbumines, les α-amylases, les globines, *etc.* (**Tableau 8**). La superfamille des gènes de globine humaine illustre bien cette situation. Elle comporte 13 membres organisés en deux groupes distincts sur les chromosomes 11 (gènes des globines α) et 16 (gènes des globines β). Les 3 gènes et les 4 pseudogènes des globines α sont séparés par des séquences de quelques centaines à plusieurs milliers de pb et constituent une série imparfaite de 25 kpb, tandis que les 5 gènes et l'unique pseudogène de la globine β occupent une région de 65 kpb environ (Esteal 1988). Une organisation similaire des gènes de globines se retrouve chez d'autres vertébrés et invertébrés et une situation comparable peut aussi être observée dans d'autres familles de gènes apparentées telles que la myoglobine de vertébrés et la leghémoglobine de certaines plantes fixatrices d'azote (Singer et Berg 1992). De même, la super famille de l'hormone de croissance est composée chez l'homme du gène de la prolactine isolé sur le chromosome 6 et d'un groupe de gènes codant pour l'hormone de croissance (2 copies) et le lactogène placentaire (3 copies) disposés en série imparfaite de 50 kpb sur le chromosome 17 (Barsch *et al.* 1983).

χ- familles de gènes des ARNt et ARNsn

Bien que très dispersés chez la plupart des organismes (Long et Dawid 1980), les nombreux gènes de 75 à 90 pb codant pour les différents ARNt (**Tableau 7**) présentent parfois une organisation

en séries parfaites ou imparfaites (Clarckson 1983). Un des nombreux loci de gènes d'ARNt (il y en aurait une cinquantaine) du génome de *D. melanogaster* (la région 42A d'environ 50 kpb) est ainsi composé d'une série imparfaite de 18 gènes codant pour 5 ARNt différents qui sont transcrits à partir des deux brins d'ADN opposés (8 copies dans un sens et 10 dans l'autre). Chez *X. laevis*, les 310 copies d'un gène d'ARNt$^{Met}$ (ARNt spécifique de la méthionine) sont en revanche organisées en une série parfaite d'environ 500 kpb. L'unité répétée de 3200 pb contient non seulement deux gènes d'ARNt$^{Met}$ séparés par un espaceur intergénique de 350 pb mais aussi plusieurs autres gènes d'ARNt (Clarckson 1983). De manière analogue, les gènes de 58 à 220 pb codant pour les petits ARNsn peuvent être totalement ou en partie dispersés chez certaines espèces et groupés en séries parfaites ou imparfaites chez d'autres (**Tableau 10**).

**Tableau 10.** Dispersion dans le génome des membres de quelques familles multigéniques codant pour les petits ARN nucléaires (ARNsn) chez les eucaryotes.

| Famille (longueur des ARNsn en nt) | Organisme | Taille[a] | Dispersion[b] |
|---|---|---|---|
| U1 (164) | Xenope | 500 | T, d |
| | Drosophile | 3-4 | d |
| | Poule | 5 | G |
| | Souris | 10 | G |
| | Homme | 30 | G |
| U2 (188-189) | Oursin | 500 | T |
| | Xenope | 500 | T |
| | Drosophile | 5 | G, d |
| | Homme | 5-40 | T |
| U3 (216) | Souris | 6 | G, d |
| U4 (142-146) | Drosophile | 4 | d |
| | Poule | 2 | G |
| U5 (116-118) | Xenope | ? | T |
| U6 (107-108) | Drosophile | 3 | G |
| | Souris | 200 | d |
| U7 (58) | Oursin | 5 | G |
| 4,5S1 (98-99) | Rat | 1000 | d |

D'après Singer et Berg (1992) et Liao *et al.* (1997).
[a] Nombre approximatif de gènes membres de la famille.
[b] G : gènes groupés en série imparfaite, T : gènes groupés en série parfaite, d : gènes dispersés

δ- Famille de gènes des histones

L'organisation des 5 gènes codant pour les histones H1, H2A, H2B, H3 et H4 est particulièrement complexe. Le nombre de copies de chacun d'entre eux varie de 2 chez la levure à environ 700 chez le triton (Hentchel et Birnstiel 1981, Singer et Berg 1992). Les différents gènes sont généralement groupés en une ou plusieurs séries imparfaites qui constituent elles-mêmes une unité de répétition de taille importante (**Tableau 11**). Chez les oiseaux et les mammifères, ces unités

polygéniques sont relativement peu nombreuses et dispersées sur plusieurs chromosomes (Ruiz-Carillo *et al.* 1983, Tripputi *et al.* 1986). Chez d'autres organismes en revanche, elles constituent une ou plusieurs séries dominantes (Hentchel et Birnstiel 1981). Les 5 gènes d'histones de la drosophile sont ainsi organisés en une seule série parfaite de 100 unités polygéniques de 5 kpb située sur le chromosome 2. De même, les gènes qui assurent la synthèse rapide d'histones pendant les stades embryonnaires très précoces de l'oursin sont associés dans une unité polygénique de 6-7 kpb répétée approximativement 500 fois (série parfaite d'environ $3.10^6$ pb). Chez le xénope, deux unités polygéniques différentes organisées en séries parfaites distinctes ont été décrites (Turner et Woodland 1983). La situation chez le triton est un peu différente (Singer et Berg 1992). Les 700 copies de l'unité polygénique de 9 kpb sont organisées en série imparfaite. Les segments de taille variables (jusqu'à 100 kpb) séparant les unités de gènes d'histones consistent eux-mêmes en de multiples répétitions en tandem d'une séquence non apparentée.

**Tableau 11.** Organisation des groupes de gènes d'histones de différentes espèces.

| Organisme | Arrangement des gènes | Taille du groupe de gènes (kpb) | Nombre de copies | Dispersion des copies[a] |
|---|---|---|---|---|
| Levure | H2A-H2B | 6 | 2 | T |
| | H3-H4 | 13 | 2 | T |
| Drosophile | H3-H4-H2A-H2B-H1 | 5 | 100 | T |
| Oursin | H1-H4-H2B-H3-H2A (gènes précoces) | 6-7 | 500 | T |
| | Les gènes tardifs sont peu nombreux (10 copies par gène) et dispersés. | / | / | / |
| Triton | H4-H2A-H2B-H3-H1 | 9 | 700 | G |
| Xenope | H3-H4-H2A-H2B-H1B-H3-H4 | 16 | 25 | T |
| | H3-H4-H2A-H2B | 6 | ? | T |
| Poule | H3-H2A-H4-H1-H2A-H2B | 14 | 10 | d |
| | H2A-H4-H2A-H2B | > 10 | ? | d |
| | H3-H4 | > 10 | ? | d |
| | Un gène H3 isolé. | / | / | / |
| Homme | H3-H4-H4-H3-H2A-H2B | 20 | 10-20 | d |
| | H4-H3-H1-H2B-H2A | 15 | 5 | d |
| Riz | H2A-H2B-H4 | 6 | ? | ? |

D'après Hentschel et Birnstiel (1981).
[a] G : copies groupées en série imparfaite, T : copies groupées en série parfaite, d : copies dispersées
? : inconnu

ε- Télomères

Les télomères ont une fonction importante de stabilisation des extrémités des chromosomes où ils se trouvent localisés (Zakian 1989). Leur organisation est relativement simple et très conservée entre les espèces. Ce sont des séries parfaites et homogènes de quelques milliers de pb constituées d'un motif répété très court de 2 à 7 pb et riche en G (Cech 1988). Les séquences télomériques sont

similaires chez des organismes aussi différents que les protozoaires ciliés [TTGGGG, Yao *et al.* 1987] et flagellés [TTAGGG, Zakian 1989], les amibes [$AG_{(1-8)}$, Zakian 1989], les levures [$TG_{(1-3)}$, Shampay et Blackburn 1988], l'homme et d'autres vertébrés [TTAGGG, Meyne *et al.* 1989] ou bien encore les plantes [TTTAGGG, Richards et Ausubel 1988, Schwarzatcher et Heslop-Harrison 1991]. Ces séquences sont identiques au sein d'une même espèce et constituent donc ensemble une famille dominante de l'ADN génomique (satellite télomérique). Chez la tomate par exemple, les séries télomériques situées aux extrémités des 12 paires de chromosomes ont chacune une taille de 30 à 60 kpb (Lapitan 1991). Le satellite télomérique représente donc approximativement 1000 kpb soit environ 0,1 % du génome de la tomate (basé sur une taille du génome d'environ 1pg/C, Bennett et Leitch 1995).

ϕ- Familles de gènes des ARNr

Les gènes codant pour les différents ARN ribosomiques sont très répétés (**Tableau 7**) et classiquement organisés en une ou plusieurs séries parfaites (Gerbi 1985, Appels et Honeycutt 1986). Chez la plupart des eucaryotes, les gènes codant pour les ARNr 18S (~ 1900-2300 pb), 5,8S (~ 155-160 pb) et 25-28S (de 3400 pb chez les champignons, les invertébrés et les plantes à plus 5000 pb chez les mammifères) sont groupés (famille des ARNr 18S, 5,8S et 25-28S ou ADNr) et séparés des séries parfaites de gènes codant pour les petits ARNr 5S d'environ 120 pb (famille des ARNr 5S ou ADNr 5S). Une liaison totale ou partielle entre ces 4 gènes a cependant été observée chez certains protozoaires (*Toxoplasma*, *Dictyostelium*), de nombreuses protophytes et algues (ex : *Cryptomonas*, *Fucus*, *Eutreptiella*, etc.), de nombreux oomycètes (ex : *Pythium*, *Phytophtora*, *etc.*) et champignons (ex : *Saccharomyces*, *Coprinus*, *Agaricus*, *Armillaria*, *etc.*), un nématode (*Meloidogyne arenaria*), plusieurs crustacés (ex : *Calanus*, *Temora*, *Balanus*, *etc.*) et quelques araignées du genre *Araneus* (Drouin et Moniz de Sá 1995, Kawai *et al.* 1997). Ils constituent alors, comme chez les procaryotes, une seule et même famille (Srivastava et Schlessinger 1991).

*- Famille des gènes des ARNr 18S, 5,8S et 25-28S (ADNr)*

Les gènes des ARNr 18S, 5,8S et 25-28S sont liés dans cet ordre au sein d'une unité polygénique (série imparfaite) répétée quelques centaines à plusieurs milliers de fois (Long et Dawid 1980, Cortadas et Pavon 1982, **Tableau 7**) en tandem ininterrompu (séries parfaites de l'ADNr). Selon l'espèce, il peut y avoir une ou plusieurs séries d'ADNr (moins d'une dizaine) généralement localisées au niveau de certaines régions chromosomiques morphologiquement distinctes (constrictions secondaires) appelées organisateurs nucléolaires (NOR : Nucleolar Organizer Region). Dans le cas des primates par exemple, il existe un seul site NOR chez le gibbon, 2 chez le gorille, 5 chez l'homme et le chimpanzé et 9 chez l'orang-outang (Long et Dawid 1980). Pendant l'interphase, ces régions sont incorporées au nucléole du noyau, structure dans laquelle les ARNr sont activement transcrits, les ARNr précurseurs (pré-ARNr) transformés et les ribosomes en grande partie assemblés (Alberts *et al.* 1989, Singer et Berg 1992).

L'unité de répétition polygénique possède une taille très variable entre les espèces (9 à 44 kpb, Long et Dawid 1980).

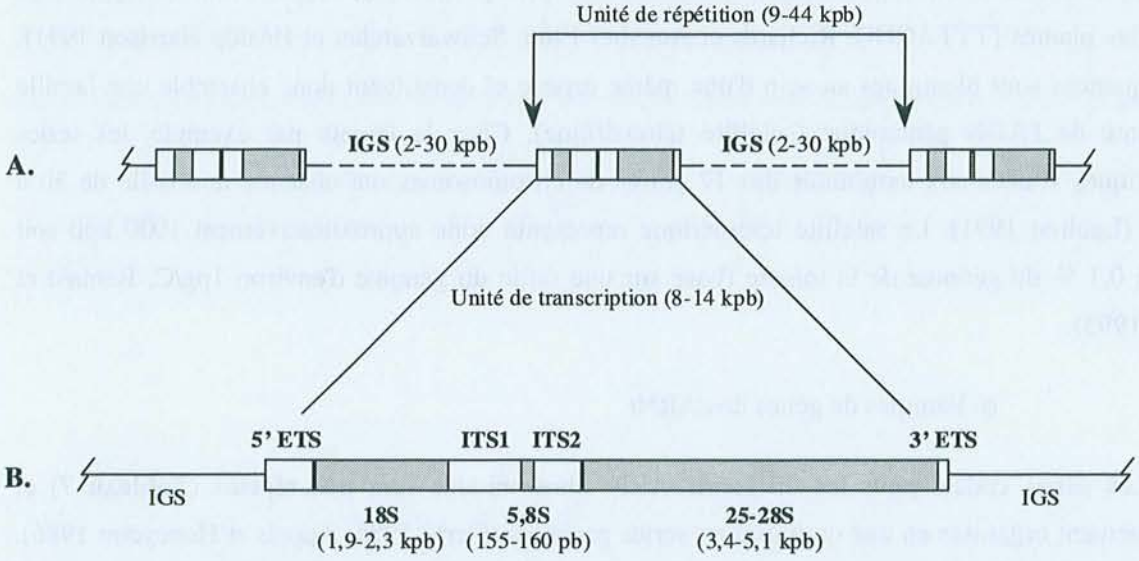



**Figure 3.** Organisation des gènes codant pour les ARNr 18S, 5,8S et 25-28S (ADNr) chez la plupart des eucaryotes supérieurs.
**A.** Les unités répétées en tandem un grand nombre de fois (quelques centaines à plusieurs milliers de copies) sont constituées d'une unité de transcription contenant les gènes d'ARNr 18S, 5,8S et 25-28S (en gris) et d'un espaceur intergénique (IGS) très variable en taille qui comporte de nombreuses répétitions internes en tandem (pointillés).
**B.** Les unités de transcription comportent deux espaceurs externes transcrits (5' ETS et 3' ETS) et deux espaceurs internes transcrits (ITS1 et ITS2) de taille variable (quelques centaines à plusieurs milliers de pb) qui sont éliminés au cours de la maturation de l'unique pré-ARNr en ARNr 18S, 5,8S et 25-28S.

Chez la plupart des eucaryotes supérieurs, elle est formée (**Figure 3**) d'une unité transcrite par l'ARN polymérase I (Pol I) de 7 kpb (levure) à 14 kpb (homme) et d'un espaceur intergénique non transcrit (IGS : InterGenic Spacer) de 2 (levure) à 30 kpb (homme). L'unité de transcription comporte deux espaceurs internes transcrits (ITS : Internal Transcribed Spacer) qui séparent les trois régions codantes (18S, 5,8S et 25-28S) et deux espaceurs externes transcrits (ETS : External Transcribed Spacer) qui précèdent le premier gène (celui de l'ARN 18S) et suivent le dernier (celui de l'ARN 25-28S). Les différences interspécifiques de taille des unités sont principalement dues aux variations localisées dans les espaceurs, en particulier dans l'IGS dont l'hétérogénéité considérable est liée à la présence quasi systématique de multiples répétitions internes en tandem (séries parfaites) de quelques dizaines à plusieurs centaines de pb qui évoluent rapidement en nombre et en séquence (Arnheim 1983, Srivastava et Schlessinger 1991). Chez les plantes, les sous-unités répétées ont une taille comprise entre 100 et 200 pb (Hamby et Zimmer 1992) et elles occupent la partie centrale des IGS (~ 1-10 kpb, Jorgensen et Cluster 1988). Dans le cas de la fève (*Vicia faba*) par exemple, plus de 20 variants de longueur de l'IGS ont été détectés (Yakura *et al.* 1984). Ils sont tous multiples d'un

fragment de 325 pb lui-même constitué d'une répétition de deux sous-unités de 155 pb séparées par 15 pb. Les sous-unités sont souvent apparentées aux séquences voisines des sites d'initiation de la transcription et donc aux promoteurs de la Pol I. Dans l'IGS de xénope (4 kpb environ), la séquence du coeur[2] du promoteur (150 pb) est ainsi répétée jusqu'à 6 fois. Ces copies partielles du promoteur sont séparées par des régions où sont entremêlés plusieurs dizaines d'éléments répétés de 60 et 81 pb apparentés au promoteur (Labhart et Reeder 1986).

*- Famille(s) de gènes des ARNr 5S (ADNr 5S)*

Les gènes codant pour les petits ARNr 5S sont généralement groupés en tandem (au moins partiellement), soit en série(s) parfaite(s) et indépendante(s) de l'ADNr (la plupart des eucaryotes supérieurs), soit intégrés dans l'IGS de l'unité polygénique de l'ADNr (beaucoup d'eucaryotes inférieurs, Drouin et moniz de Sá 1995, Kawai *et al.* 1997). Il existe cependant quelques cas de dispersion individuelle de l'ensemble des copies du gène chez les eucaryotes inférieurs comme la levure *Schizosaccharomyces pombe* (Tabata 1981) et la moisissure *Neurospora crassa* (Selker *et al.* 1981, Mentzenberg *et al.* 1985). Le nombre de copies varie considérablement, d'une centaine chez la drosophile à environ 300000 chez le triton *Notophtalmus viridescens* (Long et Dawid 1980, voir aussi **Tableau 7**).

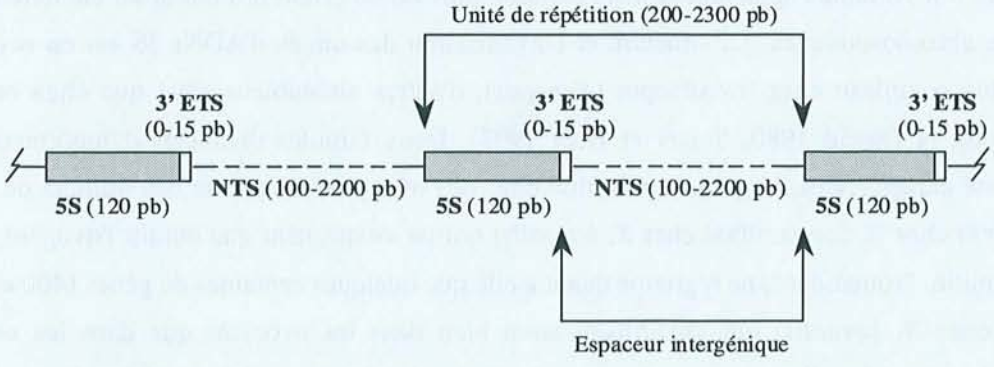



**Figure 4.** Organisation des gènes codant pour les ARNr 5S (ADNr 5S) chez la plupart des eucaryotes supérieurs.
Les unités répétées en tandem un grand nombre de fois (quelques centaines à plusieurs milliers de copies) sont constituées du gène codant pour l'ARNr 5S (en gris) et d'un espaceur intergénique dont une petite partie est parfois transcrite (3' ETS) avant d'être éliminée au cours de la maturation du pré-ARNr 5S. La zone non transcrite de l'espaceur (NTS) est très variable en taille et comporte souvent de nombreuses répétitions internes en tandem (pointillés).

La structure et l'organisation de l'ADNr 5S des eucaryotes supérieurs ont été étudiées chez un grand nombre d'espèces animales (Long et Dawid 1980, Srivastava et Schlessinger 1991, Suzuki *et al.* 1994, Pâques *et al.* 1995) et végétales (voir **Chapitre II**). Il peut y avoir une (ex : drosophile), plusieurs

[2] Séquence essentielle pour l'initiation correcte de la transcription de l'ADNr

(ex : homme) ou de nombreuses séries parfaites (ex : xénope) plus ou moins dispersées dans le génome. La taille des répétitions est généralement comprise entre 200 et 900 pb mais elle peut atteindre 1000 à plus de 2000 pb chez les xénopes et certains mammifères, dont l'homme (Long et Dawid 1980, Srivastava et Schlessinger 1991, Sorensen et Frederiksen 1991). Contrairement à l'ADNr, l'unité répétée d'ADNr 5S a généralement une structure simple (**Figure 4**) qui consiste en une région codante conservée d'environ 120 pb (gène 5S transcrit par la Pol III) associée à un espaceur intergénique en grande partie non transcrit (NTS : Non-Transcribed Spacer) et de taille très variable selon les espèces (100-2700 pb). Chez *D. melanogaster* par exemple, l'unique série d'ADNr 5S localisée sur le bras droit du chromosome 2 est constituée d'environ 160 unités de 366-385 pb comportant un gène de 120 pb et un espaceur de 246-265 pb. Comme dans l'IGS de l'ADNr, le polymorphisme de longueur de cet espaceur est directement lié à la présence d'un nombre variable de petites répétitions internes en tandem d'un motif de 7 pb (Tschudi et Pirotta 1980, Pâques *et al.* 1995). L'organisation de l'ADNr 5S est quelque peu différente chez l'homme. Il existe en effet plusieurs séries de tailles différentes et probablement aussi de nombreux variants du gène dispersés dans le génome (1700-2000 gènes au total, Sorensen et Frederiksen 1991). La série principale est localisée sur le chromosome 1 (comme chez d'autres primates, Long et Dawid 1980) et comporte 90 à 150 unités monogéniques d'environ 2300 pb contenant un microsatellite GT. Des séries mineures composées de 5 à 10 unités apparentées mais de taille plus faible (1600 pb) ont aussi été détectées à d'autres sites chromosomiques. La structure et l'organisation des unités d'ADNr 5S est en revanche beaucoup plus complexe chez les xénopes (*Xenopus*), d'autres amphibiens ainsi que chez certains poissons (Long et Dawid 1980, Singer et Berg 1992). Deux familles distinctes d'importance très inégale ont été caractérisées. La première famille dite "des ovocytes" comporte des milliers de gènes (plus de 20000 chez *X. laevis*, 9000 chez *X. borealis*) qui ne s'expriment que durant l'ovogenèse. La deuxième famille, "somatique", ne regroupe quant à elle que quelques centaines de gènes (400 chez *X. laevis*, 700 chez *X. borealis*) qui s'expriment aussi bien dans les ovocytes que dans les cellules somatiques. Les deux familles sont organisées en de nombreuses séries parfaites qui sont localisées au niveau des régions subtélomériques de la plupart des chromosomes. La structure de l'unité d'ADNr 5S somatique est classique. Elle est monogénique et sa taille varie peu (850-900 pb chez les deux espèces). En revanche, la structure des unités répétées des séries de l'ADNr 5S des ovocytes est complexe (souvent polygénique) et diffère entre espèces, même apparentées. La plupart des unités de *X. laevis* contiennent ainsi un gène et un pseudogène (chacun de 120 pb) en série parfaite et un espaceur de longueur variable (360 à 760 pb) en raison de la présence de petites répétitions internes de 15 pb. Chez *X. borealis*, les unités de 1000 à 2000 pb comportent plusieurs exemplaires d'un gène (en général deux) et un pseudogène séparés les uns des autres par un espaceur de 80 pb. Ces petites séries imparfaites alternent avec de longs espaceurs dont la longueur dépend du nombre de répétitions d'un motif de 21 pb.

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

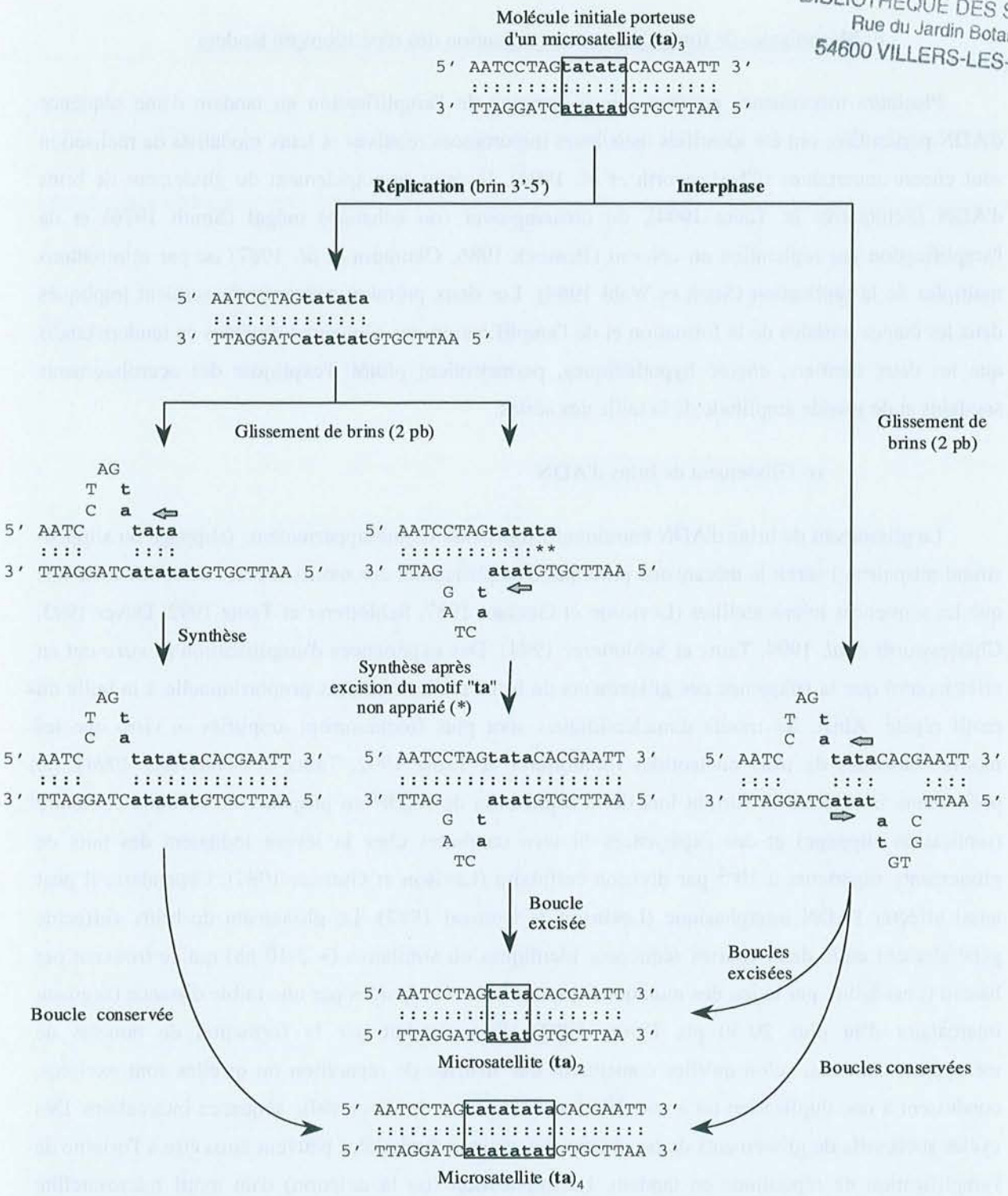



**Figure 5.** Duplication en tandem (ou délétion) d'un motif microsatellite dinucléotidique induite par des glissements de brins d'ADN au cours de la réplication ou à l'interphase. D'après Levinson et Gutman (1987).
Les glissements représentés ici (flèches grises) se produisent entre deux motifs dinucléotidiques "ta" successifs et se concrétisent par l'apparition d'une ou de plusieurs boucles de mésappariement. Lorsque le glissement se produit dans le sens 3'-5' au cours de la réplication et que la boucle de mésappariement sert de matrice de réparation, il y a formation d'une duplication. S'il se produit dans l'autre sens et que la boucle de mésappariement est excisée il y a formation d'une délétion. Dans l'ADN interphasique, le glissement engendre une délétion ou une duplication selon que les boucles de mésappariement sont ou non conservées.

c- Mécanismes de formation et d'amplification des répétitions en tandem

Plusieurs mécanismes pouvant être à l'origine de l'amplification en tandem d'une séquence d'ADN particulière ont été identifiés mais leurs importances relatives et leurs modalités de réalisation sont encore incertaines (Charlesworth *et al.* 1994). Il s'agit principalement du glissement de brins d'ADN (Schlötterer et Tautz 1994), du crossing-over (ou échange) inégal (Smith 1976) et de l'amplification par réplication en cerceau (Bostock 1986, Okumura *et al.* 1987) ou par réinitiations multiples de la réplication (Stark et Wahl 1984). Les deux premiers mécanismes seraient impliqués dans les étapes initiales de la formation et de l'amplification des séquences répétées en tandem tandis que les deux derniers, encore hypothétiques, permettraient plutôt d'expliquer des accroissements soudains et de grande amplitude de la taille des séries.

α- Glissement de brins d'ADN

Le glissement de brins d'ADN entraînant l'apparition de mésappariements (slippage ou slipped-strand mispairing) serait le mécanisme principal d'amplification des motifs répétés de petite taille tels que les séquences microsatellites (Levinson et Gutman 1987, Schlötterer et Tautz 1992, Dover 1993, Charlesworth *et al.* 1994, Tautz et Schlötterer 1994). Des expériences d'amplification *in vitro* ont en effet montré que la fréquence des glissements de brins est inversement proportionnelle à la taille du motif répété. Ainsi, les motifs dinucléotidiques sont plus fréquemment amplifiés *in vitro* que les motifs constitués de trois nucléotides (Schlötterer et Tautz 1992, Tautz et Schlötterer 1994). Le phénomène interviendrait surtout lors de la réplication de l'ADN en prophase de division cellulaire (replication slippage) et des expériences *in vivo* conduites chez la levure indiquent des taux de glissements supérieurs à $10^{-4}$ par division cellulaire (Levison et Gutman 1987). Cependant, il peut aussi affecter l'ADN interphasique (Levinson et Gutman 1987). Le glissement de brins s'effectue généralement entre deux courtes séquences identiques ou similaires (~ 2-10 pb) qui se trouvent par hasard (c'est-à-dire par le jeu des mutations) juxtaposées ou séparées par une faible distance (segment intercalaire d'au plus 20-30 pb, Dover 1993). Il se traduit par la formation de boucles de mésappariement qui, selon qu'elles constituent une matrice de réparation ou qu'elles sont excisées, conduisent à une duplication ou à une délétion du motif et de l'éventuelle séquence intercalaire. Des cycles successifs de glissements de brins dans ces régions dupliquées peuvent ainsi être à l'origine de l'amplification de répétitions en tandem. La duplication (ou la délétion) d'un motif microsatellite dinucléotidique (AT) est illustrée dans la **Figure 5**. Lorsque le glissement se produit dans le sens 3'-5' au cours de la réplication et que la boucle de mésappariement est conservée dans la molécule finale, il y a synthèse d'une duplication. Dans l'autre sens, il y a délétion si la boucle est excisée. Un mécanisme similaire peut affecter l'ADN interphasique. Il y a création de deux boucles de mésappariement (une sur chaque brin) qui induisent la formation d'une délétion ou d'une duplication selon qu'elles sont ou non excisées.

**Figure 6.** Amplification en tandem (ou délétion) d'une séquence d'ADN par crossing-over inégal entre deux chromatides soeurs. D'après Smith (1976).
Les bases nucléotidiques issues de la réplication du brin supérieur de la molécule d'ADN initiale (**A**) sont indiquées en majuscules et en gras. Les bases nucléotidiques provenant du brin inférieur sont figurées en minuscules. L'appariement local entre deux régions similaires mais non alléliques de chaque chromatide soeur initie le crossing-over inégal (**B**). Les deux molécules produites par la recombinaison réciproque sont complexes et contiennent de courtes régions hétéroduplexes impliquant un brin de chaque molécule parentale. Selon la position des points de rupture des quatre brins, ces régions hétéroduplexes peuvent contenir des paires de bases mal appariées, comme celle représentée ici dans une des molécules recombinantes (**C**). Les 4 molécules obtenues à la réplication suivante (**D**) semblent provenir d'une simple rupture et réunion à des points de crossing-over uniques (théoriques) entre les deux régions similaires des chromatides soeurs (flèches grises). Les deux molécules finales les plus longues contiennent une duplication en tandem (rectangles) située entre leur points de crossing-over théoriques respectifs. Les deux molécules finales les plus courtes contiennent des délétions.

Comme il n'implique généralement que de courtes séquences répétées en tandem et qu'il n'affecte qu'un faible nombre de répétitions à la fois (une ou deux, rarement plus, Di Rienzo *et al.* 1993), le glissement de brins ne provoque que de faibles variations de longueur à chaque étape. Pour un microsatellite donné, la distribution des fréquences des séries alléliques de différentes tailles dans une population suivent d'ailleurs relativement bien le modèle de mutation par étape progressive (stepwise mutation model, Ohta et Kimura 1973) qui implique des évènements d'amplification (ou de délétion) de faible amplitude (Charlesworth *et al.* 1994). Ce n'est pas le cas en revanche des loci qui comportent généralement des motifs répétés plus longs tels que les minisatellites et satellites. Le glissement de brins ne joue donc probablement qu'un rôle mineur dans l'évolution de ces séries (Charlesworth *et al.* 1994). Toutefois, nous avons vu précédemment que les plus grandes répétitions en tandem sont elles-mêmes souvent partiellement ou totalement constituées de sous-unités répétées de taille plus modeste. Le glissement de brins pourrait donc être important dans les étapes initiales de la formation de certaines de ces unités multimériques en provoquant des duplications d'au plus quelques dizaines de pb qui peuvent être ensuite amplifiées par d'autres mécanismes tel que le crossing-over inégal (Levinson et Gutman 1987, Stephan 1989).

β- Crossing-over inégal

Beaucoup de répétitions en tandem de taille très variable (de quelques pb à plusieurs centaines de pb) résulteraient de crossing-over inégaux (Smith 1976). Le crossing-over inégal est un processus de recombinaison complexe, globalement réciproque entre deux segments similaires non alléliques[3] (crossing-over ectopique, Jinks-Robertson et Petes 1993) d'une même molécule d'ADN, de deux chromatides soeurs, de deux chromosomes homologues ou de deux chromosomes non homologues, intervenant au cours de la mitose ou de la méiose. Sa réalisation requiert seulement la présence de petites régions présentant une similarité de séquence minimale (quelques pb) qui autorise un alignement non allélique des molécules cibles parentales (Okumura *et al.* 1987). Le crossing-over qui se réalise éventuellement au niveau d'un tel alignement illégitime produit des molécules recombinantes de tailles inégales qui présentent, soit une duplication en tandem du segment d'ADN situé approximativement entre les deux régions impliquées dans l'alignement, soit une délétion. Un exemple de crossing-over inégal se produisant entre deux chromatides soeurs est représenté dans la **Figure 6**. L'alignement illégitime des deux chromatides est lié à la présence de deux régions similaires de 7 pb (seulement une base différente entre les deux motifs) séparées par un segment de 14 pb. Selon la position des points de rupture et réunion des 4 brins d'ADN (ils peuvent être différents, Singer et Berg 1992) et leur nombre (un crossing-over est souvent multiple[4]), différentes molécules

[3] Localisés à des loci différents. Par opposition avec la recombinaison homologue classique qui se produit entre régions alléliques d'un même locus.

[4] Les molécules recombinantes issues d'un crossing-over multiple sont produites par plusieurs ruptures et réunions de brins à des points différents. Un crossing-over multiple est donc équivalent à plusieurs crossing-over simples (ruptures et réunions en un seul point) qui se produiraient successivement.

recombinantes peuvent être obtenues par échange plus ou moins complètement réciproque au niveau de ces deux petits motifs. Elles présentent souvent une courte région hétéroduplexe (dans laquelle un brin provient de chacune des deux molécules parentales) pouvant comporter des mésappariements de bases comme celui représenté dans la molécule la plus longue. Après un nouveau cycle de réplication, les 4 molécules finales semblent chacune issue d'une simple rupture et réunion en un point de crossing-over unique. La duplication en tandem de 21 pb (segment de 14 pb + motif de 7 pb) portée par les molécules les plus longues est située entre leurs points de crossing-over apparents respectifs. Les molécules les plus courtes sont quant à elles délétées de ces mêmes 21 pb.

Un segment d'ADN porteur ou non d'un ou plusieurs gènes peut facilement accumuler des mutations dans ses parties qui ne sont pas (ou peu) sous pression de la sélection naturelle (introns, espaceurs intergéniques, etc.). Cela conduit à l'apparition au hasard de petites régions similaires comme celles décritent dans la **Figure 6**. Selon leur position, un premier crossing-over inégal entre deux de ces régions peut provoquer la duplication (ou la délétion) de tout ou partie du segment initial qui se trouve alors complètement désorganisé. Les séquences plus ou moins longues ainsi dupliquées fournissent ensuite de nombreuses régions homologues susceptibles d'être à leur tour la cible de nouveaux crossing-over inégaux. Des simulations informatiques indiquent clairement que des cycles successifs de crossing-over inégaux au hasard dans ces régions répétées en tandem peuvent rapidement conduire à l'expansion (ou à la contraction) de longues séries parfaites ou imparfaites (Smith 1976, Stephan 1989). Les possibilités d'alignements illégitimes des longues séries deviennent progressivement considérables et la probabilité de réalisation de crossing-over inégaux dans ces zones augmente. Il en résulte que les duplications de séquences par crossing-over inégal se traduisent surtout par l'expansion des séries existantes et ne contribuent que plus rarement à l'amplification de nouvelles familles de séquences répétées en tandem. Les simulations réalisées indiquent en outre que l'expansion des séries est favorisée lorsque la fréquence des crossing-over est faible et que les contraintes sélectives sur la taille des séries ne sont pas trop importantes (Charlesworth *et al.* 1986). Cela pourrait expliquer pourquoi les séries dominantes sont souvent juxtaposées dans les régions télomériques et centromériques (Lapitan 1991, **Partie II-2 a et b de ce Chapitre I**) où les événements de recombinaison sont très rares tandis que les minisatellites s'accumulent plutôt dans les régions euchromatiques qui se recombinent plus fréquemment (Charlesworth *et al.* 1986).

Contrairement au glissement de brins, le crossing-over inégal peut produire des répétitions de taille relativement importante (Li et Graur 1991) et provoquer des amplifications (ou des contractions) de plusieurs unités par cycle estimées à 10-20 % de la longueur totale d'une série (Dover 1982), même si le nombre d'unités impliquées semble parfois relativement faible (estimé au plus à 6-8 unités par cycle dans l'ADNr 5S de la drosophile soit 3 à 6 % de la taille d'une série, Samson et Wegnez 1988). Ce mécanisme est donc souvent évoqué pour expliquer la formation et l'expansion des minisatellites, des satellites ou des séries de l'ADNr (Smith 1976, Warburton *et al.* 1993, Stephan et Cho 1994). Les unités les plus longues sont elles-mêmes très souvent multimériques (cas de nombreux satellites

centromériques comme le satellite α humain) ou développent de nombreuses répétitions internes en tandem (comme l'IGS de l'ADNr). L'émergence, au fur et à mesure que les unités s'allongent, de telles structures complexes est en accord avec les prédictions des modèles d'amplification par crossing-over inégal testés par simulation (Smith 1976, Stephan 1989, Stephan et Cho 1994). Considérons par exemple une série de périodicité initiale courte (AAAA, etc.). Une périodicité plus longue peut facilement se développer par le jeu des mutations et des crossing-over inégaux intervenant au hasard dans la série. Les mutations peuvent en effet conduire à une hétérogénéité locale de plusieurs unités contiguës de la série initiale (AABCDAAA par exemple). Cette région hétérogène (BCD) peut alors être dupliquée puis amplifiée de manière aléatoire par quelques cycles de crossing-over inégaux et constituer ainsi une sous-série dont la répétition est multimérique (AABCDBCDBCDAAA, etc.). Cette périodicité d'ordre supérieur se stabilise lorsque les mutations ont suffisamment réduit l'homologie entre les sous-unités de la nouvelle répétition. Les crossing-over inégaux se produisent alors préférentiellement entre les grandes unités multimériques (BCD) qui se propagent plus vite que les unités courtes initiales (A). Plus les séquences amplifiées au cours des premiers cycles de crossing-over inégaux sont de grande taille plus elles sont instables car la probabilité qu'elles renferment des petits motifs similaires augmente. Les crossing-over inégaux qui se produisent au niveau de ces petits motifs génèrent donc soit des répétitions en tandem internes aux unités, soit des unités de tailles différentes (A et B) qui peuvent ensemble constituer une nouvelle répétition dimérique (ABAB, etc.). Comme dans le cas de l'amplification de la taille des séries (**paragraphe précédent**), les simulations indiquent qu'un taux de recombinaison faible favorise la formation des unités multimériques (Charlesworth *et al.* 1986).

Finalement, la taille et la structure des répétitions dépendent essentiellement des paramètres qui gouvernent la réalisation des crossing-over inégaux et en premier lieu des caractéristiques des petites régions impliquées dans les alignements illégitimes (taille, séquence, espacement). Les simulations de Smith (1976) indiquent ainsi que plus les régions de similarité potentiellement affectées par des crossing-over inégaux sont grandes, plus elles ont de chance d'être distantes l'une de l'autre et plus les répétitions produites ont tendance à être de taille importante et/ou multimériques.

### χ- Amplification par réplication en cerceau ou par réinitiations multiples

Glissement de brins et crossing-over inégal sont des mécanismes relativement progressifs (amplification de quelques unités à la fois) qui requièrent de nombreux cycles de division cellulaire pour engendrer des séries dominantes. Ils ne peuvent donc que très difficilement expliquer les expansions (ou les contractions) soudaines de grande amplitude que certaines séries subissent au cours des divisions cellulaires. Ils ne peuvent pas rendre compte non plus des amplifications de séries apparentées qui se produisent à des sites chromosomiques éloignés de la série parentale (Singer et Berg 1992). D'autres mécanismes mieux adaptés à la compréhension des processus d'amplification non progressifs et de dispersion de séries apparentées ont donc été proposés.

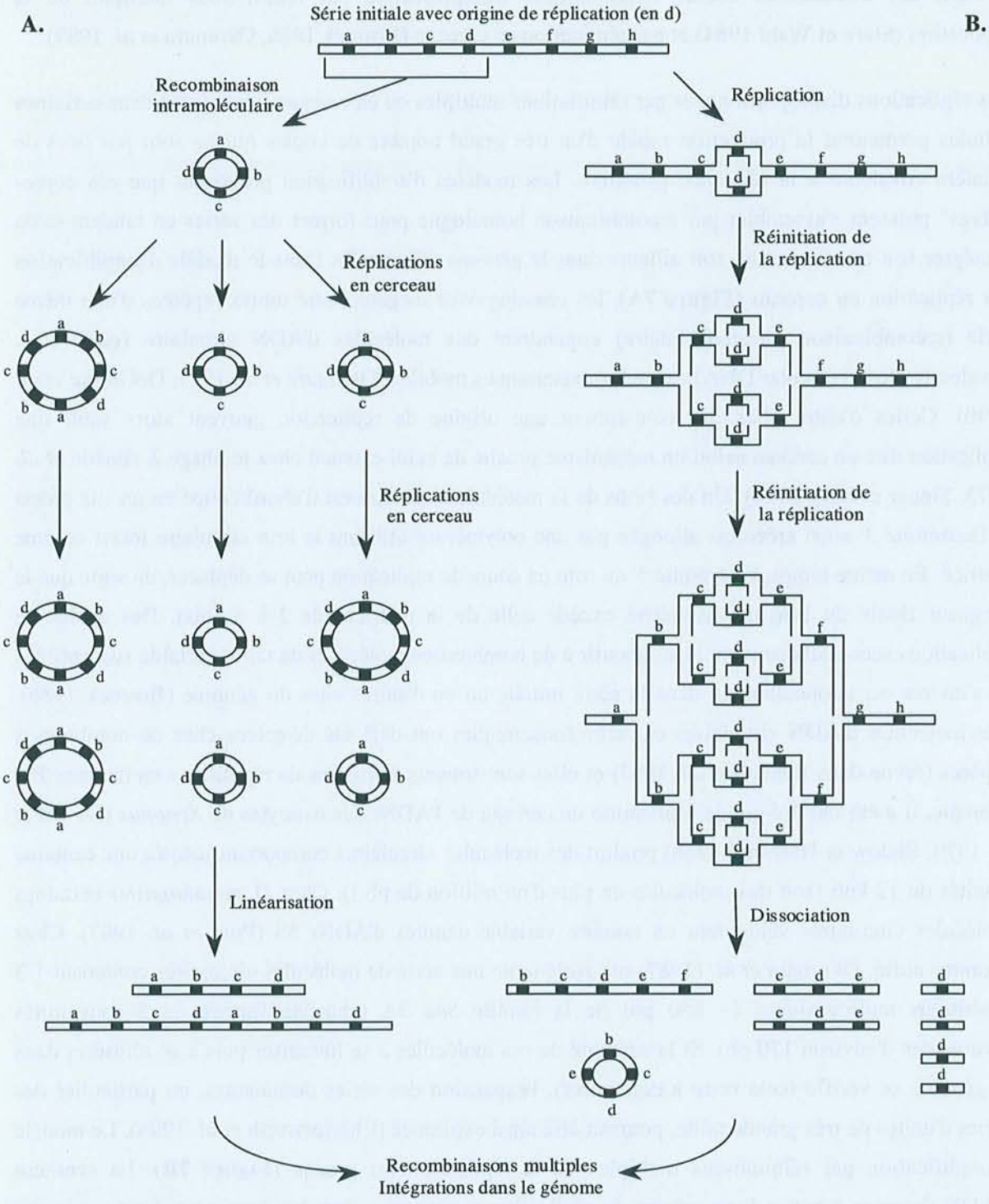



**Figure 7.** Deux modèles d'expansion des séries par réplication disproportionnée.
**A.** Amplification par réplication en cerceau. D'après Bostock (1986). Un crossing-over inégal entre deux éléments répétés de la série initiale produit une molécule circulaire qui, par cycles successifs de réplication en cerceau, engendre des molécules circulaires de taille variable.
**B.** Amplification par réinitiations multiples de la réplication. D'après Stark et Wahl (1984). La mise en route de vagues successives de réplication bidirectionnelle à partir d'une même origine dans la série initiale (représentée ici au niveau de l'élément d) engendre des molécules linéaires de taille variable qui peuvent éventuellement se circulariser par recombinaison intramoléculaire.
Dans les deux modèles, les molécules produites sont susceptibles de se recombiner et de s'intégrer au génome, soit dans la série initiale qui se trouve ainsi amplifiée, soit en d'autres loci où elles constituent de nouvelles séries.

Il s'agit des mécanismes encore hypothétiques d'amplification par réinitiations multiples de la réplication (Stark et Wahl 1984) et par réplication en cerceau (Bostock 1986, Okumura *et al.* 1987).

Les réplications disproportionnées par réinitiations multiples ou en cerceau intervenant dans certaines cellules permettent la production rapide d'un très grand nombre de copies qui ne sont pas liées de manière covalente à la séquence primitive. Les modèles d'amplification proposent que ces copies "libres" puissent s'assembler par recombinaison homologue pour former des séries en tandem et/ou s'intégrer soit au site initial, soit ailleurs dans le génome (**Figure 7**). Dans le modèle d'amplification par réplication en cerceau (**Figure 7A**), les crossing-over inégaux entre unités répétées d'une même série (recombinaison intramoléculaire) engendrent des molécules d'ADN circulaire (cccDNA : covalently closed circular DNA) extrachromosomiques mobiles (Okumura *et al.* 1987, DeGroote *et al.* 1990). Celles d'entre elles qui contiennent une origine de réplication peuvent alors subir une réplication dite en cerceau selon un mécanisme proche de celui existant chez le phage λ (Bastia *et al.* 1975, Singer et Berg 1992). Un des brins de la molécule circulaire est d'abord coupé en un site précis et l'extrémité 3' ainsi créée est allongée par une polymérase utilisant le brin circulaire intact comme matrice. En même temps, l'extrémité 5' du brin en cours de réplication peut se déplacer, de sorte que la longueur finale du brin néosynthétisé excède celle de la matrice (de 2 à 5 fois). Des cycles de réplications successifs peuvent ainsi aboutir à de nombreuses molécules de taille variable susceptibles de s'insérer occasionnellement dans la série initiale ou en d'autres sites du génome (Bostock 1986). Des molécules d'ADN circulaires extrachromosomiques ont déjà été détectées chez de nombreuses espèces (revue dans Renault *et al.* 1993) et elles sont souvent porteuses de répétitions en tandem. Par exemple, il a été montré que la réplication en cerceau de l'ADNr des ovocytes de *Xenopus* (Dawid *et al.* 1970, Endow et Hatwood 1988) produit des molécules circulaires comportant jusqu'à une centaine d'unités de 12 kpb (soit des molécules de plus d'un million de pb !). Chez *D. melanogaster* certaines molécules circulaires véhiculent un nombre variable d'unités d'ADNr 5S (Pont *et al.* 1987). Chez l'homme enfin, Okumura *et al.* (1987) ont isolé toute une série de molécules circulaires contenant 1-3 répétitions multimériques (~ 850 pb) de la famille *Sau* 3A (chacune formée de 5 sous-unités divergentes d'environ 170 pb). Si la capacité de ces molécules à se linéariser puis à se réinsérer dans le génome se vérifie (cela reste à démontrer), l'expansion des séries dominantes, en particulier des séries d'unités de très grande taille, pourrait être ainsi expliquée (Charlesworth *et al.* 1994). Le modèle d'amplification par réinitiations multiples de la réplication est proche (**Figure 7B**). La synthèse d'ADN s'amorce à partir d'une origine de réplication et progresse dans les deux sens. Lorsque le site d'initiation est de nouveau disponible, une nouvelle vague de réplication bidirectionnelle débute au même site et ainsi de suite. Un grand nombre de copies "emboîtées" de molécules de différentes tailles sont ainsi engendrées (modèle en "peau d'oignon"). Ce type de réplication disproportionnée a par exemple été observé dans les cellules folliculaires qui entourent l'ovocyte chez *D. melanogaster* et qui sont responsables de la synthèse des différentes protéines du chorion (enveloppe de l'oeuf). Il n'existe qu'une copie de chacun des gènes de ces protéines mais environ 18 h avant le début de leur expression, ils subissent une amplification en peau d'oignon d'environ 15 fois (Spradling et Orr-

Weaver 1987). Selon ce modèle, des recombinaisons multiples entre copies libérées de ce système complexe peuvent finalement former des séries et/ou des molécules circulaires extrachromosomiques de tailles variables susceptibles de s'intégrer dans le génome (Stark et Wahl 1984).

### 3- Evolution des familles de séquences répétées

La plupart des familles de séquences répétées présentent une homogénéité intraspécifique remarquable (revue dans Elder et Turner 1995) et il est difficile d'imaginer qu'elle soit le résultat d'une sélection agissant individuellement sur chaque membre de la famille. C'est particulièrement évident dans le cas des familles de séquences non fonctionnelles et, concernant les familles composées de nombreuses séquences fonctionnelles identiques, l'altération de quelques copies est probablement sans conséquence sur le métabolisme cellulaire d'un individu. Avec le temps, les unités répétées devraient donc accumuler des mutations, en particulier dans les régions moins importantes comme les introns ou les espaceurs intergéniques. Bien qu'elle puisse parfois simplement correspondre à de récents évènements d'amplification (Liao 1999), cette homogénéité traduit donc plutôt l'existence d'interactions entre les répétitions d'une famille qui tend, au niveau d'un génome, à les rendre toutes semblables. Ce phénomène correspond à ce qui est connu dans la littérature comme étant le processus d'évolution concertée des membres d'une famille de séquences répétées.

#### a- Mise en évidence du processus d'évolution concertée

Brown *et al.* (1972) furent les premiers à montrer que les membres d'une famille de séquences répétées n'évoluent généralement pas de manière indépendante. En comparant les cartes de restriction des unités d'ADNr de *X. laevis* et *X. borealis* (xénopes africains), ces auteurs révélèrent que les régions codantes (en particulier les gènes des ARNr 18S et 28S) sont très conservées en taille et en séquence tandis que les espaceurs intergéniques (IGS) sont au contraire extrêmement divergents. Ces premières observations indiquaient donc clairement que les espaceurs intergéniques ne sont pas (ou très peu) sous pression de la sélection (contrairement aux gènes) et qu'ils évoluent rapidement par fixation aléatoire des mutations (dérive génétique au hasard ou dérive neutre). En même temps et paradoxalement, Brown *et al.* (1972) constatèrent que les différentes copies des IGS étaient étonnamment semblables aussi bien au sein d'un individu, qu'entre les individus de chaque espèce de xénope. Ils en conclurent que cette remarquable uniformité intraspécifique était liée à l'existence de mécanismes moléculaires d'interactions entre les unités d'ADNr qui permettent à une mutation de se propager rapidement à l'ensemble des membres de la famille (évolution ou transfert horizontal[5]) avant même que de nouvelles mutations n'interviennent dans les unités. Cette homogénéisation intragénomique des répétitions d'une famille ou évolution concertée (Zimmer *et al.* 1980) qui conduit

[5] Par opposition à l'évolution verticale qui correspond à la propagation d'une mutation par reproduction sexuée dans une population.

à l'homogénéité intraspécifique (**Figure 8A**) s'est par la suite avérée être un phénomène biologique majeur oeuvrant chez la plupart des êtres vivants (des procaryotes aux eucaryotes).

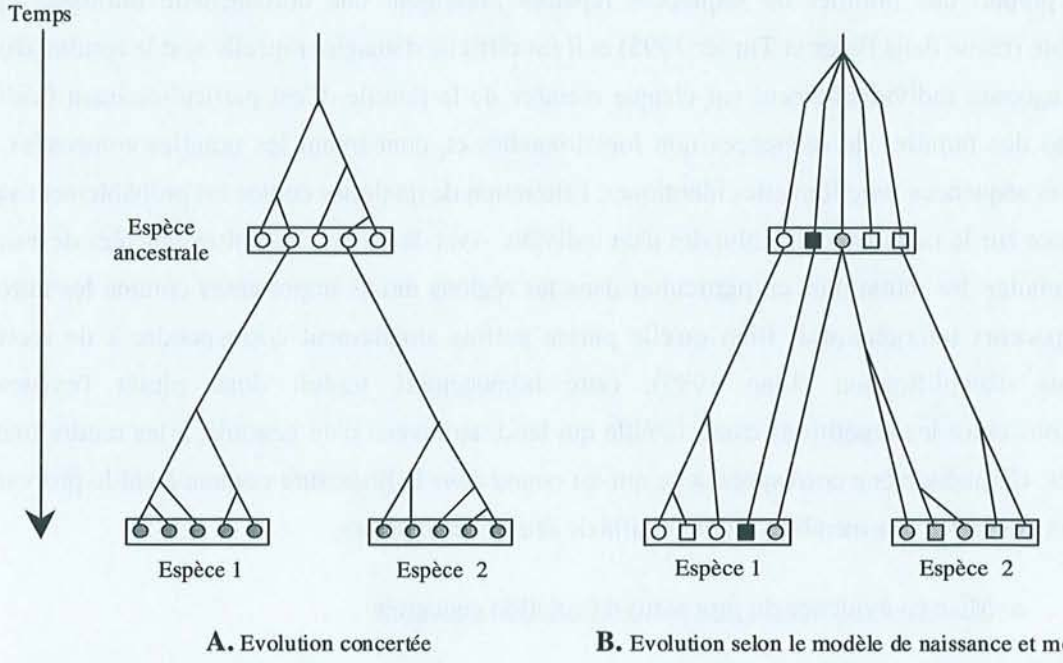



**Figure 8.** Deux modèles d'évolution des familles multigéniques. Adapté de Nei *et al.* (1997).
Dans le modèle d'évolution concertée (**A**), l'homogénéité intraspécifique est liée à l'existence de mécanismes intragénomiques d'interaction entre les différents gènes dupliqués de la famille. L'évolution selon le modèle de naissance et mort (**B**) aboutit à une plus grande hétérogénéité intraspécifique. Les gènes dupliqués peuvent diverger et donner naissance à des groupes de gènes fonctionnels ou de pseudogènes plus ou moins homogènes.

o Gène fonctionnel □ Pseudogène

b- Importance et fonction éventuelle de l'évolution concertée

L'évolution concertée a été mise en évidence dans de nombreuses familles de séquences répétées en tandem ainsi que dans des familles de séquences répétées dispersées (revues dans Ohta 1980, Dover 1982, Arnheim 1983, Elder et Turner 1995, Liao 1999). Très efficace dans les familles multigéniques à répétition de dosage comme l'ADNr (Arnheim *et al.* 1980, Schlötterer et Tautz 1994), l'ADNr 5S (Cronn *et al.* 1996), la famille des gènes des ARNsn U2 (Liao *et al.* 1997), de l'ubiquitine (Nenoi *et al.* 1998) ou des histones (Coen *et al.* 1982), elle semble en revanche plus ou moins inopérante dans certaines familles multigéniques à répétition de variants. Certains résultats tendent en

effet à montrer que le processus d'homogénéisation aurait pu s'exercer précocément, dans les premières étapes de la genèse de ces familles à répétition de variants, avant de s'estomper progressivement au cours de leur histoire évolutive. Par exemple, les gènes de la super-famille des globines de plusieurs mammifères évoluent actuellement de manière indépendante bien qu'ils aient apparemment subi une évolution concertée dans la phase initiale de leur amplification (Goodman 1999). Dans d'autres cas comme ceux des familles multigéniques du complexe majeur d'histocompatibilité et des immunoglobulines des vertébrés, l'analyse des séquences disponibles (Nei *et al.* 1997) a conduit à des conclusions incompatibles avec le modèle d'évolution concertée jusque là évoqué par d'autres auteurs (Parham et Ohta 1996, Yun *et al.* 1997). Ces familles évolueraient plutôt selon un processus dit de "naissance et mort" (birth-and-death process, Nei *et al.* 1997) au cours duquel les gènes se dupliquent, divergent et se maintiennent sous forme de variants fonctionnels ou non fonctionnels (pseudogènes), ou bien sont éliminés (**Figure 8B**). Ce modèle présente l'avantage de rendre compte de la persistance de nombreux pseudogènes (voir les **Tableaux 8 et 9**) et des différences de séquence parfois considérables entre les différents groupes de gènes de ces familles à répétition de variants. Des observations similaires ont été réalisées au niveau de la fraction répétée dispersée de l'ADN non codant. Ainsi, certains éléments *Alu* (SINE) ou rétrotransposons de classe I chez les primates sont ainsi très divergents et ne semblent pas évoluer de manière concertée (Koop *et al.* 1986, Liao *et al.* 1998).

L'évolution concertée des familles de séquences répétées n'est donc pas un phénomène universel. Cependant, le fait que malgré son effet homogénéisateur défavorable à l'émergence de nouveaux gènes par duplication et divergence (Ohno 1970), elle s'exerce de manière très efficace dans certaines familles multigéniques (familles à répétition de dosage) chez de nombreux organismes, conduit à penser qu'elle puisse avoir une fonction biologique et qu'elle soit donc aussi soumise à une certaine pression de la sélection naturelle. Liao (1999) a par exemple suggéré que l'évolution concertée pourrait permettre à la cellule de contrôler la qualité des composants élémentaires nécessaires à la structure et au bon fonctionnement de complexes macromoléculaires importants comme les ribosomes ou les nucléosomes. Il est toutefois intéressant de remarquer que les membres des familles multigéniques à répétition de dosage sont typiquement organisés en un petit nombre de séries parfaites tandis que les familles multigéniques à répétition de variants présentent généralement un profil de dispersion beaucoup plus important (séries imparfaites, plusieurs loci). Les mécanismes moléculaires sous-jacents à l'évolution concertée pourraient donc simplement agir plus facilement sur des répétitions groupées que sur des répétitions dispersées (Ohta et Dover 1983).

c- Mécanismes de l'évolution concertée

L'évolution concertée d'une famille de séquences répétées suppose de multiples interactions entre ses différents membres pendant des périodes de temps considérables (nombreuses générations). Chez les animaux, elle s'accomplit essentiellement dans la lignée cellulaire germinale qui est très tôt

déterminée au cours du développement. En revanche, chez les plantes, elle intervient sans doute dès la phase végétative car les cellules germinales ne se différencient que tardivement à partir des cellules somatiques (Swoboda *et al.* 1994).

Selon que les séquences sont strictement répétées en tandem ou plus ou moins dispersées dans le génome, l'évolution concertée nécessite des homogénéisations intrachromosomiques (intramoléculaires ou entre chromatides soeurs) au cours de la mitose et/ou des homogénéisations interchromosomiques (entre chromosomes homologues ou non homologues) au cours de la méiose. L'étude de l'ADNr de la levure a démontré l'existence de chacun de ces deux processus d'homogénéisation à plus ou moins longue distance (Petes 1980, Klein et Petes 1981). Il est probable que plusieurs mécanismes moléculaires sont impliqués. Il s'agirait principalement du crossing-over inégal (Smith 1976, **Figure 6**) et de la conversion génique (Dover 1982) pour lesquels il existe de nombreux et solides arguments théoriques et expérimentaux (Petes 1980, Ohta 1980, 1984, Nagylaki 1984, Stephan 1989, Hillis *et al.* 1991, Liao 1999). Cependant, tous les mécanismes capables de provoquer des expansions ou des contractions des familles de séquences répétées comme la transposition, le glissement de brins et l'amplification en cerceau ou par réinitiations multiples de la réplication, sont théoriquement des acteurs possibles de l'évolution concertée (Dover 1982, Ozenberger et Roeder 1991, Thompson-Stewart *et al.* 1994).

α- Homogénéisations par crossing-over inégal

Les duplications ou délétions de taille importante provoquées par crossing-over inégal entre deux loci très éloignés d'une même molécule, de deux chromatides soeurs ou de deux chromosomes homologues ont probablement de sérieux effets délétères. Il en va de même de la plupart des réarrangements survenant entre deux loci situés sur des chromosomes non homologues. Le crossing-over inégal est donc largement inadapté pour rendre compte de l'homogénéité des familles de séquences très dispersées. En revanche, il représente un mécanisme très efficace d'homogénéisation intralocus des séries alléliques d'une famille de séquences répétées en tandem. Des cycles successifs de crossing-over inégaux entre chromatides soeurs ou entre chromosomes homologues peuvent provoquer aussi bien l'expansion des séries alléliques que leur contraction (**Partie II-2-c-β de ce Chapitre I**) et conduire, par dérive génétique, à l'élimination d'une unité variante ou au contraire à sa propagation puis à sa fixation dans une série (**Figure 9A**). Les simulations entreprises par différents auteurs indiquent que l'homogénéisation se réalise ainsi très rapidement (Smith 1976) et même si la fréquence des crossing-over inégaux est faible par rapport au taux de mutation (Stephan 1989, Stephan et Cho 1994). Ce type de mécanisme est d'ailleurs suffisant pour expliquer la remarquable homogénéité des satellites centromériques et télomériques pourtant situés dans des régions où la recombinaison est rare (Charlesworth *et al.* 1994). Les variations importantes de taille des séries multigéniques (ADNr 5S, Cronn *et al.* 1996) ou de l'ADN non codant (minisatellites, Nakamura *et al.* 1987) généralement observées dans les populations ou entre espèces proches sont également

compatibles avec l'existence d'un tel processus d'homogénéisation. Il en va de même des changements rapides de la taille des séries de l'ADNr pendant le cycle de vie du basidiomycète *Coprinus cinereus* (Pukkila et Skrzynia 1993) ou de l'ADNr 5S chez des plants de lin soumis à un stress environnemental (Schneeberger et Cullis 1991).

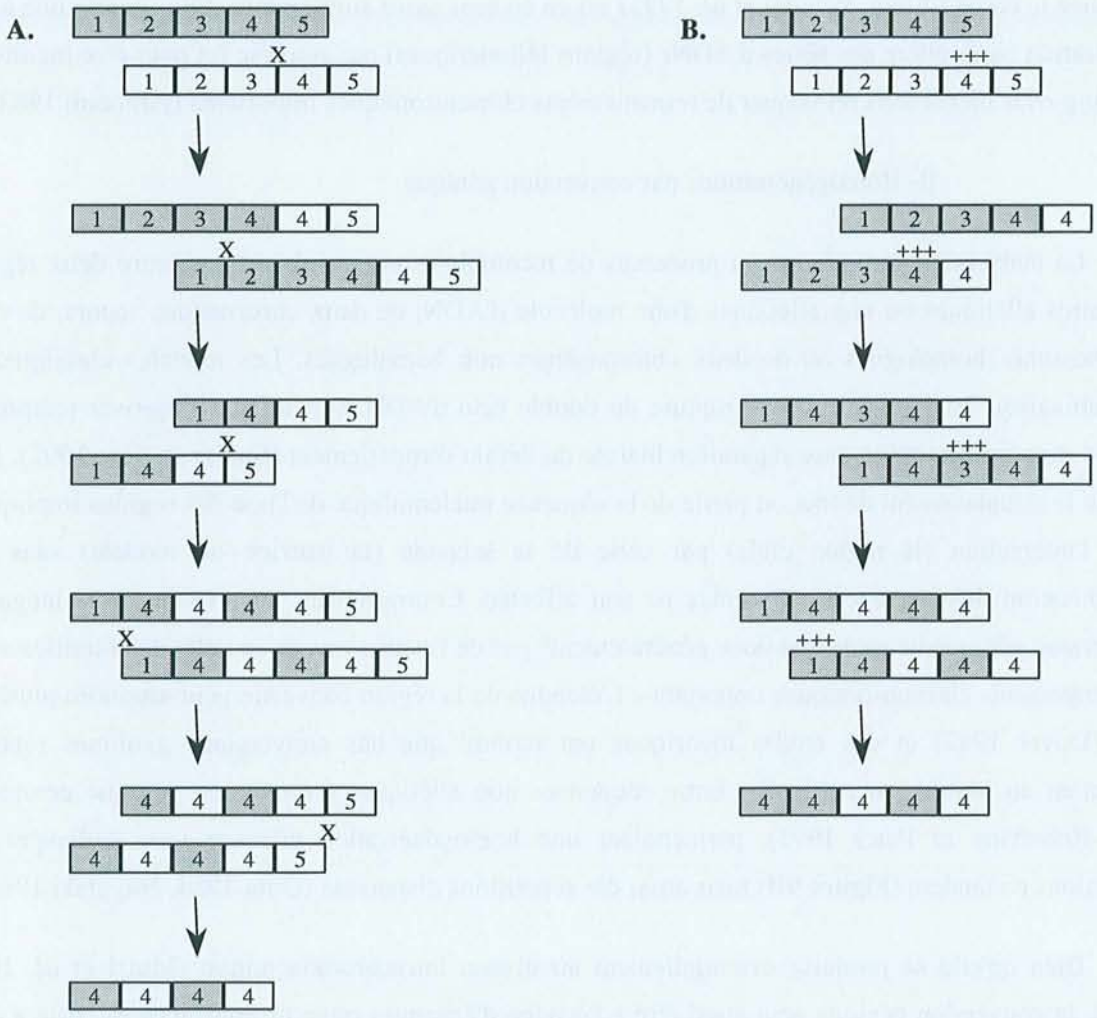



**Figure 9.** Homogénéisation hypothétique d'une série de 5 unités par cycles successifs de crossing-over inégal (**A**) ou de conversion génique ectopique (**B**) entre deux chromatides soeurs (en gris et en blanc). Les deux mécanismes peuvent aussi fonctionner entre chromosomes homologues. Modifié à partir de Li et Graur (1991).
**A.** Les crossing-over inégaux au hasard (X) provoquent des expansions et des contractions de la série initiale au gré des transmisions des molécules recombinantes. Un variant (ici l'unité N°4) peut ainsi se propager dans la série.
**B.** Les conversions géniques ectopiques au hasard (+++) entraînent progressivement l'homogénéisation de la série. Un biais directionnel peut exister en faveur de l'un des membres de la famille (par exemple ici l'unité N°4) et faciliter ainsi sa fixation (séquence fondatrice).

Les crossing-over inégaux semblent beaucoup plus fréquents entre chromatides soeurs qu'entre chromosomes homologues (démonstration au locus d'ADNr de levure, Petes 1980). L'analyse détaillée

des régions qui flanquent les séries multigéniques codant pour les ARNsn U2 chez l'homme confirme ce résultat et permet même de conclure à l'absence de crossing-over inégaux entre chromosomes homologues (Liao *et al.* 1997). L'observation d'échanges entre les loci d'ADNr situés sur des chromosomes non homologues chez l'homme et d'autres primates (5 loci, Arnheim *et al.* 1980) ainsi que chez le coton (2 loci, Wendel *et al.* 1995) est en ce sens assez surprenante. Elle semble liée à une localisation particulière des séries d'ADNr (régions télomériques) qui autorise (et peut-être facilite) le crossing-over inégal sans provoquer de remaniements chromosomiques importants (Arnheim 1983).

β- Homogénéisations par conversion génique

La conversion génique est un processus de recombinaison non réciproque entre deux régions similaires alléliques ou non alléliques d'une molécule d'ADN, de deux chromatides soeurs, de deux chromosomes homologues ou de deux chromosomes non homologues. Les modèles classiques de recombinaison impliquent soit une rupture du double brin d'ADN, soit un crossing-over réciproque simple sans isomérisation avec réparation biaisée du défaut d'appariement (Singer et Berg 1992). Il en résulte le remplacement de tout ou partie de la séquence nucléotidique de l'une des régions impliquées dans l'interaction (la région cible) par celle de la seconde (la matrice ou modèle) sans que l'organisation des séquences adjacentes ne soit affectée. Contrairement au crossing-over inégal, la conversion génique ne provoque donc généralement[6] pas de fluctuations de la taille des familles ou de réarrangements chromosomiques importants. L'étendue de la région convertie peut atteindre plusieurs kpb (Dover 1982) et des études théoriques ont montré que des conversions géniques répétées survenant au hasard, en particulier entre séquences non alléliques (conversion génique ectopique, Jinks-Robertson et Petes 1993), permettaient une homogénéisation efficace non seulement des répétitions en tandem (**Figure 9B**) mais aussi des répétitions dispersées (Ohta 1984, Nagylaki 1984).

Bien qu'elle se produise essentiellement au niveau intrachromosomique (Murti *et al.* 1992, 1994), la conversion génique peut aussi être à l'origine d'échanges interchromosomiques. Elle a ainsi été identifiée comme responsable de l'homogénéisation de gènes dispersés chez les bactéries (Abdulkarim et Hughes 1996) et la levure (Amstutz *et al.* 1985). Ohta et Dover (1983) ont en outre montré que des événements rares de conversion génique survenant au hasard entre chromosomes homologues ou non homologues suffisaient pour expliquer l'homogénéisation d'une famille de séquences dispersées sur quelques chromosomes. De même, l'évolution concertée des petites familles d'éléments mobiles très dispersés peut être simplement expliquée par les effets aléatoires conjugués des conversions géniques et des duplications par transposition. Roeder et Fink (1982) ont ainsi calculé que l'homogénéisation de la famille des rétrotransposons viraux Ty1 (**Partie II-1-b-α de ce Chapitre I**) de la levure se réalise par des conversions géniques qui surviennent à des fréquences 10 à 100 fois supérieures à celles des transpositions. Il est important de souligner par ailleurs que la conversion

---

[6] Certaines conversions géniques semblent cependant se réaliser en association avec des crossing-over chez la levure (Jinks-Robertson et Petes 1993).

génique n'est pas toujours aléatoire et qu'elle peut parfois se réaliser dans un sens préférentiel (Dover 1982). Chez la levure par exemple, certains loci sont des matrices plus efficaces que d'autres.

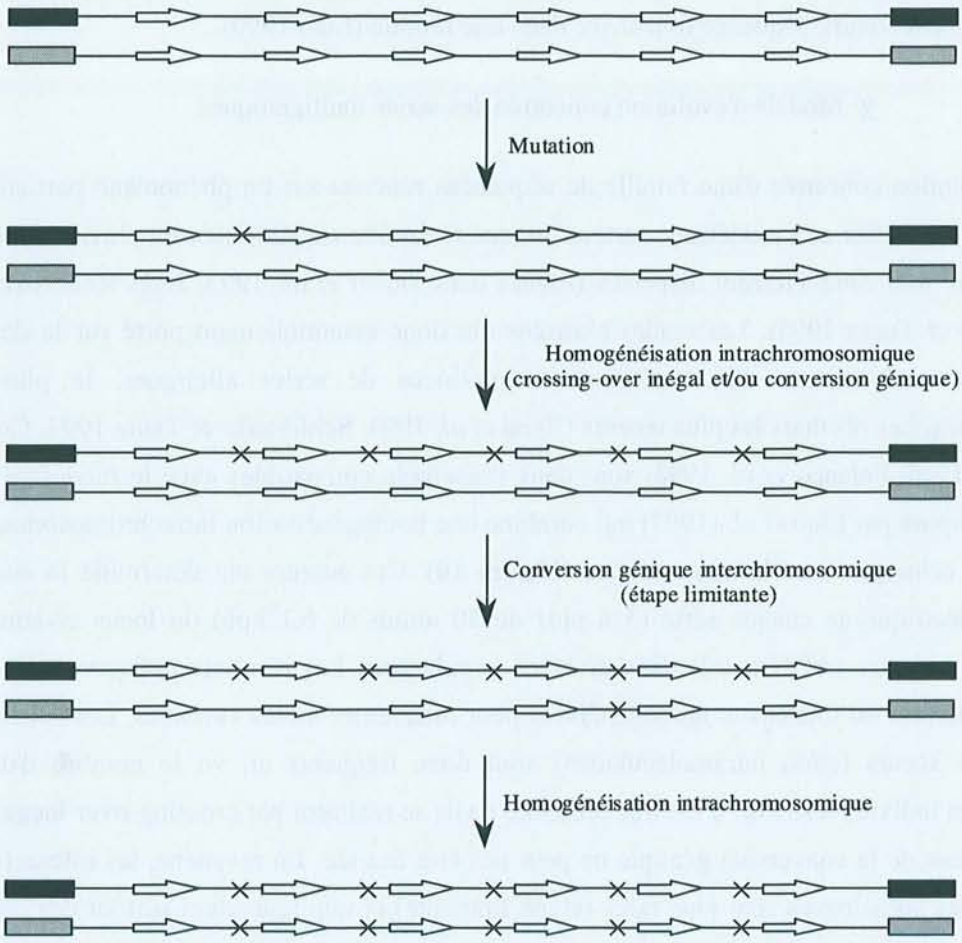



**Figure 10.** Modèle d'évolution concertée intralocus des séries multigéniques déduit de l'analyse du locus codant pour les ARNsn U2 humains. D'après Liao *et al.* (1997).
Deux séries alléliques (situées sur des chromosomes homologues) ainsi que leurs régions flanquantes (en noir et hachurées, respectivement) sont représentées. Les régions codantes de chaque série sont figurées par des flèches blanches et les espaceurs par des lignes. Une mutation (X) intervenant dans une répétition de la série allélique supérieure est rapidement fixée par des mécanismes d'homogénéisation intrachromosomique (recombinaisons intramoléculaires ou entre chromatides soeurs). Elle est ensuite transférée à la série allélique inférieure par de rares conversions géniques entre chromosomes homologues (pas d'échanges des régions flanquantes) où elle subit un nouveau cycle d'homogénéisation intrachromosomique.

Il semble de même que des conversions géniques biaisées interviennent entre les différents loci d'ADNr chez le lézard *Heteronotia binoei* (Hillis *et al.* 1991). Les études théoriques de Nagylaki et Petes (1982) et Walsh (1985) montrent qu'un tel biais directionnel, même minime, augmente considérablement la probabilité de fixation d'un membre d'une famille (séquence dite fondatrice), en

particulier dans les très grandes familles de séquences dispersées comme les SINE *Alu* humaines. Des caractéristiques structurales particulières favorisant la conversion génique, comme la présence de microsatellites (Hibner *et al.* 1991), pourraient ainsi permettre à un variant de s'établir momentanément comme séquence fondatrice dans une famille (Liao 1999).

### χ- Modèle d'évolution concertée des séries multigéniques

L'évolution concertée d'une famille de séquences répétées est un phénomène particulièrement complexe à quantifier et à modéliser, surtout lorsque la famille est sudivisée en plusieurs loci ou que ses membres sont complètement dispersés (revues dans Dover *et al.* 1993, Jinks-Robertson et Petes 1993, Elder et Tuner 1995). Les études réalisées ont donc essentiellement porté sur la détection et l'analyse des mécanismes d'homogénéisation intralocus de séries alléliques, le plus souvent multigéniques. Les résultats les plus récents (Reed *et al.* 1994, Schlötterer et Tautz 1994, Copenhaver et Pikaard 1996, Polanco *et al.* 1998) sont dans l'ensemble compatibles avec le modèle d'évolution concerté proposé par Liao *et al.* (1997) qui combine une homogénéisation intrachromosomique rapide et de rares échanges interchromosomiques (**Figure 10**). Ces auteurs ont déterminé la composition allélique spécifique de chaque série (5 à plus de 30 unités de 6,1 kpb) du locus codant pour les ARNsn U2 humains (ADNsn U2) dans diverses populations. Les résultats indiquent que les séries sont partiellement ou totalement homogénéisées pour différentes unités variantes. Les échanges entre chromatides soeurs (et/ou intramoléculaires) sont donc fréquents et, vu le nombre d'unités très variable d'un individu à l'autre, il est très probable qu'ils se réalisent par crossing-over inégal même si la contribution de la conversion génique ne peut pas être écartée. En revanche, les interactions entre chromosomes homologues sont plus rares (étape limitante) et impliqueraient surtout des conversions géniques car aucun échange entre les marqueurs polymorphes des régions qui flanquent chaque série n'a pu être observé. Il n'est toutefois pas exclu que les échanges interchromosomiques soient plus fréquents et/ou se réalisent par crossing-over inégal dans d'autres familles de taille plus importante (voir Coen et Dover 1983 et Wendel *et al.* 1995). Les résultats récents de Polanco *et al.* (1998) concernant le locus d'ADNr localisé sur la paire de chromosomes sexuels de *D. melanogaster* sont cependant troublants. L'analyse du polymorphisme des ITS montre que les unités d'ADNr subissent une forte homogénéisation intrachromosomique cohérente avec le modèle de Liao *et al.* (1997). L'analyse des IGS des mêmes unités conduit à une conclusion inverse : les unités d'ADNr évolueraient par de fréquents échanges interchromosomiques du type crossing-over inégal. Les auteurs expliquent ce paradoxe par des homogénéisations indépendantes de ces deux régions de l'unité d'ADNr. En particulier, ils invoquent un mécanisme hypothétique d'échange réciproque simple brin lié à la présence de nombreuses sous-répétitions dans l'IGS. Le modèle d'évolution concertée intralocus développé par Liao *et al.* (1997) est donc clairement insuffisant dans ce cas.

d- Conséquences de l'évolution concertée

On considère classiquement que la fixation d'une mutation dans une population d'individus sexués s'opère uniquement par sélection naturelle (évolution adaptative) et/ou dérive génétique au hasard (évolution non adaptative). Dans le cas des familles de séquences répétées, elle est également (et surtout) la conséquence des mécanismes de l'évolution concertée. En effet, la fréquence des échanges intra- et interchromosomiques (crossing-over inégal et conversion génique) est suffisamment importante par rapport au taux de mutation (de l'ordre de $10^3$ à $10^6$ fois plus élevée, Liao 1999) pour permettre la diffusion progressive d'une ou plusieurs unités variantes à l'ensemble des membres d'une famille au cours de l'évolution (homogénéisation intrachromosomique). En revanche, elle est beaucoup trop faible pour que la fixation puisse se réaliser à l'échelle de la vie d'un individu (Dover 1982). Les effets du brassage chromosomique qui intervient entre les individus à chaque génération dépassent de beaucoup ceux de l'homogénéisation intragénomique. Ceci explique que la proportion de variants soit similaire dans chaque individu d'une population pour une famille donnée (Ohta et Dover 1984). Ce processus de propagation concomitante d'un variant dans une famille (homogénéisation) et dans une population au cours des générations (fixation) a été nommé "molecular drive" par Dover (1982). Sa dynamique dépend essentiellement de la fréquence des échanges interchromosomiques (entre chromosomes homologues ou non homologues) et de l'efficacité du brassage chromosomique (fonction du temps de génération). Conjugué aux effets éventuels de la sélection naturelle et/ou de la dérive génétique au hasard, il peut conduire à une cohésion génétique élevée (homogénéité) dans de vastes unités de reproduction panmictiques (populations, espèces, complexes d'espèces).

Etant donné la variété des mécanismes d'homogénéisation susceptibles d'intervenir dans les génomes eucaryotes (**Partie II-3-c** ci-dessus), leur possible superposition, leur fréquence variable, les biais éventuels, les différentes familles (taille, structure des unités, subdivisions, etc.) et niveaux (intra- et interchromosomiques) sur lesquels ils peuvent s'exercer, il n'est pas surprenant de constater (Dover 1982, 1993, Liao 1999) que l'évolution concertée peut conduire à l'apparition rapide de polymorphismes (ou discontinuités) entre les populations (différenciation génétique) ou entre les espèces (**Figure 8A,** revue dans Elder et Turner 1995). Par exemple, les marqueurs RFLP de l'ADNr permettent généralement de distinguer des espèces, même très proches, d'étudier la structuration géographique des populations ou bien encore l'introgression entre espèces dans les zones d'hybridation (Avise 1994). Dans certains cas même, il est possible que l'évolution concertée soit un facteur puissant de la spéciation (Dover 1982). Les signaux de contrôle de la transcription des gènes des ARNr par la Pol I (promoteur situé dans l'IGS) interagissent ainsi avec des facteurs protéiques qui ont une spécificité prononcée (L'ADNr de souris ne peut pas être transcrit par une cellule humaine par exemple). Selon Heix et Grummt (1995), le transfert horizontal d'une mutation ayant affecté la région promotrice des unités d'ADNr aurait induit la coévolution adaptative des gènes des facteurs protéiques spécifiques. De même, des travaux récents (Swanson et Vacquier 1998) indiquent que

l'homogénéisation des gènes codant pour les récepteurs ovulaires d'une protéine spermatique est la base moléculaire de l'apparition des barrières reproductives dans le genre *Haliotis* (les ormeaux).

**4- Conclusion**

La fraction répétée du génome nucléaire des eucaryotes présente donc une architecture particulièrement complexe (**Figure 11**). Elle est composée de nombreuses familles de séquences fonctionnelles (codantes ou non codantes) ou de fonction inconnue qui sont isolées les unes des autres, juxtaposées, ou plus ou moins imbriquées. Leur importance est très inégale et leur structure variable. Les unités répétées peuvent ainsi être simples ou multimériques, très courtes ou au contraire très longues, dispersées ou groupées en tandem (séries parfaites ou imparfaites). Malgré cette diversité d'organisation, la plupart des familles de séquences répétées ont en commun de subir une homogénéisation intragénomique (évolution concertée) plus ou moins efficace (aléatoire ou dirigée selon les mécanismes moléculaires en cause) qui conduit à une plus grande cohésion génétique dans les populations ("molecular drive"). Paradoxalement, cette évolution concertée qui s'exerce à des niveaux de variation intraindividuel inexistants dans le cas des séquences uniques (**Figure 1**), peut aussi contribuer à la divergence rapide des populations et des espèces. Les familles de séquences répétées sont donc susceptibles de fournir de nombreux marqueurs génétiques polymorphes. Toutefois, l'interprétation de ces variations nécessite une solide évaluation préalable de l'intensité de l'évolution concertée (Ohta et Dover 1983).

Dans le cadre de la recherche de marqueurs nucléaires adaptés à l'identification des plantes (typage) et à l'étude des variations génétiques (diversité, génétique des populations, relations phylogénétiques), les familles de séquences fonctionnelles (familles multigéniques) constituent les cibles les plus intéressantes en raison de leur universalité (non garantie dans le cas des familles de l'ADN non fonctionnel). D'innombrables études ont en particulier démontré l'intérêt de la famille des gènes codant pour les ARNr 18S, 5,8S et 28S (ADNr). D'une part, la grande taille de cette famille et son organisation en série d'unités répétées en tandem facilitent son analyse. D'autre part, la structure particulière de ses longues répétitions (**Figure 3**) qui comportent à la fois des régions très conservées (les gènes) et des régions variables (ITS) ou très variables (IGS) est favorable à l'obtention de multiples marqueurs informatifs à divers niveaux taxinomiques (Hamby et Zimmer 1992, Avise 1994, Henry 1997), de l'individu (IGS) au genre (ITS et IGS), et du genre aux subdivisions supérieures du règne végétal (polymorphismes des gènes des ARNr 18S par exemple).

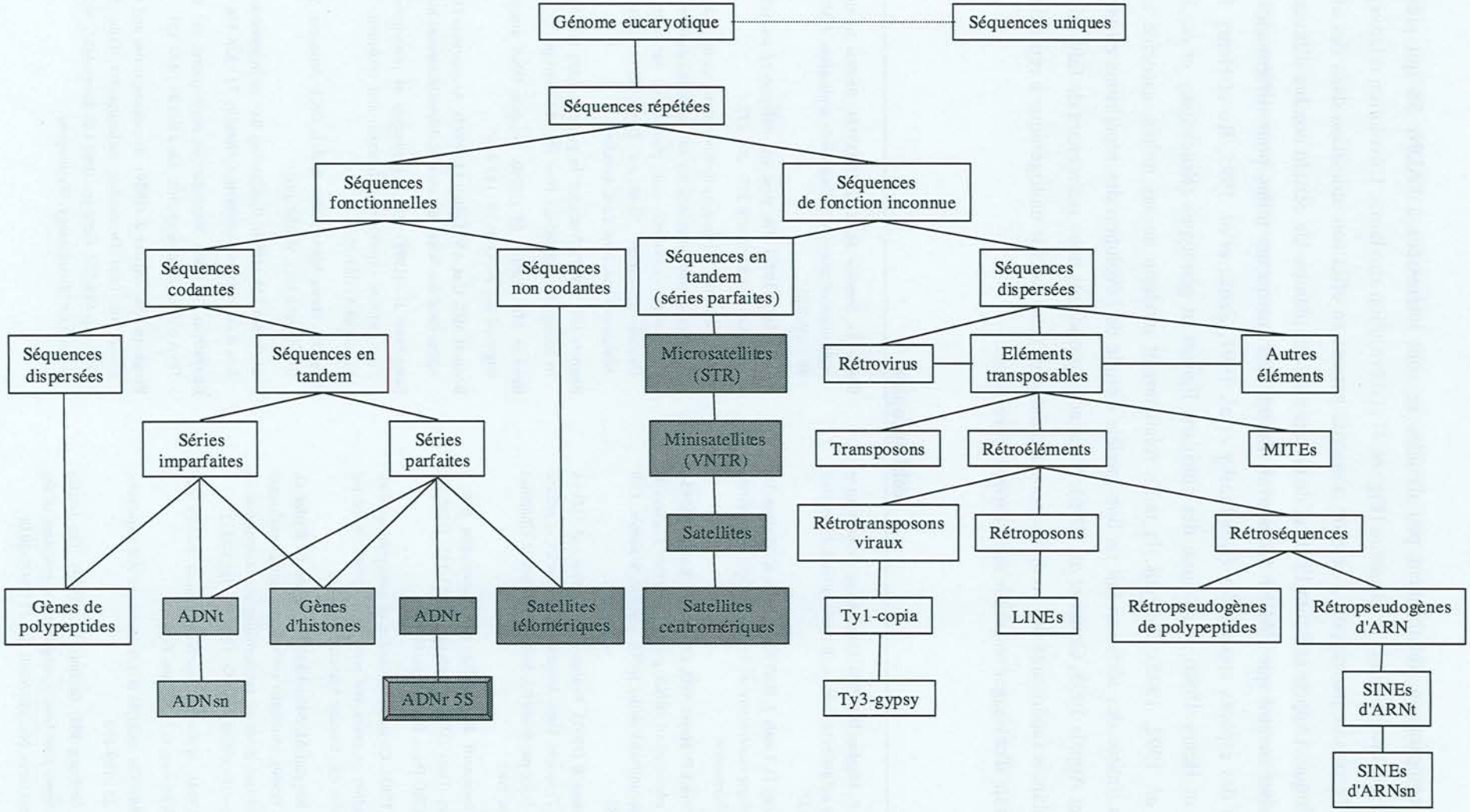



**Figure 11.** Architecture simplifiée de la fraction répétée du génome nucléaire des eucaryotes. L'échelle de gris est inversement proportionnelle à la dispersion des membres d 'une famille.

En comparaison, relativement peu d'études se sont intéressées à l'ADNr 5S qui présente pourtant des caractéristiques de structure (**Figure 4**) et d'évolution similaires. L'évolution relativement rapide de l'espaceur par rapport à la zone transcrite permet en effet son utilisation dans des études phylogénétiques (Appels et Baum 1991) et de typage chez les plantes. Un certain nombre d'études ont en particulier montré que l'ADNr 5S pouvait fournir des marqueurs utiles pour différencier non seulement des espèces apparentées (Kolchinsky *et al.* 1991, Zanke *et al.* 1995, Ko et Henry 1996, Shepherd et Henry 1998), mais aussi des cultivars, lignées et génotypes (Kolchinsky *et al.* 1991, Röder *et al.* 1992, 1995). De plus, la taille relativement modeste de ses unités, associée à une dispersion limitée des séries en fait un bon modèle d'étude de l'évolution des répétitions en tandem (Kellogg et Appels 1995, Cronn *et al.* 1996). Nous nous proposons donc maintenant de faire le bilan chez les plantes supérieures des connaissances acquises sur cette famille multigénique à répétition de dosage, avant d'envisager son étude sur 4 espèces forestières.

## Bibliographie


**Abdulkarim F, Hughes D (1996).** Homologous recombination between the *tuf* genes of *Salmonella typhimurium*. J Mol Biol 260 : 506-522

**Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD (1989).** *Biologie moléculaire de la cellule*. Deuxième édition française. Flammarion

**Amstutz H, Munz P, Heyer WD, Leupoid U, Kohli J (1985).** Concerted evolution of tRNA genes: intergenic conversion among three unlinked serine tRNA genes in *S. pombe*. Cell 40 : 879-886

**Appels R, Baum B (1991).** Evolution of the Nor and 5SDNA loci in the *Triticeae*. Dans *Molecular systematics of plants*, pp 92-116. Edité par Soltis PS, Soltis DE, Doyle JJ. Chapman & Hall, New York

**Appels R, Honeycutt RL (1986).** rDNA: evolution over a billion years. Dans *DNA systematics*, pp 81-135. Edité par Dutta SK. CRC Press, Boca raton, Florida

**Arnheim N (1983).** Concerted evolution of multigene families. Dans *Evolution of genes and proteins*, pp 38-61. Edité par Nei M, Koehn RK. Sinauer Associates, Boston

**Arnheim N, Krystal M, Schmickel R, Wilson G, Ryder O, Zimmer E (1980).** Molecular evidence for genetic exchanges among ribosomal genes on nonhomologous chromosomes in man and apes. Proc Natl Acad Sci USA 77 : 7323-7327

**Avise JC (1994).** *Molecular markers, natural history and evolution*. Chapman & Hall, New York

**Bains W (1986).** The multiple origins of human Alu sequences. J Mol Evol 23 : 189-199

**Barsch GS, Seeburg PH, Gelinas RE (1983).** The human growth hormone gene family: structure and evolution of the chromosomal locus. Nucleic Acids Res 11 : 3939-3958

**Bastia D, Sueoka M, Cox E (1975).** Studies in the late replication of phage λ: rolling circle replication. J Mol Biol 98 : 305-320

**Bennett MD (1977).** The time and duration of meiosis. Phil Trans Roy Soc London B 277 : 201-277

**Bennett MD (1982).** Nucleotypic basis of the spatial ordering of the chromosomes in eukaryotes and implications of the order for genome evolution and phenotypic variation. Dans *Genome Evolution*. Edité par Dover GA, Flavell RB. Academic Press, New York, Londre

**Bennett MD (1987).** Variation in genomic form in plants and its ecological implications. New Phytol 106 (sup) : 177-200

**Bennett MD, Leitch IJ (1995).** Nuclear DNA amounts in angiosperms. Ann Bot 76 : 113-176

**Bennett MD, Cox AV, Leitch IJ (1997).** Angiosperm DNA C-values database. www.rbgkew.org.uk/cval/database1.html l

**Bennetzen JL (1996).** The contribution of retroelements to plant genome organization, function and evolution. Trends Microbiol 4 : 347-353

**Berg DE, Howe MM (1989).** *Mobile DNA*. American Society for Microbiology, Washington

**Beukeboom LW (1994).** Bewildering Bs : an impression of the first B chromosome conference. Heredity 73 : 328-336

**Bharathan G (1996).** Reproductive development and nuclear DNA content in angiosperms. Am J Bot 83 : 440-451

**Bingham PM, Zachar Z (1989).** Retrotransposons and the FB transposon from *Drosophila melanogaster*. Dans *Mobile DNA*, pp 485-502. Edité par Berg DE, Howe MM. American Society for Microbiology, Washington

**Bingham PM, Kidwell MG, Rubin GM (1982).** The molecular basis of P-M hybrid dysgenesis : the role of the P element, a P strain-specific transposon family. Cell 29 : 995-1004

**Bostock CJ (1986).** Mechanisms of DNA sequence amplification and their evolutionary consequences. Phil Trans Roy Soc London 312 B : 261-273

**Brandes A, Heslop-Harrison JS, Kamm A, Kubis S, Doudrick RL, Schmidt T (1997).** Comparative analysis of the chromosomal and genomic organization of *Tyl-copia-like* retrotransposons in pteridophytes, gymnosperms and angiosperms. Plant Mol Biol 33 : 11-21

**Britten RJ (1994).** Evolutionary selection against change in many Alu repeat sequences interspersed throughout primate genomes. Proc Natl Acad Sci USA 91 : 5992-5996

**Britten RJ, Kohne DE (1968).** Repeated sequences in DNA. Science 161 : 529-540

**Britten RJ, Baron WF, Stout DB, Davidson EH (1988).** Sources and evolution of human Alu repeated sequences. Proc Natl Acad Sci USA 85 : 4770-4774

**Brown MS, Goldstein JL (1986).** A receptor-mediated pathway for cholesterol homeostasis. Science 232 : 34-47

**Brown DD, Wensink PC, Jordan E (1972).** A comparison of the ribosomal DNAs of *Xenopus laevis* and *Xenopus mulleri*: evolution of tandem genes. J Mol Biol 63 : 57-73

**Brutlag DL (1980).** Molecular arrangement and evolution of heterochromatic DNA. Ann Rev Genet 14 : 121-144

**Burt A, Trivers R (1998).** Selfish DNA and breeding system in flowering plants. Proc Roy Soc London 265B : 141-146

**Cappello J, Cohen SM, Lodish HF (1984).** *Dictyostelium* transposable element *DIRS-1* preferentially inserts into *DIRS-1* sequences. Mol Cell Biol 4 : 2207-2213

**Cavalier-Smith T (1985).** Eukaryotic gene numbers, noncoding DNA and genome size. Dans *The evolution of genome size*, pp 69-103. Edité par Cavalier-Smith T. Wiley, Chichester, UK

**Cech TR (1988).** G-strings at chromosome ends. Nature 332 : 777-778

**Charlesworth B, Langley C, Stephan W (1986).** The evolution of restricted recombination and the accumulation of repeated DNA sequences. Genetics 112 : 947-962

**Charlesworth B, Sniegowski P, Stephan W (1994).** The evolutionary dynamics of repetitive DNA in Eukaryotes. Nature 371 : 215-220

**Clare J, Farabaugh P (1985).** Nucleotide sequence of a yeast *Ty* element: evidence for an usual mechanism of gene expression. Proc Natl Acad Sci USA 82 : 2829-2833

**Clarkson SG (1983).** Transfer RNA genes. Dans *Eukaryotic genes: their structure, activity and regulation*, pp 239-261. Edité par Maclean N, Gregory SP, Flavell RA. Butterworth, Londre

**Coen ES, Carpenter R (1986).** Transposable elements in *Antirrhinum majus*: generators of genetic diversity. Trends Genet 2 : 292-296

**Coen ES, Dover GA (1983).** Unequal exchanges and the co-evolution of X and Y rDNA arrays in *Drosophila melanogaster*. Cell 73 : 849-855

**Coen E, Strachan T, Dover G (1982).** Dynamics of concerted evolution of ribosomal DNA and histone gene families in the *melanogaster* species subgroup of *drosophila*. J Mol Biol 158 : 17-35

**Copenhaver GP, Pikaard CS (1996).** Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in *Arabidopsis thaliana* suggesting local spreading of variants as the mode for gene homogenization during concerted evolution. Plant J 9 : 273-282

**Cortadas J, Pavon MC (1982).** The organization of ribosomal genes in vertebrates. EMBO J 1 : 1075-1080

**Cronn RC, Zhao X, Paterson AH, Wendel JF (1996).** Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. J Mol Evol 42 : 685-705

**Cullis CA (1990).** DNA rearrangements in response to environmental stress. Adv Genet 28 : 73-97

**Daniels GR, Deininger PL (1985).** Several major mammalian SINE families are derived from tRNA genes. Nature 317 : 819-822

**Dawid IB, Brown DD, Reeder RH (1970).** Composition and structure of chromosomal and amplified ribosomal DNAs of *Xenopus laevis*. J Mol Biol 51 : 341-360

**DeGroote F, Renault S, Picard G (1990).** Nucleotide sequence of circular DNA molecules homologous to the 240 bp tandem repeats of the intergenic spacer of *Drosophila melanogaster* ribosomal DNA. Biol Cell 69 : 69-70

**Deininger PL (1989).** SINEs : short interspersed repeated DNA elements in higher eukaryotes. Dans *Mobile DNA*, pp 619-636. Edité par Berg DE, Howe MM. Washington DC, American society for microbiology

**Deininger PL, Batzer MA (1993).** Evolution of retrotransposons. Dans *Evolutionary biology* Vol 27, pp 157-196. Edité par Hetch MK, MacIntyre RJ, Clegg MT. Plenum, New York

**Deumling B, Clermont L (1989).** Changes in DNA content and chromosomal size during cell culture and plant regeneration of *Scilla siberica*: selective chromatin diminution in response to environmental conditions. Chromosoma 97 : 439-448

**Dhillon SS (1987).** DNA in tree species. Dans *cell and tissue culture in forestry* vol 1, pp 298-313. Edité par Bonga JM, Durzan DJ. Martinus Nijhoff publisher

**Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1993).** Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA 91 : 3166-3170

**Dib C, Fauré S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J (1996).** A comprehensive genetic map of the human genome based on 5264 microsatellites. Nature 380 : 152-154

**Doolittle WF, Sapienza C (1980).** Selfish genes, the phenotype paradigm and genome evolution. Nature 284 : 601-603

**Dounce AL (1955).** The isolation and composition of cell nucleoli. Dans *the nucleic acid*, pp 93-153. Edité par Chargaff E, Davidson JN. Academic Press, New York

**Dover GA (1980).** Ignorant DNA ? Nature 285 : 618-619

**Dover GA (1982).** Molecular drive: a cohesive mode of species evolution. Nature 299 : 111-117

**Dover GA (1993).** Evolution of genetic redundancy for advanced players. Curr Opin Genet Dev 3 : 902-910

**Dover GA, Linares AR, Bowen T, Hancock JM (1993).** Detection and quantification of concerted evolution and molecular drive. Methods Enzymol. 224 : 525-541

**Dow BD, Ashley MV, Howe HF (1995).** Characterization of highly variable (GA/CT)n microsatellites in the bur oak, *Quercus macrocarpa*. Theor Appl Genet 91 : 137-141

**Drouin G, Moniz de Sá M (1995).** The concerted evolution of 5S ribosomal genes linked to the repeat units of other multigene families. Mol Biol Evol : 481-493

**Elder JF Jr, Turner BJ (1995).** Concerted evolution of repetitive DNA sequences in eukaryotes. Q. Rev. Biol. 70 : 297-320

**Ellis THN, Lee D, Thomas CM, Simpson PR, Cleary WG, Newman MA, Burcham KWG (1988).** 5S rRNA genes in *Pisum* : sequence, long range and chromosomal organization. Mol Gen Genet 214 : 333-342

**Endow SA, Hatwood KC (1988).** Magnification: gene amplification by an inducible system of sister chromatid exchange. Trends Genet 3 : 348-351

**Engels WR (1989).** P elements in *Drosophila melanogaster*. Dans *Mobile DNA*, pp 437-484. Edité par Berg DE, Howe MM. American Society for Microbiology, Washington

**Engels WR (1992).** The origin of P elements in *Drosophila melanogaster*. BioEssays 14 : 681-686

**Esteal S (1988).** Rate constancy of globin gene evolution in placental mammals. Proc Natl Acad Sci USA 85 : 7622-7626

**Fedoroff NV (1989).** Maize transposable elements. Dans *Mobile DNA*, pp 375-412. Edité par Berg DE, Howe MM. American Society for Microbiology, Washington.

**Finnegan DJ (1989).** The I factor and I-R dysgenesis in *Drosophila melanogaster*, pp 503-518. F and related elements in *Drosophila melanogaster*, pp 519-522. Dans *Mobile DNA*. Edité par Berg DE, Howe MM. American Society for Microbiology, Washington

**Flamm WG (1972).** Highly repetitive sequences of DNA in chromosomes. Int Rev Cytol 32 : 1-51

**Flavell RB (1986).** Repetitive DNA and chromosome evolution in plants. Phil Roy Soc London 312 B : 227-242

**Flavell RB, Jones J, Lonsdale DM, O'Dell M (1983).** Higher plant genome structure and the dynamics of genome evolution. Dans *Advance in gene technology*, pp 47-59. Edité par Downey K, Voellmy RN, Ahmed F, Schultz J. Academic Press, New York

**Flavell AJ, Smith DB, Kumar A (1992a).** Extreme heterogeneity of *Tyl-copia* group retrotransposons in plants. Mol Gen Genet 231 : 233-242

**Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A (1992b).** *Tyl-copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic Acids Res 20 : 3639-3644

**Gerbi SA (1985).** Evolution of ribosomal DNA. Dans *Molecular evolutionary genetics*, pp 419-517. Edité par MacIntyre RJ. Plenum Press, New York

**Gierl A, Cuypes H, Lütticke S, Pereira A, Schwarz-Sommer Z (1988).** Structure and function of the En/Spm transposable element system of *Zea mays*: identification of the Suppressor component of En. Dans *Proceedings of the international symposium on plant transposable elements*, pp 115-119. Edité par Nelson O. Plenum, New York

**Goldblatt P (1980).** Polyploidy in angiosperms: dicotyledons. Dans *Polyploidy: biological relevance*, pp 219-240. Edité par Lewis WH. Plenum Press, New York

**Goodman M (1999).** The genomic record of humankind's evolutionary roots. Am J Hum Genet 64 : 31-39

**Govindaraju DR, Cullis CA (1991).** Modulation of genome size in plants: the influence of breeding systems and neighbourhood size. Evol Trends in Plants 5 : 43-51

**Greilhuber J (1998).** Intraspecific variation in genome size: a critical reassessment. Ann Bot 82 (sup) : 27-35

**Griffiths AJF, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM (1996).** *An introduction to genetic analysis*. Freeman WH et Cie, New York

**Grime JP (1986).** Predictions of terrestrial vegetation responses to nuclear winter conditions. Int J Env Studies 28 : 11-19

**Grime JP (1996).** Testing predictions of the human impacts of global change on terrestrial ecosystems. Asp Appl Biol 45 : 3-13

**Grime JP, Mowforth MA (1982).** Variation in genome size. An ecological interpretation. Nature 299 : 151-153

**Hamby RK, Zimmer EA (1992).** Ribosomal RNA as a phylogenetic tool in plant systematics. Dans *Molecular systematics of plants*, pp 50-91. Edité par Soltis PS, Soltis DE, Doyle JJ. Chapman & Hall, New York

**Hansen LJ, Chalker DL, Sandmeyer SB (1988).** *Ty3*, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. Mol Cell Biol 8 : 5245-5256

**Heix J, Grummt I (1995).** Species specificity of transcription by polymerase I. Curr Opin Genet Dev 5 : 652-656

**Henry RJ (1997).** *Practical applications of plant molecular biology*. Chapman & Hall, Londre

**Hentschel CC, Birnstiel ML (1981).** The organization and expression of histone gene families. Cell 25 : 301-313

**Heslop-Harrison JS, Brandes A, Taketa S, Schmidt T, Vershinin AV, Alkhimova EG, Kamm A, Doudrick RL, Schwarzacher T, Katsiotis A, Kubis S, Kumar A, Pearce SR, Flavell AJ, Harrison GE (1997).** The chromosomal distributions of *Tyl-copia* group retrotransposable elements in higher plants and their implications for genome evolution. Genetica 100 : 197-204

**Hibner BL, Burke WD, Eickbush TH (1991).** Sequence identity in an early chorion multigene family is the result of localized gene conversion. Genetics 128 : 595-606

**Higashiyama T, Noutoshi Y, Fujie M, Yamada T (1997).** Zepp, a LINE-like retrotransposon accumulated in the *Chlorella* telomeric region. EMBO J 16 : 3715-3723

**Hillis DM, Moritz C, Porter CA, Baker RJ (1991).** Evidence for biased gene conversion in concerted evolution of ribosomal DNA. Science 251 : 308-310

**Hull R, Covey SN (1996).** Retroelements: propagation and adaptation. *Virus genes* 11 : 105-118

**Hutchison III CA, Hardies SC, Loeb DD, Shehee WR, Edgell MH (1989).** LINES and related retrotransposons: Long interspersed repeated sequences in the eukaryotic genome. Dans *Mobile DNA*, pp 593-618. Edité par Berg DE, Howe MM. American Society for Microbiology, Washington

**Isagi Y, Suhandono S (1997).** PCR primers amplifying microsatellite loci of *Quercus myrsinifolia* Blume and their conservation between oak species. Mol Ecol 6 : 897-899

**Jasienski M, Bazzaz FA (1995).** Genome size and high CO2. Nature 376 : 559-560

**Jeffreys AJ, Wilson V, Thein SL (1985).** Hypervariable minisatellite regions in human DNA. Nature 314 : 67-73

**Jinks-Robertson S, Petes TD (1993).** Experimental determination of rates of concerted evolution. Methods Enzymol 224 : 631-646

**Jones RN (1985).** Are B chromosome "selfish" ? Dans *The evolution of genome size*, pp 397-425. Edité par Cavalier-Smith T. Wiley, Chichester, UK

**Jorgensen RA, Cluster PD (1988).** Modes and tempos in the evolution of nuclear ribosomal DNA: new characters for evolutionary studies and new markers for genetic and population studies. Ann Mo Bot Gard 75 : 1238-1247

**Kawai H, Nakayama T, Inouye I, Kato A (1997).** Linkage of 5S ribosomal DNA to other rDNAs in the chromophytic *algae* and related taxa. J Phycol 33 : 505-511

**Kellogg EA, Appels R (1995).** Intraspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. Genetics 140 : 325-343

**Kidwell MG, Lisch D (1997).** Transposable elements as sources of variation in animals and plants. Proc Natl Acad Sci USA 94 : 7704-7711

**Klein HL, Petes TD (1981).** Interchromosomal gene conversion in yeast. Nature 289 : 144-148

**Ko HL, Henry RJ (1996).** Specific 5S ribosomal RNA primers for plant species identification in admixtures. Plant Mol Biol Rep 14 : 33-43

**Kolchinsky A, Kolesnikova M, Ananiev E (1991).** "Portraying" of plant genomes using polymerase chain reaction amplification of ribosomal 5S genes. Genome 34 : 1028-1031

**Koop BF, Miyamoto MM, Embury JE, Goodman M, Czelusniak J, Slightom JL (1986).** Nucleotide sequence and evolution of the orangutan epsilon globin gene region and surrounding Alu repeats. J Mol Evol 24 : 94-102

**Koshoo TN (1959).** Polyploidy in gymnosperms. Evolution 13 : 24-39

**Kubis S, Schmidt T, Heslop-Harrison JSP (1998a).** Repetitive DNA elements as a major component of plant genomes. Ann Bot 82 (sup) : 45-55

**Kubis SE, Heslop-Harrison JS, Desel C, Schmidt T (1998b).** The genomic organization of non-LTR retrotransposons (LINEs) from three *Beta* species and five other angiosperms. Plant Mol Biol 36 : 821-831

**Kuff EL, Lueders KK (1988).** The intracisternal A-particule gene family: structure and functional aspects. Adv Cancer Res 51 : 183-276

**Labhart P, Reeder RH (1986).** Characterization of three sites of RNA 3' end formation in the *Xenopus* ribosomal spacer. Cell 45 : 431-433

**Lagercrantz U, Ellegren H, Andersson L (1993).** The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res 21 : 1111-1115

**Lakich D, Kazazian HH, Antonarakis SE, Gitschier J (1993).** Inversions disrupting the factor VIII gene are a common cause of Haemophilia-A. Nature Genet 5 : 236-241

**Lapitan NLV (1991).** Organization and evolution of higher plant nuclear genomes. Genome 35 : 171-181

**Leeton PRJ, Smyth DR (1993).** An abundant LINE-like element amplified in the genome of *Lilium speciosum*. Mol Gen Genet 237 : 97-104

**Leitch IJ, Chase MW, Bennett MD (1998).** Phylogenetic analysis of DNA C-values provides evidence for a small ancestral genome size in flowering plants. Ann Bot 82 (sup) : 85-94

**Levinson G, Gutman GA (1987).** Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4 : 203-221

**Lewin B (1975).** Units of transcription and translation: sequence components of heterogeneous nuclear DNA and messenger RNA. Cell 4 : 77-93

**Lewis WH (1980).** Polyploidy in angiosperms: monocotyledons. Dans *Polyploidy: biological relevance*, pp 241-273. Edité par Lewis WH. Plenum Press, New York

**Li WH, Graur D (1991).** *Fundamentals of molecular evolution*. Sinauer, Sunderland, Massachusetts

**Liao D (1999).** Concerted evolution: molecular mechanism and biological implications. Am J Hum Genet 64 : 24-30

**Liao D, Pavelitz T, Weiner AM (1998).** Characterization of a novel class of interspersed LTR elements in primate genomes: structure, genomic distribution, and evolution. J Mol Evol 46 : 649-660

**Liao D, Pavelitz T, Kidd JR, Kidd KK, Weiner AM (1997).** Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the *RNU2* locus) involves rapid intrachromosomal homogenization and rare interchromosomal gene conversion. EMBO J 16 : 588-598

**Lindauer A, Fraser D, Brüderlein M, Schmitt R (1993).** Reverse transcriptase families and a *copia*-like retrotransposon, *Osser*, in the green alga *Volvox carteri*. FEBS letters 319 : 261-266

**Lindgren V, Bernstein LB, Weiner AM, Francke U (1985).** Human U1 small nuclear RNA pseudogenes do not map to the site of the U1 genes in 1p36 but are clustered in 1q12-q22. Mol Cell Biol 5 : 2172-2180

**Long EO, Dawid IB (1980).** Repeated genes in Eukaryotes. Ann Rev Biochem 49 : 727-764

**MacGillivray CW, Grime JP (1995).** Genome size predicts frost resistance in British herbaceous plants : implications rates of vegetation response to global warming. Funct Ecol 9 : 320-325

**Mackay TFC (1986).** Transposable element-induced fitness mutations in *Drosophila melanogaster*. Genet Res 48 : 77-87

**Manninen I, Schulmann AH (1993).** *BARE-1*, a *copia*-like retroelement in barley (*Hordeum vulgare* L.). Plant Mol Biol 22 : 829-846

**Manning JE, Schmid CW, Davidson N (1975).** Interspersion of repetitive and nonrepetitive DNA sequences in the *Drosophila melanogaster* genome. Cell 4 : 141-155

**Marlor RL, Parkhurst SM, Corces VG (1986).** The *Drosophila melanogaster gypsy* transposable element encodes putative gene products homologous to retroviral protein. Mol Cell Biol 6 : 1129-1134

**McClintock B (1951).** Chromosome organisation and genic expression. Cold Spring Harbor Symposium on Quantitative Biology 16 : 13-47

**Mentzenberg RL, Stevens JN, Selker EU, Morzycka-Wroblewska E (1985).** Identification and chromosomal distribution of 5S rRNA genes in *Neurospora crassa*. Proc Natl Acad Sci USA 82 : 2067-2071

**Meyne J, Ratliff RL, Moyzis RK (1989).** Conservation of the human telomere sequence (TTAGGG)n among vertebrates. Proc Natl Acad Sci USA 86 : 7049-7053

**Moore G, Lucas H, Batty N, Flavell R (1991).** A family of retrotransposons and associated genomic variation in wheat. Genomics 10 : 461-468

**Morgante M, Olivieri AM (1993).** PCR-amplified microsatellites as markers in plant genetics. Plant J 3 : 175-182

**Mount SM, Rubin GM (1985).** Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between *copia* and retroviral proteins. Mol Cell Biol 5 : 1630-1638

**Murray BG (1998).** Nuclear DNA amounts in gymnosperms. Ann Bot 82 (sup) : 3-15

**Murti JR, Bumbulis M, Schimenti JC (1992).** High-frequency germ line gene conversion in transgenic mice. Mol Cell Biol 12 : 2545-2552

**Murti JR, Bumbulis M, Schimenti JC (1994).** Gene conversion between unliked sequences in the germline of mice. Genetics 137 : 837-843

**Muskavitch MAT, Hogness DS (1982).** An expandable gene that encodes a *Drosophila* glue protein is not expressed in variants lacking remote upstream sequences. Cell 29 : 1041-1051

**Nagylaki T (1984).** Evolution of multigene families under interchromosomal gene conversion. Proc Natl Acad Sci USA 81 : 3796-3800

**Nagylaki T, Petes TD (1982).** Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. Genetics 100 : 315-337

**Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M. Martin C, Fujimoto E, Hoff M, Kulmin E, White R (1987).** Variable number of tandem repeat (VNTR) markers for human genetic mapping. Science 235 : 1616-1622

**Narayan RKJ (1998).** The role of genomic constraints upon evolutionary changes in genome size and chromosome organization. Ann Bot 82 (sup) : 57-66

**Nei M (1987).** *Molecular evolutionary biology*. Colombia University Press, New York

**Nei M, Gu X, Sitnikova T (1997).** Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci USA 94 : 7799-7806

**Nenoi M, Mita K, Ichimura S, Kawano A (1998).** Higher frequency of concerted evolutionary events in rodents than in man at the polyubiquitin gene VNTR locus. Genetics 148 : 867-876

**Nevers P, Saedler H (1977).** Transposable genetic elements as agents of gene instability and chromosomal rearrangements. Nature : 109-115

**Noma K, Ohtsubo H, Ohtsubo E (1998).** Non-LTR retrotransposons (LINEs) as ubiquitous components among plant kingdom. Plant and animal genome VI, San Diego, Abstract guide 103

**Ohno S (1970).** *Evolution by gene duplication.* Springer-Verlag, Berlin

**Ohno S (1972).** So much "junk" DNA in our genome. Brookhaven Symp Biol 23 : 366-370

**Ohno S (1981).** Original domain for the serum-albumin family arose from repeated sequences. Proc Natl Acad Sci USA 78 : 7657-7661

**Ohta T (1980).** *Evolution and variation of multigene families.* Springer-Verlag, Berlin

**Ohta T (1984).** Some models of gene conversion for treating the evolution of gene families. Genetics 106 : 517-528

**Ohta T, Dover GA (1983).** Population genetics of multigene families that are dispersed into two or more chromosomes. Proc Natl Acad Sci USA 80 : 4079-4083

**Ohta T, Dover GA (1984).** The cohesive population genetics of molecular drive. Genetics 108 : 501-521

**Ohta T, Kimura M (1973).** A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet Res 22 : 201-204

**Okada N (1991).** SINEs. Curr Opin Genet Dev 1 : 498-504

**Okumura K, Kiyama R, Oishi M (1987).** Sequence analyses of extrachromosomal *Sau*3A and related family DNA: analysis of recombination in the excision event. Nucleic Acids Res 15 : 7477-7489

**Orgel LE, Crick FHC (1980).** Selfish DNA: the ultimate parasite. Nature 284 : 604-607

**Orgel LE, Crick FHC, Sapienza C (1980).** Selfish DNA. Nature 288 : 645-646

**Ozenberger BA, Roeder GS (1991).** A unique pathway of double strand break repair operates in tandemly repeated genes. Mol Cell Biol 11 : 1222-1231

**Pages MJM, Roizes GP (1984).** Nature and organization of the sequence variations in the long-range periodicity calf satellite DNA I. J Mol Biol 173 : 143-157

**Pâques F, Samson ML, Jordan P, Wegnez M (1995).** Structural evolution of the *Drosophila* 5S ribosomal genes. J Mol Evol 41 : 615-621

**Parham P, Ohta T (1996).** Population biology of antigen presentation by MHC class I molecules. Science 272 : 67-74

**Pearce SR, Harrison G, Li D, Heslop-Harrison JS, Flavell A, Kumar A (1996).** The *Tyl-copia* group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localization. Mol Gen Genet 250 : 305-315

**Pech M, Igo-Kemenes T, Zachau HG (1979).** Nucleotide sequence of a highly repetitive component of rat DNA. Nucleic Acids Res 7 : 417-432

**Petes TD (1980).** Molecular genetics of yeast. Ann Rev Biochem 49 : 845-876

**Pimpinelli S, Berloco M, Fanti L, Dimitri P, Bonaccorsi S, Marchetti E, Caizzi R, Caggesse C, Gatti M (1995).** Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. Proc Natl Acad Sci USA 92 : 3804-3808

**Polanco C, González AI, de la Fuente Á, Dover GA (1998).** Multigene family of ribosomal DNA in *Drosophila melanogaster* reveals contrasting patterns of homogenization for IGS and ITS spacer regions: a possible mechanism to resolve this paradox. Genetics 149 : 243-256

**Pont G, DeGroote F, Picard G (1987).** Some extrachromosomal circular DNAs from *Drosophila* embryos are homologous to tandemly repeated genes. J Mol Biol 195 : 447-451

**Price HJ (1976).** Evolution of DNA content in higher plants. Bot Rev 42 : 27-52

**Price HJ, Johnston JS (1996).** Influence of light on DNA content of *helianthus anuus*. Proc Natl Acad Sci USA 93 : 11264-11267

**Pukkila PJ, Skrzynia C (1993).** Frequent changes in the number of reiterated ribosomal RNA genes throughout the life cycle of the basidiomycete *Coprinus cinereus*. Genetics 133 : 203-211

**Reed KM, Beukeboom LW, Eickbush DG, Werren JH (1994).** Junctions between repetitive DNAs on the PSR chromosome of *Nasonia vitripennis*: association of palindromes with recombination. Mol Evol 38 : 352-362

**Rees H, Jones RN (1972).** The origin of the wide species variation in nuclear DNA content. Internat Rev Cytol 32 : 53-92

**Renault S, DeGroote F, Picard G (1993).** Identification of short tandemly repeated sequences in extrachromosomal circular DNAs from *Drosophila melanogaster* embryos. Genome 36 : 244-254

**Richards EJ, Ausubel F (1988).** Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. Cell 53 : 127-136

**Röder MS, Sorrells ME, Tanksley SD (1992).** 5S ribosomal gene clusters in wheat: pulsed field gel electrophoresis reveals a high degree of polymorphism. Mol Gen Genet 232 : 215-220

**Röder MS, Sorrells ME, Tanksley SD (1995).** Pulsed-field gel analysis of 5S and satellite DNA in Barley. Genome 38 : 153-157

**Roeder GS, Fink GR (1982).** Movement of yeast transposable elements by gene conversion. Proc Natl Acad Sci USA 79 : 5621-5625

**Rubin G (1983).** Dispersed repetitive DNAs in *Drosophila*. Dans *Mobile genetic elements*, pp 329-361. Edité par Shapiro JA. Academic Press, New York

**Ruiz-Carillo A, Affolter M, Renaud J (1983).** Genomic organization of the gene coding for the six main histones of the chicken: complete sequence of the H5 gene. J Mol Biol 170 : 843-859

**Samson ML, Wegnez M (1988).** Bipartite structure of the 5S ribosomal gene family in a *drosophila melanogaster* strain and its evolutionary implications. Genetics 118 : 685-691

**SanMiguel P, Bennetzen JL (1998).** Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann Bot 82 (Sup) : 37-44

**SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996).** Nested retrotransposons in the intergenic regions of the maize genome. Science 274 : 765-768

**Schlötterer C, Tautz D (1992).** Slippage synthesis of simple sequence DNA. Nucleic Acids Res 20 : 211-215

**Schlötterer C, Tautz D (1994).** Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. Curr Biol 4 : 777-783

**Schmid CW, Deininger PL (1975).** Sequence organization of the human genome. Cell 6 : 345-358

**Schmidt T, Kubis S, Heslop-Harrison JS (1995).** Analysis and chromosomal localization of retrotransposons in sugar beet (*Beta vulgaris* L.): LINEs and *Tyl-copia*-like elements as major components of the genome. Chromosome Res 3 : 335-345

**Schneeberger RG, Cullis CA (1991.** Specific DNA alterations associated with the environmental induction of heritable changes in flax. Genetics 128 : 619-630.

**Schwarzatcher T, Heslop-Harrison JS (1991).** *In situ* hybridization to plant telomeres using synthetic oligomers. Genome 34 : 317-333

**Schwarz-Sommer Z, Leclercq L, Goebel E, Saedler H (1987).** Cin4, an insert altering the structure of the *A1* gene in *Zea mays*, exhibits properties of nonviral retrotransposons. EMBO J 13 : 3873-3880

**Selker EU, Yanofsky C, Driftmier K, Metzenberg RL, AlznerDeWeerd B, Rajbhandary UL (1981).** Dispersed 5S RNA genes in *N. crassa* : structure, expression and evolution. Cell 24 : 819-828

**Shampay J, Blackburn EH (1988).** Generation of telomere-length heterogeneity in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 85 : 534-538

**Shepherd M, Henry R (1998).** Monitoring of fluorescence during DNA melting as a method for discrimination and detection of PCR products in variety identification. Mol Breeding 4 : 509-517

**Singer M, Berg P (1992).** *Genes & Genomes*. Edition Vigot, Paris.

**Smit AFA (1996).** The origin of interspersed repeats in the human genome. Curr Opin Genet Dev 6 : 743-748

**Smith GP (1976).** Evolution of repeated DNA sequences by unequal crossover. Science 191 : 528-535

**Sorensen PD, Frederiksen S (1991).** Characterization of human 5S RNA genes. Nucleic Acids Res 19 : 4147- 4151

**Sparrow AH, Miksche JP (1961).** Correlation of the nuclear volume and DNA content with higher plant tolerance to chronic radiation. Science 134 : 282-283

**Sparrow AH, Nauman AF (1976).** Evolution of the genome size by DNA doublings. Science 192 : 524-529

**Spradling A, Orr-Weaver T (1987).** Regulation of DNA replication during *Drosophila* development. Ann Rev Genet 21 : 373-403

**Srivastava AK, Schlessinger D (1991).** Structure and organization of ribosomal DNA. Biochimie 73 : 631-638

**Stark G, Wahl GM (1984).** Gene amplification. Ann Rev Bioch 53 : 447-491

**Steinkellner H, Fluch S, Turetschek E, Lexer C, Streiff R, Kremer A, Burg K, Glössl J (1997).** Identification and characterization of (GA/CT)n microsatellite loci from *Quercus petraea*. Plant Mol Biol 33 : 1093-1096

**Stephan W (1989).** Tandem-repetitive noncoding DNA: forms and forces. Mol Biol Evol 6 : 198-212

**Stephan W, Cho S (1994).** Possible role of natural selection in the formation of tandem noncoding DNA. Genetics 136 : 333-341

**Suoniemi A, Tanskanen J, Schulman AH (1998).** *Gypsy*-like retrotransposons are widespread in the plant kingdom. Plant J 13 : 699-705

**Suzuki H, Moriwaki K, Sakurai S (1994).** Sequences and evolutionary analysis of mouse 5S rDNAs. Mol Biol Evol 11 : 704-710

**Swanson WJ, Vacquier VD (1998).** Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. Science 281 : 710-712

**Swift H (1950).** The constancy of desoxyribose nucleic acid in plant nuclei. Genetics 36 : 643-654

**Swoboda P, Gal S, Hohn B, Puchta H (1994).** Intrachromosomal homologous recombination in whole plants. EMBO J 13 : 484-489

**Syvanen M (1984).** The evolutionary implications of mobile genetic elements. Ann Rev Genet 18 : 271-293

**Tabata S (1981).** Nucleotide sequences of the 5S ribosomal genes and their adjacent regions in *Schizosaccharomyces pombe*. Nucleic Acids Res 9 : 6429-6437

**Tautz D (1989).** Hypervariability of simples sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17 : 6463-6471

**Tautz D (1993).** Notes on the definition and nomenclature of tandemly repetitive sequences. Dans *DNA fingerprinting:*

*state of science*. Edité par Pena S, Chakraborty R, Epplen JT, Jeffreys AJ. Birkhaüser Verlag.

**Tautz D, Schlötterer C (1994)**. Simple sequences. Curr Opin Genet Dev 4 : 832-837

**Thomas CA (1971)**. The genetic organization of chromosomes. Ann Rev Genet 5 : 237-256

**Thompson-Stewart D, Karpen GH, Spradling AC (1994)**. A transposable element can drive the concerted evolution of tandemly repetitious DNA. Proc Natl Acad Sci USA 91 : 9042-9046

**Tripputi P, Emanuel BS, Croce CM, Green LG, Stern GS, Stern JL (1986)**. Human histone genes map to multiple chromosomes. Proc Natl Acad Sci USA 83 : 3185-3188

**Tschudi C, Pirotta V (1980)**. Sequence and heterogeneity in the 5S RNA gene cluster of *Drosophila melanogaster*. Nucleic Acids Res 8 : 441-451

**Turner PC, Woodland HR (1983)**. Histone gene number and organization in *Xenopus* : *Xenopus borealis* has a homogeneous major cluster. Nucleic Acids Res 11 : 971-986

**Ullu E, Tschudi C (1984)**. Alu sequences are processed 7SL RNA genes. Nature 312 : 171-172

**Van't Hof J, Sparrow AH (1963)**. A relationship between DNA content, nuclear volume, and minimum mitotic cycle time. Proc Natl Acad Sci USA 49 : 897-902

**Volkaert H, Greenwood M, Hutchison K (1992).** Analysis of microsatellite DNA sequences in *Larix spp.* Abstract. University of Maine, Orono, Maine USA.

**Voytas DF, Cummings MP, Konieczny A, Ausubel FM, Rodermel SR (1992)**. *Copia*-like retrotransposons are ubiquitous among plants. Proc Natl Acad Sci USA 89 : 7124-7128

**Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW, Collins FS (1991)**. A *de novo* Alu insertion results in neurofibromatis type-1. Nature 353 : 864-866

**Walsh JB (1985)**. Interaction of selection and biased gene conversion in a multigene family. Proc Natl Acad Sci USA 82 : 153-157

**Wang AC, Fudenberg HH (1974)**. Gene expansion and antibody variability in nonsense DNA in 1 species, sense DNA in another species. Am J Hum Genet 26 : 90a

**Wang Z, Weber JL, Zhong G, Tanksley SD (1994)**. Survey of plant short tandem DNA repeats. Theor Appl Genet 88 : 1-6

**Warburton PE, Waye JS, Willard HF (1993)**. Nonrandom localization of recombination events in human alpha satellite repeat unit variants: implications for higher-order structural characteristics within centromeric heterochromatin. Mol Cell Biol 13 : 6520-6529

**Weber JL (1990)**. Informativeness of human (dC-dA)n, (dG-dT)n polymorphisms. Genomics 7 : 524-530

**Wendel JF, Schnabel A, Seelanan T (1995)**. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc Natl Acad Sci USA 92 : 280-284

**Wessler SR, Bureau TE, White SE (1995)**. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr Opin Genet Dev 5 : 814-821

**White SE, Habera LF, Wessler SR (1994)**. Retrotransposons in the flanking regions of normal plant genes: A role for *copia-like* elements in the evolution of gene structure and expression. Proc Natl Acad Sci USA 91 : 11792-11796

**Widegren B, Arnason U, Akusjarvi G (1985)**. Characteristics of conserved 1,579-bp highly repetitive component in the killer whale, *Orcinus orca*. Mol Biol Evol 2 : 411-419

**Wright DA, Ke N, Smalle J, Hauge BM, Goodman HM, Voytas DF (1996)**. Multiple non-LTR retrotransposons in the genome of *Arabidopsis thaliana*. Genetics 142 : 569-578

**Yakura K, Kato A, Tanifuji S (1984)**. Length heterogeneity of the large spacer of *Vicia faba* is due to the differing number of a 325 bp repetitive sequence element. Mol Gen Genet 193 : 400-405

**Yao MC, Zheng K, Yao CH (1987)**. A conserved nucleotide sequence at the sites of developmentally regulated chromosomal breakage in *Tetrahymena*. Cell 48 : 779-788

**Yun TJ, Melvold RW, Pease LR (1997)**. A complex major histocompatibility complex D locus variant generated by an usual recombination mechanism in mice. Proc Natl Acad Sci USA 94: 1384-1389

**Zakian VA (1989)**. Sructure and function of telomeres. Ann Rev Genet 23 : 579-604

**Zanke C, Borisjuk N, Ruoss B, Schilde-Rentschler L, Ninnemann H, Hemleben V (1995)**. A specific oligonucleotide of the 5S rDNA spacer and species-specific elements identify symmetric somatic hybrids between *Solanum tuberosum* and *S. pinnatisectum*. Theor Appl Genet 90 : 720-726

**Zimmer EA, Martin SL, Beverley SM, Kan YW, Wilson AC (1980)**. Rapid duplication and loss of genes coding for the α chains of hemoglobin. Proc Natl Acad Sci USA 77 : 2158-2162

# CHAPITRE II : L'ADNr 5S nucléaire chez les plantes supérieures

Comme chez la plupart des eucaryotes supérieurs (**Partie II-2-b-ϕ** du **Chapitre I**), l'ADNr 5S de l'ensemble des plantes supérieures étudiées jusqu'à présent est organisé en séries parfaites qui ne sont pas imbriquées avec les séries de l'ADNr. Les deux familles peuvent être juxtaposées sur un même bras chromosomique (*Gossypium hirsutum*, Hanson *et al.* 1996), disposées chacune sur un bras (*Larix decidua*, Lubaretz *et al.* 1996) ou sur des chromosomes différents (*Thapsia garganica*, Rasmussen et Avato 1998). Les unités répétées en tandem ont une structure classique consistant en un unique gène 5S d'environ 120 pb associé avec un espaceur intergénique de taille variable (**Figure 4**). Dans ce **Chapitre II** nous examinerons plus particulièrement la fonction, l'organisation macromoléculaire et l'évolution de cette famille de gènes.

## I- Fonction et expression de l'ADNr 5S

Le petit ARNr 5S est un composant essentiel de la grande sous-unité 60S des ribosomes eucaryotiques où il forme un complexe ribonucléoprotéique (RNP) avec les ARNr 5,8S, 25-28S et 45 à 50 protéines différentes (**Figure 12**). Il se lie en particulier avec la protéine ribosomique L5 pour former la particule RNP 5S reconnue par le facteur d'élongation EF2 (Nygard et Nilsson 1987), mais son rôle précis n'est pas encore connu. En dehors des ribosomes, l'ARNr 5S a de nombreuses autres fonctions (revue dans Barciszewska *et al.* 1994). C'est un élément constitutif de diverses particules RNP (7S, 42S). Il participe aussi au complexe multienzymatique des aminoacyl-ARNt synthétases. Enfin, il stimule les réactions d'aminoacylation des ARNt. Comme les gènes des ARNt et de certains ARNsn (U6, 4,5S1) et ARNsc (7SL, 7SK, 4,5S), les gènes 5S sont transcrits par la Pol III (~ 650 kDa) localisée dans le nucléoplasme non nucléolaire (Pieler *et al.* 1987). En général, les transcrits primaires ont la même taille que le gène (peu différente de 120 nt) et correspondent à l'ARNr 5S mature. Il n'y a donc pas de maturation post-transcriptionnelle (Singer et Berg 1992, Szymanski *et al.* 1998). Les familles de l'ADNr 5S somatique et des ovocytes de *Xenopus* (**Partie II-2-b-ϕ** du **Chapitre I**) ont fourni un excellent modèle d'étude de la régulation de la transcription des gènes 5S par la Pol III (Korn 1982, Pieler *et al.* 1985, 1987, Wolffe et Brown 1988, Hayes et Tullius 1992, Sturges *et al.* 1995). Des séquences de contrôle ont été détectées au sein des gènes et dans les espaceurs. Par ailleurs, la Pol III ne se fixe pas directement sur l'ADN. La présence de plusieurs protéines additionnelles (des facteurs de transcription) est en effet nécessaire pour réaliser une initiation correcte et efficace. Etant donné que la séquence et la structure des ARNr 5S sont remarquablement conservées chez les eucaryotes (Szymanski *et al.* 1998), on peut considérer que les fonctions, signaux et processus de régulation identifiés chez *Xenopus* ont probablement une portée assez générale. Ils ont d'ailleurs été en grande partie retrouvés chez plusieurs autres organismes modèles comme *Neurospora* (Tyler 1987), *Drosophila* (Sharp et Garcia 1988), *Bombyx* (Smith *et al.* 1995), *Saccharomyces* (Lee *et al.* 1995), l'homme (Nielsen *et al.* 1993) et *Caenorhabditis* (Nelson *et al.* 1998). Concernant les

plantes en revanche, les informations sont très rares. Une seule étude, partielle, concernant le riz (*Oryza sativa*) est actuellement disponible (Hariharan *et al.* 1987).

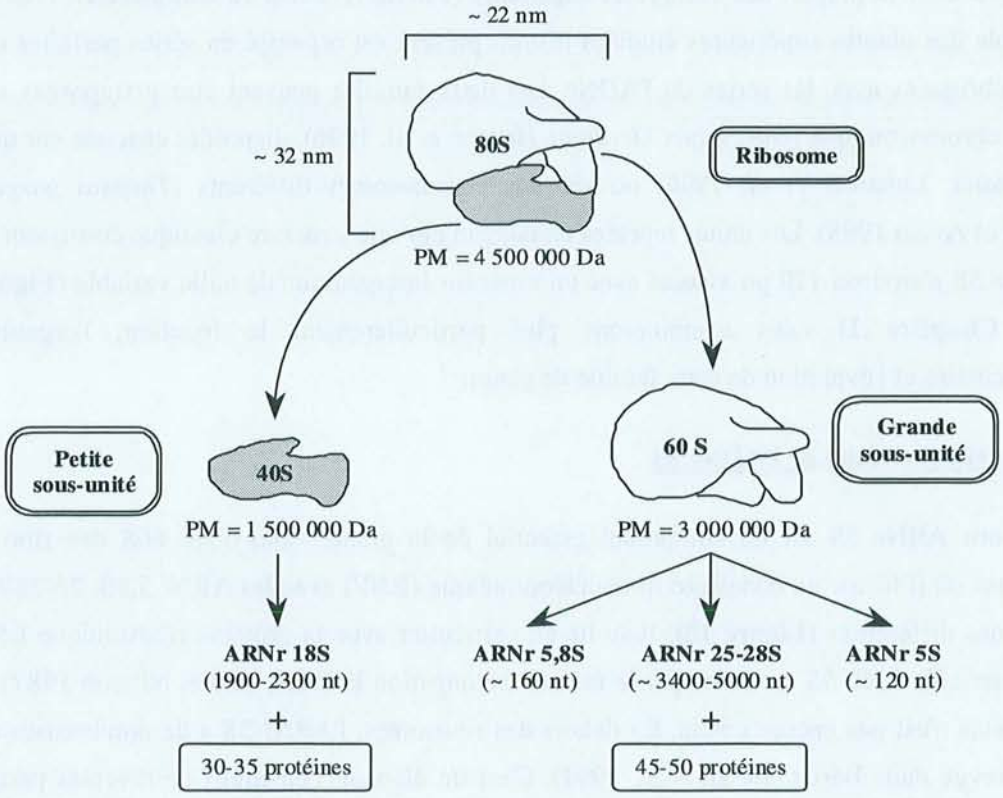



**Figure 12.** Composition typique des ribosomes eucaryotiques. D'après Singer et Berg (1992).

### 1- Signaux de contrôle de la transcription par la Pol III

Contrairement aux gènes transcrits par la Pol I (l'ADNr), mais comme les gènes transcrits par la Pol II (essentiellement les gènes de protéines), les signaux déterminant la régulation de l'initiation et de la terminaison de la transcription des divers types de gènes de la classe III sont très conservés ou analogues entre espèces, même très éloignées (Singer et Berg 1992). Les ARNr 5S, mais aussi les ARNt et d'autres petits ARN cellulaires, sont donc correctement transcrits par des systèmes hétérologues.

a- Séquences d'initiation de la transcription

Comme la plupart des gènes de la classe III, l'initiation de la transcription des gènes 5S est sous le contrôle d'un puissant promoteur intragénique (ICR : Internal Control Region). Chez le xénope, l'activité de cette région de contrôle interne est suffisante pour obtenir un bon niveau de transcription (Korn 1982) mais elle est renforcée par des éléments de régulation supplémentaires situés dans

l'espaceur immédiatement en amont du site d'initiation de la transcription (Reynolds et Azer 1988). Chez la plupart des autres organismes étudiés, ces régions promotrices de l'espaceur se sont avérées tout aussi indispensables que l'ICR (voir Nelson *et al.* 1998).

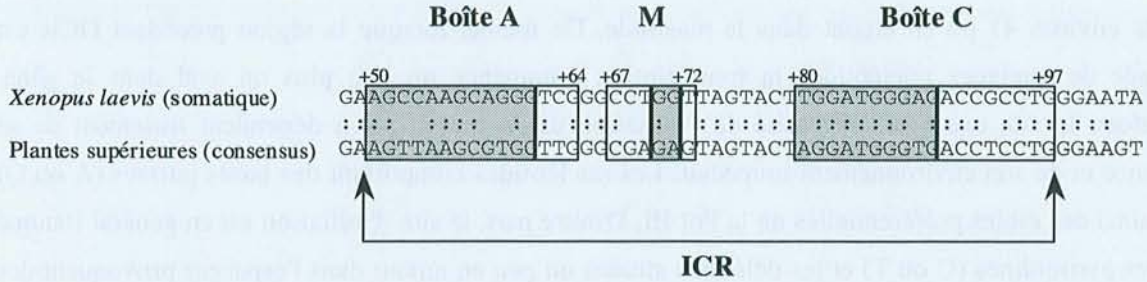



**Figure 13.** Similarité de séquence (:) entre la région de contrôle interne (ICR) de l'initiation de la transcription des gènes 5S de *Xenopus laevis* (ADNr 5S somatique) et la région correspondante des plantes supérieures (séquence consensus, Szymanski *et al.* 1998). Les séquences fonctionnelles déterminées par mutagénèse dirigée (boîte A, élément intermédiaire M et boîte C) sont encadrées et leur position indiquée par rapport à l'extrémité 5' du gène 5S (120 pb). Les nucléotides grisés ont une très forte influence sur l'activité de transcription par la Pol III (D'après Pieler *et al.* 1985, 1987).

α- La région de contrôle interne (ICR)

Chez *Xenopus laevis*, des expériences de mutagénèse *in vitro* introduisant des délétions, des insertions ou des changements de séquence dans les gènes 5S ont permis de déterminer précisément l'étendue de la région de contrôle interne (Pieler *et al.* 1985, 1987, **Figure 13**). Elle couvre approximativement les nucléotides +50 à +97 (48 pb) et comporte 3 séquences fonctionnelles distinctes de 15 pb (boîte A, nt +50 à +64), 6 pb (élément M, nt +67 à +72) et 18 pb (boîte C, nt +80 à +97). La boîte A est un élément commun des ICR de tous les gènes de la classe III, tandis que la boîte C et l'élément M sont caractéristiques des ICR de gènes 5S. Les mutations ponctuelles affectant ces trois régions ainsi que les variations de l'espacement entre les boîtes A ou C ont un impact considérable sur l'activité de transcription. En revanche, les changements de séquence intervenant dans les régions intercalées entre la boîte A et l'élément M (nt +65 et +66) et entre l'élément M et la boîte C (nt +73 à +79) n'ont qu'un effet limité. Des analyses du même type ont montré que la structure des ICR est tout à fait comparable chez *Bombyx mori* (Smith *et al.* 1995) et *Saccharomyces cerevisiae* (Lee *et al.* 1995). Des éléments régulateurs supplémentaires existent toutefois chez *Neurospora crassa* (nt +19 à +30, Tyler 1987) et *Drosophila melanogaster* (nt +3 à +18 et nt +37 à +44, Sharp et Garcia 1988). Chez les plantes supérieures, on trouve aussi une région promotrice intragénique comme le démontre, par exemple, la possibilité de transcrire *in vitro* un gène 5S de riz dissocié de son espaceur et tronqué de 30 pb à son extrémité 5' (Hariharan *et al.* 1987). La similarité de séquence

entre l'ICR de *Xenopus* et la région correspondante des plantes est par ailleurs excellente (environ 73 %, **Figure 13**).

La Pol III initie la transcription environ 50 pb en amont de l'ICR (Korn 1982). Par exemple, la transcription *in vitro* d'un clone plasmidique comportant la région +41 à +87 du gène de *Xenopus* débute environ 41 pb en amont dans le plasmide. De même, lorsque la région précédant l'ICR est allongée de quelques nucléotides la transcription commence un peu plus en aval dans le gène. Toutefois, le site exact et l'efficacité de l'initiation de la transcription dépendent fortement de sa séquence et de son environnement immédiat. Les nucléotides comportant des bases purines (A ou G) sont ainsi des cibles préférentielles de la Pol III. D'autre part, le site d'initiation est en général flanqué par des pyrimidines (C ou T) et les délétions situées un peu en amont dans l'espaceur provoquent des variations de sa position.

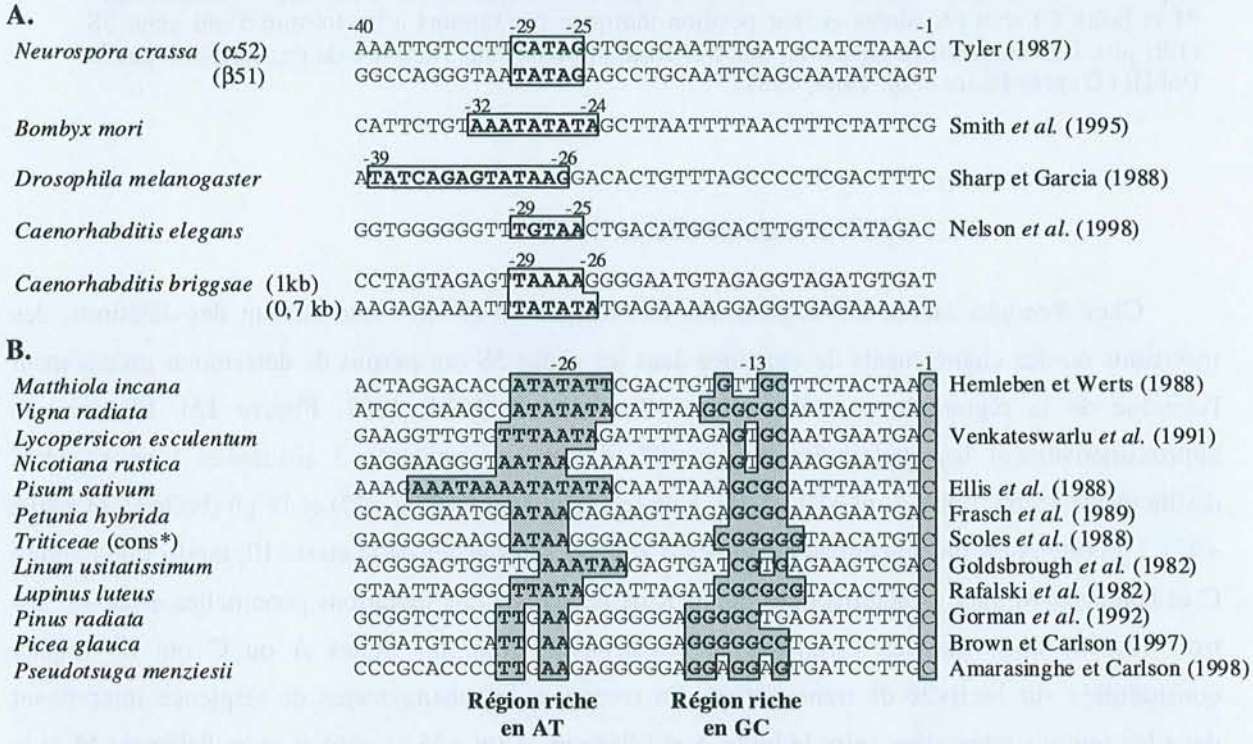



**Figure 14.** Signaux d'initiation de la transcription situés en amont des gènes 5S chez les eucaryotes.
**A.** Résultat des expériences de mutagénèse dirigée chez quelques organismes animaux. Les éléments indispensables sont encadrés et leur position dans l'espaceur est indiquée par rapport à l'extrémité 5' du gène 5S.
**B.** Alignement des régions correspondantes de quelques plantes supérieures. Les signaux potentiels d'initiation de la transcription proposés par Venkateswarlu *et al.* (1991) sont encadrés en gris (positions -26, -13 et -1). La région riche en AT des plantes correspond probablement aux séquences régulatrices identifiées chez les animaux. La région riche en GC semble être caractéristique des espaceurs de plantes. La cytosine en position -1 est très conservée.
* Séquence consensus établie à partir de 31 séquences de 14 espèces (9 genres).

β- Les séquences régulatrices de l'espaceur

Korn (1982) a montré chez *Xenopus laevis* que, si l'élimination de la plus grande partie de l'espaceur n'affecte pas l'efficacité de l'initiation de la transcription de l'ADNr 5S somatique, en revanche, les mutants portant des délétions dans la région qui précède immédiatement le gène (environ 40 pb) sont moins bien transcrits par la Pol III. Ces résultats, confirmés chez la plupart des autres organismes examinés, y compris les plantes (Hariharan *et al.* 1987), mettent bien en évidence le rôle régulateur de cette petite région de l'espaceur. Chez *Neurospora* (Tyler 1987), *Bombyx* (Smith *et al.* 1995), *Drosophila* (Sharp et Garcia 1988) ou *Caenorhabditis* (Nelson *et al.* 1998), il a même été montré qu'elle était aussi importante que l'ICR : son élimination interdit ou limite en effet de manière drastique la transcription (l'activité de la Pol III est réduite jusqu'à 30 fois chez *Neurospora crassa*, Tyler 1987). Les études par mutagénèse dirigée ont permis d'identifier précisément le signal en cause (**Figure 14**). Il s'agit d'un motif riche en A/T qui ressemble beaucoup à la boîte TATA nécessaire à la transcription par la polymérase II (30 pb en amont des gènes, Singer et Berg 1992). Sa taille varie de 5 pb chez *N. crassa* à 14 pb chez *D. melanogaster* et sa composition en base peut varier au sein d'une même espèce (ex : clones α52 et β51 de *N. crassa* et clones de 1 kb et 0,7 kb de *C. briggsae*). Toutefois, sa position est remarquablement conservée, 24 à 25 pb en amont du gène. Des séquences similaires sont observées chez les plantes autour de la position -26. Elles sont exclusivement constituées de nucléotides A et T, sauf dans le cas des trois seuls gymnospermes étudiés (motif TTG/CAA). Certaines d'entre elles ont une structure TATA partielle (*Matthiola incana*, *Pisum sativum*, *Lupinus luteus*) ou totale (*Vigna radiata*), tandis que d'autres consistent essentiellement en une succession de A (*Linum usitatissimum*) ou de T (*Lycopersicon esculentum*). L'absence de séquence consensus évidente indique que cette région riche en A/T n'a probablement pas la fonction des boîtes TATA (Venkateswarlu *et al.* 1991, Sastri *et al.* 1992). Elle serait plutôt similaire à la boîte -10 des promoteurs bactériens qui servent de site initial au déroulement des hélices d'ADN (Reznikoff *et al.* 1985). Les plantes possèdent par ailleurs une région riche en G/C aux environs de la position -13 qui pourrait aussi avoir un rôle régulateur (Venkateswarlu *et al.* 1991). Sa composition est en effet similaire à celle d'un promoteur bien connu de la transcription par la Pol II [le motif "GC" : (G/T)GGGCGG(G/A)(G/A) (C/T)] qui est impliqué dans la transcription de certains gènes viraux de classe III (voir Sorensen et Frederiksen 1991). On peut remarquer enfin la présence d'une pyrimidine C immédiatement en amont du gène (position -1) qui semble universelle chez les plantes et joue probablement un rôle dans l'initiation de la transcription (Venkateswarlu *et al.* 1991). Chez la levure en effet, l'activité de la Pol III est abaissée de près de 15 % par des mutations à ce niveau (Challice et Segall 1989).

b- Séquences de terminaison de la transcription

Alors que l'initiation de la transcription par la Pol III exige des séquences de régulation complexes dans le gène et dans l'espaceur ainsi que des facteurs protéiques (voir ci-dessous), la

terminaison requiert seulement la présence d'un groupe de résidus désoxythymidylates (poly T) près de l'extrémité 3' du gène (Ciliberto *et al.* 1983, Singer et Berg 1992). Chez *Xenopus borealis* par exemple, un groupe de 4 T occupe les positions + 119 à + 122 par rapport au site d'initiation de la transcription et la terminaison intervient le plus souvent à la position + 121. Des expériences de transcription *in vitro* de différentes matrices (Bogenhagen et Brown 1981) ont montré que la réduction de la taille du poly T abolit la terminaison. De même, l'efficacité de celle-ci est réduite si des séries de résidus A suivent ou précèdent le poly T. En revanche, la terminaison est améliorée lorsque le poly T est flanqué de résidus G ou C. La taille des poly T présents en aval des gènes 5S d'eucaryotes (**Figure 15**) est très variable (de 4 pb chez *P. glauca* à 21 pb chez *S. alba*) et il sont parfois discontinus ou multiples (ex : *V. radiata,* les *Triticeae*). Certains sont très proches de l'extrémité 3' du gène (ex : *B. mori*, *N. rustica*, *P. sativum*), d'autres plus éloignés dans l'espaceur (ex : *D. melanogaster*, *M. incana*, les 3 gymnospermes). Dans ce cas, les transcrits primaires subissent probablement une maturation post-transcriptionnelle car les ARNr 5S matures ont généralement une taille très peu différente de 120 nt (Szymanski *et al.* 1998). Ce phénomène a d'ailleurs été vérifié chez *D. melanogaster* (transcrits primaires de plus de 130 nt, Tschudi et Pirotta 1980, Pâques *et al.* 1995). Chez les plantes, il existe également souvent des séquences riches en AT plus éloignées dans l'espaceur (38 pb en amont du poly T chez les *Triticeae*) qui sont aussi susceptibles de jouer un rôle dans la terminaison de la transcription (Scoles *et al.* 1988).

| | +120 … +140 | |
|---|---|---|
| *Xenopus borealis* | gc**tttt**GCACTGCTAGCCGAATGC | Bogenhagen et Brown (1981) |
| *Saccharomyces cerevisiae* | atc**tttttttttttt**CCGCGCGCA | MacMahon *et al.* (1984) |
| *Bombyx mori* | c**ttttttt**GCCTGTGAAAACCAA | Smith *et al.* (1995) |
| *Drosophila melanogaster* | gcctCGTCCACAAC**ttttt**GCTGC | Tschudi et Pirrotta (1980) |
| *Matthiola incana* | acccCAC**tttttttttt**C**ttttt**GT | Hemleben et Werts (1988) |
| *Vigna radiata* | acctC**ttttttt**CACA**ttttttt**C | |
| *Lycopersicon esculentum* | ccctCC**ttttt**GTTGAAA**ttt**GAT | Venkateswarlu *et al.* (1991) |
| *Nicotiana rustica* | ccc**ttttttt**GGCAAAA**ttt**GTAT | |
| *Pisum sativum* | ctc**ttttttt**GCGA**ttttttt**AAATA | Ellis *et al.* (1988) |
| *Petunia hybrida* | cccgC**ttttt**GCCAAAATAACGTC | Frasch *et al.* (1989) |
| *Triticeae* (cons*) | tccc**ttttt**AA**ttttttttt**GCGC | Scoles *et al.* (1988) |
| *Linum usitatissimum* | cccc**TCCttttt**GCAA**ttttt**CTCC | Goldsbrough *et al.* (1982) |
| *Lupinus luteus* | ctc**ttttttt**ATAGGGAGGTGCGC | Rafalski *et al.* (1982) |
| *Sinapis alba* | acccT**ttttttttttttttttt**T | Capesius (1991) |
| *Pinus radiata* | cctcCCCCCCC**ttttttt**GTGCGGC | Gorman *et al.* (1992) |
| *Picea glauca* | ccttCCCCCC**tttt**GCATGGCTCC | Brown et Carlson (1997) |
| *Pseudotsuga menziesii* | cctcCGCACC**ttttttt**GCATGCCT | Amarasinghe et Carlson (1998) |

**Gène 5S** ← → **Espaceur**

**Figure 15.** Signaux de terminaison de la transcription des gènes 5S chez les eucaryotes. Le motif poly T fonctionnel caractérisé chez *Xenopus borealis* par des études de mutagénèse *in vitro* est encadré et sa position indiquée par rapport à l'extrémité 5' du gène (en minuscules). Les séquences similaires observées chez d'autres animaux et chez les plantes sont figurées en gras. * Séquence consensus établie à partir de 31 séquences de 14 espèces (9 genres).

**2- Modulation de l'activité de la Pol III par des facteurs de transcription**

La présence de la Pol III n'est pas suffisante pour initier la transcription de l'ADNr 5S. Il a été en effet montré chez le xénope qu'au moins trois autres protéines sont nécessaires (Wolffe et Brown 1988). Il s'agit du facteur de transcription TFIIIA spécifique des gènes 5S, et des facteurs de transcription TFIIIB et TFIIIC qui sont communs à tous les gènes de la classe III. Présents chez tous les eucaryotes, ces facteurs n'ont pas une forte spécificité et peuvent être aisément employés dans les systèmes de transcription hétérologues. Leur mode d'action est connu. Ils interagissent directement avec les régions de contrôle de l'ADNr 5S (ICR et région riche en A/T de l'espaceur) selon une séquence ordonnée (**Figure 16**) pour former des complexes de préinitiation stables reconnus par la Pol III. Le TFIIIA est une métalloprotéine de 38 kDa possédant neuf doigts à zinc homologues (20 kDa) formés chacun d'environ 30 acides aminés. Ces structures reployées forment le domaine de liaison avec l'ICR. Le contact s'établit précisément avec les nt 80-86 de la boîte C et dans une moindre mesure avec les nt 70-71 de l'élément intermédiaire M (Pieler *et al.* 1985, McBryant *et al.* 1995). La protéine couvre alors une large région du gène 5S (nt 45-96, Korn 1982), son carboxyle terminal étant orienté vers l'extrémité 5' du gène. Ce premier complexe protéine/ICR est cependant relativement instable. Il est rapidement et fortement consolidé par le TFIIIC qui interagit non seulement avec le TFIIIA, mais aussi avec l'ICR, probablement au niveau de la boîte A (Pieler *et al.* 1985, Sturges *et al.* 1995). Le facteur TFIIIB se lie alors aux deux précédents et au signal régulateur riche en A/T situé en amont du gène. Il serait directement responsable de l'initiation de la transcription. Le complexe de préinitiation qui résulte de cette cascade d'événements est très stable et peut être maintenu pendant plusieurs cycles de transcription par la Pol III. La production des ARNr 5S est simplement régulée par un mécanisme de rétrocontrôle (feed-back). Les ARNr 5S synthétisés sont en effet capables de se lier au TFIIIA pour former une particule RNP 7S de réserve. Ils sont donc des compétiteurs de l'ADNr 5S pour la fixation du TFIIIA (Christiansen *et al.* 1987, McBryant *et al.* 1995).

**3- Conséquence : conservation de la structure des ARNr 5S chez les plantes**

Les ARNr 5S sont suffisamment conservés pour permettre un alignement de séquence entre des organismes appartenant à des groupes taxinomiques très éloignés, des bactéries aux vertébrés et aux plantes (Szymanski *et al.* 1998). L'interprétation phylogénétique de ces comparaisons révèle une séparation nette entre le groupe des eubactéries (y compris les chloroplastes et les mitochondries), les archaebactéries et les eucaryotes (Hori *et al.* 1985, Hori et Osawa 1987). Il n'est pas possible en revanche d'établir des relations évolutives à des niveaux plus fins, chez les métazoaires (Halanych 1991), ou bien chez les plantes (Steele *et al.* 1991, voir la **Partie III-3-c-α** ci-dessous). Cette grande stabilité de la séquence des ARNr 5S est probablement en partie la conséquence des mécanismes de régulation de la transcription des gènes 5S que nous venons de décrire, peut-être en liaison avec des phénomènes de correction ("editing") des transcrits primaires (Szymanski *et al.* 1995).

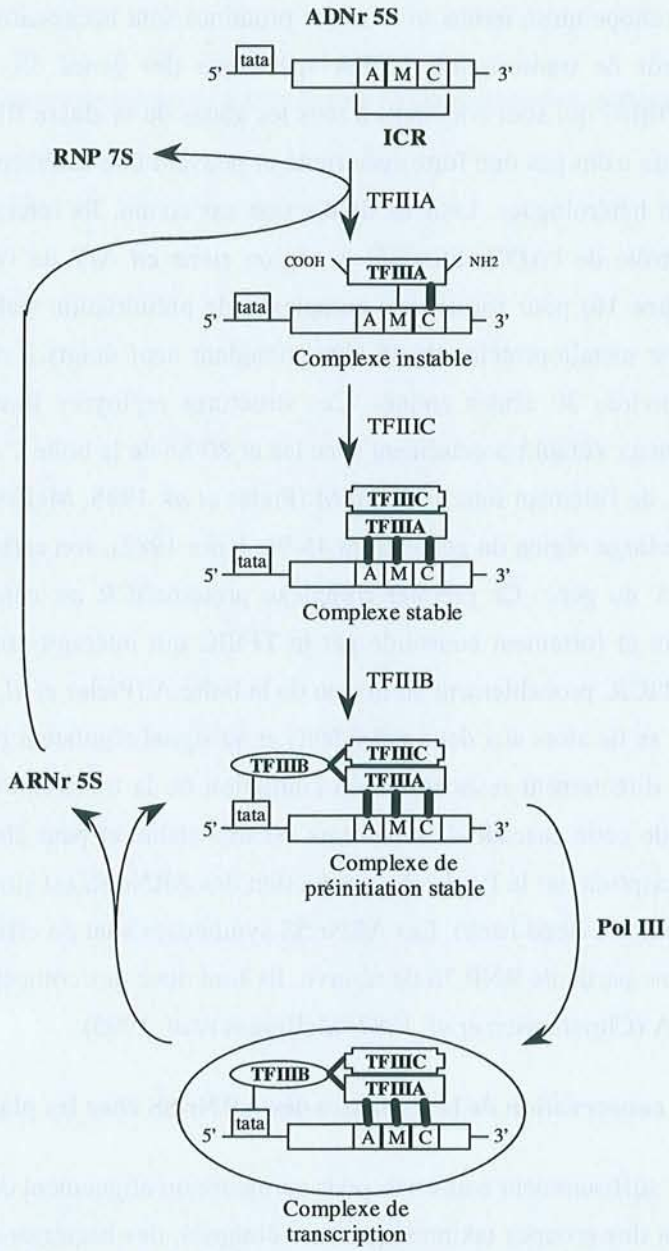



**Figure 16.** Etapes de l'initiation de la transcription de l'ADNr 5S par la polymérase III (Pol III) chez les eucaryotes (d'après Sturges *et al.* 1995, Singer et Berg 1992, Wolffe et Brown 1988, Pieler *et al.* 1985). L'interaction des facteurs de transcription (TFIIIA, TFIIIB et TFIIIC) avec la région de contrôle interne du gène 5S (ICR) et la région régulatrice riche en A/T (tata) de l'espaceur permet la formation d'un complexe de préinitiation stable qui est conservé pendant plusieurs cycles de transcription. TFIIIA interagit surtout avec la boîte C de l'ICR, tandis que TFIIIC stabilise le complexe, probablement par des liaisons avec TFIIIA et la boîte A de l'ICR. Enfin, le positionnement de TFIIIB en amont du gène nécessite la présence des deux facteurs précédents. Il est en relation avec la région régulatrice riche en A/T de l'espaceur et guide probablement la Pol III vers le bon site d'initiation. L'ARNr 5S est capable de se lier au TFIIIA pour former une particule RNP 7S de réserve (régulation en "feed-back").

L'existence d'une région promotrice intragénique (ICR) qui interagit avec différents facteurs de transcription doit en effet imposer des contraintes structurales considérables, à la fois sur le gène et sur l'ARNr 5S. Le fait que ces gènes soient facilement transcrits par des machineries enzymatiques hétérologues est clairement révélateur de la forte pression sélective à laquelle ils sont soumis. La nécessité de l'association entre le facteur TFIIIA et l'ARNr 5S pour former la particule RNP 7S introduit probablement aussi des contraintes de structure supplémentaires car les régions impliquées sont en grande partie externes à l'ICR (McBryant *et al.* 1995). Enfin, bien évidemment, les éventuels changements de séquence pouvant survenir au sein des ARNr 5S doivent préserver les structures secondaire et tertiaire indispensables au maintien de leur activité au sein des ribosomes.

L'étude des différences nucléotidiques entre les 24 séquences (23 espèces) connues d'ARNr 5S de plantes angiospermes (Szymanski *et al.* 1998) confirme leur très grande conservation en taille (118-120 nt) et en séquence (seulement 21 sites polymorphes). Le modèle de structures secondaire et tertiaire proposé par Barciszewska *et al.* (1994) montre que les variations sont essentiellement situées près des extrémités du gène, en dehors des zones d'interaction des facteurs de transcription (en particulier TFIIIA) avec l'ICR ou l'ARNr 5S (**Figure 17**). Les 9 séquences d'ARNr 5S déterminées chez les gymnospermes présentent une distribution des variations de séquence similaire (voir **Tableau 16** et **Partie III-1** du **Chapitre IV**).

## II- Organisation génomique de l'ADNr 5S

### 1- Localisation génomique

#### a- Localisation chromosomique

Les techniques d'hybridation *in situ* à l'aide de sondes marquées (souvent fluorescentes) sont en plein essor depuis quelques années. Elles ont déjà permis d'obtenir la localisation chromosomique fine de l'ADNr 5S chez 8 gymnospermes et un grand nombre d'angiospermes d'intérêt agronomique (**Tableau 12**). L'hybridation *in situ* permet de révéler avec fiabilité non seulement les sites majeurs composés de milliers de copies (**Partie III-1** ci-dessous), mais aussi les sites mineurs avec des limites de détection de l'ordre de la dizaine de kpb. Les espèces diploïdes ont généralement moins de 5 loci d'ADNr 5S. Plusieurs gymnospermes (*Larix decidua*, 3 espèces de *Picea*, *Pseudotsuga menziesii*) et angiospermes (*Beta vulgaris*, *Hypochaeris*, *Lycopersicon esculentum*, *Malus x domestica*, *etc.*) ne possèdent même qu'un seul locus.

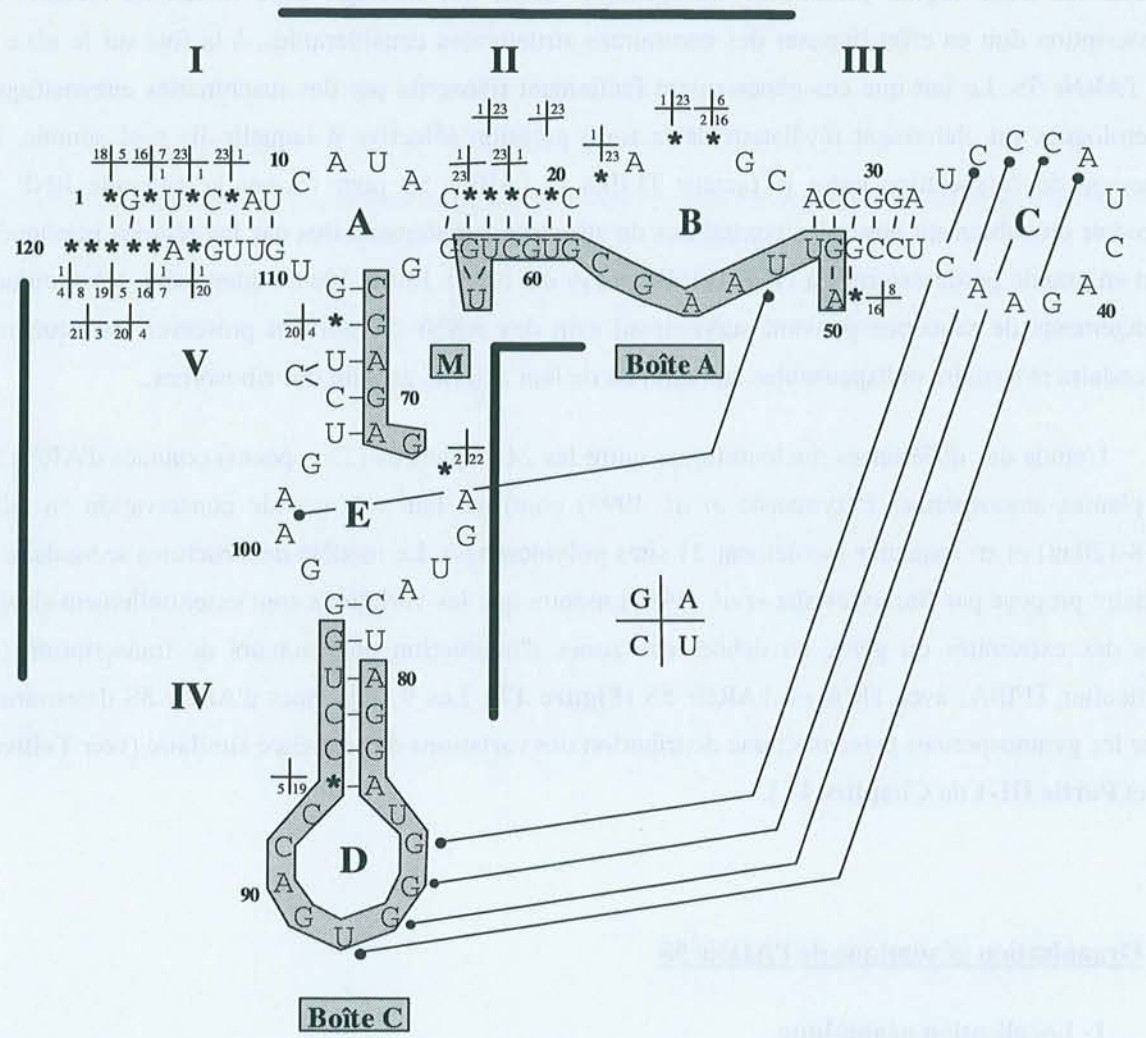



**Figure 17.** Conservation de la structure des ARNr 5S chez les plantes angiospermes.
Le modèle de structure secondaire proposé par Barciszewska *et al.* (1994) comprend cinq régions bicaténaires (I-V), 3 boucles internes (A, B et E) et 2 boucles externes (C et D). Les interactions tertiaires entre les boucles C/D d'une part, et B/E d'autre part, sont figurées par des traits fins. Les positions nucléotidiques variables entre les 24 molécules (23 espèces) connues sont indiquées par des étoiles. La distribution des bases G, A, C et U au niveau de ces sites spécifiques figure en regard.
Les zones encadrées en gris correspondent à l'ICR (boîte A, élément M et boîte C) des gènes 5S de *Xenopus* (Pieler *et al.* 1987). Les barres noires désignent les zones d'interaction entre les ARNr 5S et le facteur de transcription TFIIIA (Mc Bryant *et al.* 1995).

Une dispersion importante de séries d'ADNr 5S de taille très variable sur la plupart ou sur tous les chromosomes a toutefois été rapportée chez *Linum usitatissimum* (Schneeberger *et al.* 1989) et *Pinus radiata* (Gorman *et al.* 1992) et une situation similaire existerait chez différentes espèces du genre *Secale* (discuté par Kellogg et Appels 1995) où, en plus des 2-3 sites majeurs, il y aurait de nombreux sites mineurs disséminés (Reddy et Appels 1989). Il est peu probable que tous ces loci soient

fonctionnels. L'analyse des séquences des gènes 5S chez *L. usitatissimum* indique en effet que la plupart sont probablement des pseudogènes (Schneeberger et Cullis 1992). De même, parmi les 10 sites comptabilisés chez *Hordeum vulgare* (Leitch et Heslopp-Harrison 1993, Fukui *et al.* 1994), 4 seulement sont détectés par une sonde constituée de la région codante (*a priori* les séries fonctionnelles) alors que tous sont révélés par une sonde composée de l'unité entière. Certains loci pourraient donc être composés uniquement de séries d'espaceurs intergéniques et représenteraient selon Fukui *et al.* (1994) les vestiges d'anciens loci fonctionnels. Certains de ces sites silencieux sont d'ailleurs situés sur les mêmes chromosomes que les loci fonctionnels et parfois à proximité de ceux-ci (ex : les deux loci en position subterminale du chromosome 3).

Chez les angiospermes, le nombre de loci est souvent (mais pas toujours) plus élevé lorsque le niveau de ploïdie (x) augmente. Dans le genre *Saccharum* par exemple (x = 8 ou 10), 6 à 14 loci ont pu être détectés (D'Hont *et al.* 1998), tandis que chez *Triticum aestivum* (x = 6), les séries de l'ADNr 5S occupent 6 loci distribués sur tous les chromosomes homéologues des groupes 1 et 5 (Mukai *et al.* 1990). De même, le nombre de loci de l'espèce tétraploïde *A. vaviloviana* (8 sites, Katsiotis *et al.* 1997) est le double de celui de l'espèce diploïde *Avena strigosa* (4 sites) et une situation comparable se retrouve dans le genre *Gossypium* (Hanson *et al.* 1996). Lorsque plusieurs sites sont observés dans un génome, il est rare qu'ils aient la même importance. Les génomes diploïdes ne possèdent bien souvent qu'un site majeur et plusieurs sites mineurs (*P. elliotii*, *P. sylvestris*, *A. thaliana*, *P. sativum*, *etc.*) alors que dans les génomes polyploïdes, les différents loci sont généralement de taille encore plus variable (*Saccharum*, *Triticum aestivum*). En dehors de ces différences liées à la ploïdie, le nombre de sites est stable au sein d'une même espèce. En revanche, il peut varier entre les espèces d'un même genre (ex : *Aegilops*, *Hordeum*, *Secale*, *Trifolium*).

**Tableau 12.** Nombre et distribution des loci d'ADNr 5S chez les plantes supérieures.

| Organisme[a] | Ploïdie | N[b] | Chromosome[c] | Position[d] | Référence |
|---|---|---|---|---|---|
| Gymnospermes | | | | | |
| *Larix decidua* | 2x = 24 | 1 | 3 | ST (bc) | Lubaretz *et al.* (1996) |
| *Picea abies* | 2x = 24 | 1 | 2 | I (bl) | |
| *P. glauca + P. sitchensis* | 2x = 24 | 1 | 5 | I | Brown et Carlson (1997) |
| *Pinus elliottii* | 2x = 24 | 3 | 1 (m) | T(bl) | Doudrick *et al.* (1995) |
| | | | 2 (m) | T (bl) | |
| | | | 9 (M) | I | |
| *P. radiata* | 2x = 24 | d | Tous | | Gorman *et al.* (1992) |
| *P. sylvestris* | 2x = 24 | 4 | (M) | I | Lubaretz *et al.* (1996) |
| *Pseudotsuga menziesii* | 2x = 26 | 1 | 3 | ST (bl) | Amarasinghe et Carlson (1998) |

**Tableau 12 (suite).** Nombre et distribution des loci d'ADNr 5S chez les plantes supérieures.

| Organisme[a] | Ploïdie | N[b] | Chromosome[c] | Position[d] | Référence |
|---|---|---|---|---|---|
| Angiospermes | | | | | |
| *Aegilops* (12 sp) | 2x = 14 | 1-2 | 1 (M ou Ø) | ST (bc ou bl) | Badaeva *et al.* (1996) |
| | | | 5 (M, m ou Ø) | I (bc) | |
| *Arabidopsis thaliana* | 2x = 10 | 3 | 5 (M) | P (bl) | Murata *et al.* (1997) |
| | | | 4 (m) | P (bc) | |
| | | | 3 (m) | I (bl) | |
| *Avena* strigosa | 2x = 14 | 4 | | | Katsiotis *et al.* (1997) |
| *A. vaviloviana* | 4x = 28 | 8 | | | |
| *Beta vulgaris* | 2x = 18 | 1 | 4 | P (bc) | Schondelmaier *et al.* (1997) |
| *Gossypium* (4 sp) | 2x = 26 | 1 | | P (bc) | Hanson *et al.* (1996) |
| *G. hirsutum* | 4x = 52 | 2 | (M) | P (bl) | |
| | | | (m) | P (bc) | |
| *Glycine max* (2 var) | 2x = 40 | 1 | | | Danna *et al.* (1996) |
| *G. soja* | 2x = 40 | 1 | | | |
| *Hordeum* (9 sp sauvages) | 2x = 14 | 1-3 | | T, ST, I ou P | Taketa *et al.* (1999) |
| | 4x = 28 | | | | |
| *H. vulgare* (2 var) | 2x = 14 | 4 | 1 (i) | P (bc) | Leitch et |
| | | + | 2 (M) | ST (bl) | Heslopp-Harrison (1993) |
| | | ***6**** | 3 (i) | ST (bl) + ***ST (bl)*** | Fukui *et al.* (1994) |
| | | | 4 (m) | T (bl) + ***ST (bl)*** | |
| | | | ***6 (m)*** | ***P + T*** | |
| | | | ***7 (m)*** | ***P (bc) + I (bl)*** | |
| *Hypochaeris* (10 sp) | 2x = 8-12 | 1 | 2, 3, 4 ou 6 | T, I ou P (bc ou bl) | Cerbah *et al.* (1998) |
| *H. cretensis* | 2x = 6 | 3 | 1 | P (bl) | |
| | | | 1 | T (bc) | |
| | | | 2 | T | |
| *Linum usitatissimum* | 2x = 30 | d | La plupart | | Schneeberger *et al.* (1989) |
| *Lycopersicon esculentum* | 2x = 24 | 1 | 1 | P (bc) | Lapitan *et al.* (1991) |
| *Malus x domestica* | 2x = 34 | 1 | | P (bc) | Schuster *et al.* (1997) |
| *Neonotonia wightii* | 2x = 22 | 1 | | | Danna *et al.* (1996) |
| *Oryza sativa* ssp *indica* | 2x = 24 | 1 | 9 | T (bc) | Song et gustafson (1993) |
| *O. sativa* ssp *japonica* | 2x = 24 | 1 | 11 | P (bc) | Kamisugi *et al.* (1994) |
| *Phaseolus vulgaris* | 2x = 22 | ≥ 3 | | | Danna *et al.* (1996) |
| *Pisum sativum* | 2x = 14 | 3 | (M) | I | Ellis *et al.* (1988) |
| | | | (m) | T | |
| | | | (m) | T | |
| *Populus* (5 sp) | 2x = 38 | 2 | | | Prado *et al.* (1996) |
| *Saccharum officinarum* | 10x = 80 | 8 | (M, i, m) | I | D'Hont *et al.* (1998) |
| *S. robustum* (3 clones) | 10x = 60-80 | 6-8 | (M, i, m) | I | |
| *S. spontaneum* (4 clones) | 8x = 64-112 | 8-14 | (M, i, m) | I | |
| *Secale* (4 sp) | 2x = 14 | 2-3 | 1 (M) | I (bc) | Cuadrado et Jouve (1997) |
| | | + | 3 (M ou Ø) | ST (bc) | Reddy et Appels (1989) |
| | | d | 5 (M) | ST (bc) | |
| *Thapsia garganica* | 2x = 22 | 1 | | P (bl) | Rasmussen et Avato (1998) |
| *Trifolium* (6 sp + 2 ssp) | 2x = 16 | 1-2 | (M) | P (bl) | Ansari *et al.* (1999) |
| | 4x = 32 | | (i, m) | P (bl, bc) | |
| *Triticum aestivum* | 6x = 42 | 6 | 1A (m) | T | Mukai *et al.* (1990) |
| | | | 1B (M) | T (bc) | |
| | | | 1D (i) | T (bc) | |
| | | | 5A (i) | ST (bc) | |
| | | | 5B (M) | ST (bc) | |
| | | | 5D (i) | ST (bc) | |
| *T. tauschii* | 2x = 14 | 2 | 1D (M) | | Lagudah *et al.* (1989) |
| | | | 5D (M) | | |
| *Vicia faba* | 2x = 12 | 2 | | | Knälmann et Burger (1977) |

[a] sp : espèce ; ssp : sous-espèce ; var : variété. [b] Nombre de sites. d : sites nombreux et dispersés.
[c] Numéro du chromosome porteur. m : site mineur ; i : site intermédiaire ; M : site majeur ; Ø : site absent.
[d] T : terminale ; ST : subterminale ; I : intercalaire ; P : proximale ; bl : bras long ; bc : bras court.
* Sites détectés avec une sonde correspondant à l'espaceur.

De même, selon les espèces d'un même genre, un locus particulier peut occuper différents chromosomes (ex : *Hypochaeris*) ou différents bras d'un chromosome (ex : *Aegilops*, *Trifolium*). Cependant, l'ADNr 5S occupe le plus souvent des positions proches du centromère (positions

proximales) ou des télomères (positions subterminales ou terminales). En raison de la fréquence relativement faible des événements de recombinaison dans ces régions (Charlesworth *et al.* 1994), une telle localisation privilégiée est probablement plus stable au cours de l'évolution (Hanson *et al.* 1996).

b- Cartographie génétique

La localisation génomique de régions de l'ADN déterminant des caractères quantitatifs intéressants (rendement, qualité, croissance, etc.) est souvent établie par l'analyse de liaison avec des marqueurs, le plus souvent anonymes (RFLP, RAPD, microsatellites, etc.), placés sur des cartes génétiques (Henry 1997). Cette connaissance permet alors d'élaborer des stratégies de sélection assistée par marqueurs et de clonage de gènes (construction de banques d'ADN spécifiques par exemple). Le succès de telles entreprises dépend largement de la précision avec laquelle les distances physiques entre marqueurs sont définies. Or, de ce point de vue, les cartes génétiques sont souvent imprécises car les événements de recombinaison (dont la fréquence sert à évaluer les distances entre marqueurs) ne sont pas distribués aléatoirement le long des chromosomes. La plupart interviennent dans les régions distales, ce qui se traduit en particulier par des compressions de marqueurs autour du centromère (Schondelmaier *et al.* 1997). Une manière d'accéder plus précisément aux distances physiques consiste à intégrer dans les cartes génétiques des marqueurs cytologiques de référence comme les organisateurs nucléolaires (NORs), les bandes C ou des sites d'hybridation *in situ* de familles de séquences répétées. En raison de leur universalité et de leur localisation physique relativement aisée (**paragraphe précédent**), les loci d'ADNr 5S (ou d'autres familles de séquences répétées en tandem) sont considérés comme d'excellents jalons permettant d'assigner un groupe de liaison génétique à un chromosome et d'établir la correspondance entre distances physique et génétique (Lapitan *et al.* 1991). La cartographie génétique des séries de l'ADNr 5S est toutefois souvent compliquée soit par l'absence de polymorphisme, soit au contraire par des profils de bandes trop complexes pour autoriser l'analyse de ségrégation dans le cadre d'un croisement (Moran *et al.* 1992, Schondelmaier *et al.* 1997). De plus, d'éventuels échanges intra- et interchromosomiques (évolution concertée, **Partie II-3-d** du **Chapitre I**) peuvent conduire à une hérédité non-mendélienne plus ou moins persistante de ce type de marqueur (Dover 1982). A notre connaissance, la localisation génétique de l'ADNr 5S n'a été obtenue que pour la tomate (*Lycopersicon esculentum*, Lapitan *et al.* 1991), l'orge (*Hordeum vulgare*, Kanazin *et al.* 1993, Laurie *et al.* 1993) et la betterave sucrière (*Beta vulgaris*, Schondelmaier *et al.* 1997). Kanazin *et al.* (1993) ont par exemple exploité une différence d'intensité d'amplification des espaceurs pour localiser deux loci d'ADNr 5S sur les groupes de liaison correspondant aux chromosomes 2 (locus ADNr 5S-A, unités courtes) et 3 (locus ADNr 5S-B, unités longues, mais voir Baum et Johnson 1994) de l'orge. Le locus ADNr 5S-B est lié à un QTL (Quantitative Trait Loci) majeur impliqué dans le rendement en grain. La localisation génétique du locus ADNr 5S-A a aussi été obtenue par Laurie *et al.* (1993) grâce à l'analyse d'un profil RFLP complexe généré par l'endonucléase *Taq* I. La comparaison des positions physiques (subterminale, 73 % de la longueur du bras long) et génétiques (proximale, 33 % de la longueur de la carte

correspondant au bras long) de ce locus révèle une sérieuse discordance qui s'explique par un déficit de recombinaison dans la région du centromère. La situation est différente chez *L. esculentum* (Lapitan *et al.* 1991) et *B. vulgaris* (Schondelmaier *et al.* 1997). Les localisations physique (hybridation *in situ*) et génétique (analyses de polymorphismes RFLP *Taq* I ou *Xba* I, respectivement) de l'unique locus identifié dans les deux cas, sont en effet tout à fait colinéaires (position proche du centromère des chromosomes 1 de la tomate et IV de la betterave).

## 2- Méthylation de l'ADNr 5S

La méthylation de l'ADN semble plus importante chez les plantes que chez les animaux (Gruenbaum *et al.* 1981). Au moins 30 % des résidus "cytosine" du génome des plantes supérieures seraient méthylés (Cm : 5-méthylcytosine), essentiellement au niveau des motifs symétriques dinucléotidiques CG et des trinucléotides CWG et CCG (Gruenbaum *et al.* 1981, Bezdek *et al.* 1992, Jeddoloh et Richards 1996). La méthylation de l'ADN interviendrait dans le contrôle de la transcription par les polymérases I et II, la condensation de la chromatine et l'inactivation des chromosomes X (Fulnecek *et al.* 1998).

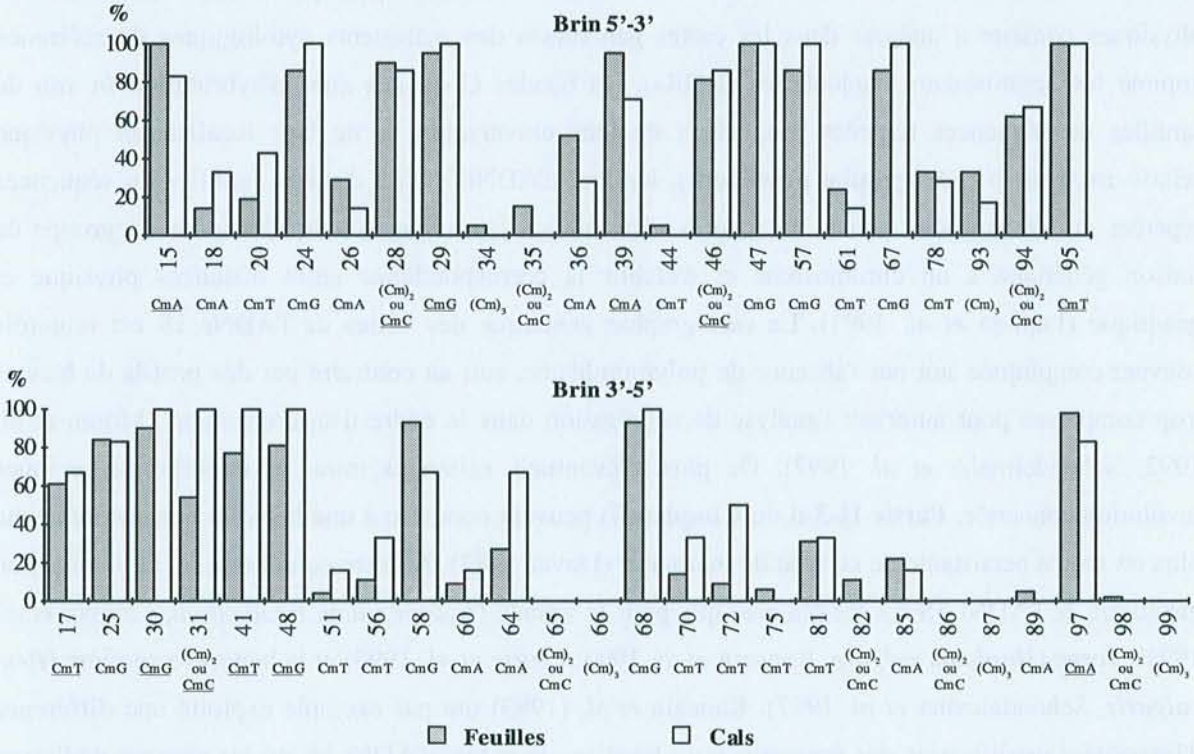



**Figure 18.** Fréquence de méthylation (%) des résidus "cytosine" (Cm) à différentes positions nucléotidiques (en abscisse) sur les brins 5'-3' et 3'-5' des gènes 5S de *Nicotiana tabacum* et pour différents tissus (feuilles et cals). Adapté de Fulnecek *et al.* (1998). Les positions sont numérotées par rapport au site d'initiation de la transcription. La nature du site cible (dans le sens 5'-3') est indiquée sous chaque position. Les motifs CNG sont soulignés.

Son rôle n'est toutefois pas encore bien compris pour les gènes de la classe III, comme l'ADNr 5S. Plusieurs études chez les animaux suggèrent que la densité et la distribution de la méthylation influent sur un signal de positionnement des nucléosomes (Dong *et al.* 1990) et modulent l'accessibilité de l'ADNr 5S (Hodges-Garcia et Hagerman 1992), en particulier dans la région de l'ICR (Bogenhagen *et al.* 1980). Goldsbrough *et al.* (1982) indiquent aussi que la plupart des transitions C→T seraient la conséquence directe d'une désamination des Cm. Environ 78 % des mutations détectées entre 5 clones d'ADNr 5S chez le lin (*Linum usitatissimum*) sont en effet des transitions C→ T et concernent des motifs CG (45 % des cas) ou CWG (18 % des cas).

Chez les plantes supérieures, les unités d'ADNr 5S sont particulièrement riches en Cm. De récents travaux de séquençage génomique concernant *Nicotiana tabacum* (Fulnecek *et al.* 1998) ont estimé leur fréquence moyenne à 53 % sur le brin 5'-3' (21 clones, 21 positions étudiées) et 34 % sur le brin 3'-5' (48 clones, 26 positions étudiées) de la région transcrite (**Figure 18**). Les profils de méthylation observés se sont révélés tout à fait comparables entre différentes plantes et différents tissus (feuille et cal). Aucun des gènes analysés ni aucune position "cytosine" ne semblent épargnés par ce phénomène. Comme dans l'ensemble du génome, la méthylation affecte préférentiellement les dinucléotides CG (CmG = CmW > CCm > CmCm >> CmC) et les trinucléotides CNG (CmGG = CCmG = CmCmG = CmWG >> CmCG). Elle a pour conséquence d'inactiver près de 100 % des sites de restriction par l'endonucléase *Eco*R II (sensible à la méthylation de la cytosine interne, CCmWGG), 95 % des sites *Hpa* II (CCmGG) et 80 % des sites *Msp* I (CmCGG). Des estimations du même ordre ont été obtenues chez le lin (Goldsbrough *et al.* 1982) à deux sites de restriction par *Hpa* II ou *Msp* I localisés dans le gène (position + 28) et l'espaceur (+ 139). La fréquence de coupure par *Hpa* II n'est ainsi que de 2,6 % à ces deux positions, tandis que *Msp* I n'est efficace que dans 65 % et 55 % des cas, respectivement. La méthylation inactive aussi environ 16 % des sites de restriction *Bam*H I (GGATCmC) situés en position + 30 dans le gène. Ce site est très utilisé pour caractériser et cloner les unités d'ADNr 5S chez les plantes supérieures car il est très conservé et généralement unique. Les profils de restriction obtenus par RFLP de l'ADN génomique se présentent systématiquement sous la forme d'une échelle de multimères (monomères, dimères, trimères, etc.) caractéristique des séquences répétées en tandem. Même si les mutations ponctuelles ne peuvent pas être totalement écartées (voir Kellogg et Appels 1995 et **Partie III-3** ci-dessous), cette digestion incomplète des séries de l'ADNr 5S est généralement attribuée aux effets de la méthylation différentielle des unités au site *Bam*H I (Mascia *et al.* 1981, Rafalski *et al.* 1982, Ellis *et al.* 1988, Dvorák *et al.* 1989, Reddy et Appels 1989, Gottlob-McHugh *et al.* 1990, Campell *et al.* 1992, Johnson *et al.* 1992, McIntyre *et al.* 1992, Bhatia *et al.* 1993, Zanke *et al.* 1995, Amarasinghe et Carlson 1998). Plusieurs auteurs soulignent toutefois que la méthylation du site *Bam*H I est probablement moins importante près des bordures des séries car, dans les banques d'ADN génomique, les régions flanquantes sont plus facilement détectées que les régions centrales (Campell *et al.* 1992, Schneeberger *et al.* 1989, Ellis *et al.* 1988).

**Tableau 13.** Nombre d'unités d'ADNr 5S chez les plantes supérieures.

| Plante | Source | N[a] | Référence |
|---|---|---|---|
| Gymnospermes | | | |
| *Picea glauca* | | 1170 | Brown et Carlson (1997) |
| *Pinus radiata* | Génotype 50048 x 80055 | 1500 | Gorman *et al.* (1992) |
| | 14 individus | 4000-9250 | Moran *et al.* (1992) |
| | Unités courtes | 3850-7500 | |
| | Unités longues | 150-1750 | |
| Angiospermes | | | |
| *Arabidopsis thaliana* | 4 écotypes | 1000 | Campell *et al.* (1992) |
| *Glycine max* | Cultivar BSR101 (tiges, feuilles et racines) | 600 | Danna *et al.* (1996) |
| | Cultivar PI 437.654 | 4600 | |
| | Cultivar Bragg | 3000-4000 | Kolchinsky et Gresshoff (1992) |
| *Glycine soja* | Cultivar PI 468.916 | 3100 | Danna *et al.* (1996) |
| | Cultivar PI 468.397 | 3000-4000 | Kolchinsky et Gresshoff (1992) |
| *Gossypium robinsonii* (C) | | 1470 | Cronn *et al.* (1996) |
| *G. herbaceum* (A) | | 1707 | |
| *G. arboreum* (A) | | 3775 | |
| *G. gossypioides* (D) | | 572 | |
| *G. trilobum* (D) | | 710 | |
| *G. aridum* (D) | | 842 | |
| *G. thurberi* (D) | | 1035 | |
| *G. laxum* (D) | | 1380 | |
| *G. armourianum* (D) | | 2102 | |
| *G. harknessii* (D) | | 2192 | |
| *G. schwendimanii* (D) | | 2307 | |
| *G. raimondii* (D) | | 2365 | |
| *G. lobatum* (D) | | 2915 | |
| *G. klotzschianum* (D) | | 3465 | |
| *G. turneri* (D) | | 4247 | |
| *G. davidsonii* (D) | | 5140 | |
| *G. hirsutum* (AD) | | 5595 | |
| *G ossypium* [2(AD)]* | | 7007 | |
| *G. mustelinum* (AD) | | 10922 | |
| *G. tomentosum* (AD) | | 11145 | |
| *G. bardadense* (AD) | | 11757 | |
| *Hordeum vulgare* | Lignée haploïde doublée $2^{S}3S$ | 4600 | Kanazin *et al.* (1993) |
| | Lignée haploïde doublée $2^{M}3M$ | 5500 | |
| | Lignée haploïde doublée $2^{S}3M$ | 1800 | |
| | Lignée haploïde doublée $2^{M}3S$ | 6900 | |
| *Linum usitatissimum* | Lignée P1 | 58500 | Goldsbrough *et al.* (1981) |
| | Lignée L1 | 40150 | |
| | Lignée S1 | 26150 | |
| | Lignée LH | 24800 | |
| | Lignée SH | 50950 | |
| *Lycopersicon esculentum* | | 1000 | Lapitan *et al.* (1991) |
| *Matthiola incana* | | 3600 | Hemleben et Werts (1988) |
| *Neonotonia wightii* | | 1300 | Danna *et al.* (1996) |
| *Phaseolus vulgaris* | | 1495 | |
| | Série A | 530 | |
| | Série B | 720 | |
| | Série C | 245 | |
| *Pisum sativum* | 3 cultivars | 3000-7000 | Ellis *et al.* (1988) |
| *Teramnus labialis* | | 300-450 | Danna *et al.* (1996) |
| *Triticum aestivum* | Variété Chinese Spring | 5000 | Appels et al. (1980) |
| | Variété Chinese Spring | 7000-8000 | Röder *et al.* (1992) |
| | Unités courtes (1B) | 3244-3610 | |
| | Unités courtes (1D) | 1098-1476 | |
| | Unités longues (5B) | 1670-1940 | |
| | Unités longues (5D) | 1140 | |
| *Vigna radiata* | | 4300 | Hemleben et Werts (1988) |

[a] Nombre d'unités par génome haploïde (c)

* Allopolyploïde synthétique doublé (colchicine) issu d'un hybride F1 stérile entre *G. arboreum* et *G. thurberi*

## III- Evolution de l'ADNr 5S

### 1- Variations du nombre d'unités

Le nombre d'unités d'ADNr 5S par génome haploïde est extrêmement labile chez les plantes supérieures (**Tableau 13**). Il varie d'environ 570 copies chez *Gossypium gossypioides* (Cronn *et al.* 1996) à plus de 58000 copies dans une lignée de *Linum usitatissimum* (Goldsbrough *et al.* 1982). Les contractions et expansions de séries interviennent apparemment rapidement au sein des genres et des espèces. Cronn *et al.* (1996) ont ainsi montré qu'au sein d'un ensemble de 20 espèces diploïdes (génomes C, A et D) et allopolyploïdes (génome AD) issues de l'ensemble de l'aire de répartition du genre *Gossypium* en Europe et en Asie (A), en Amérique (D et AD) et en Australie (C), l'amplitude des variations est d'environ 20 fois (572-11757). Le nombre de copies varie du simple au double chez les espèces des groupes A (1707-3775, moyenne : 2740) et AD (5595-11757, moyenne : 9285) et d'environ 10 fois dans le groupe D (572-5140, moyenne : 2250). De même, dans la sous-section monophylétique *Erioxylum* de ce dernier groupe, des différences importantes peuvent être notées entre les espèces *G. aridum* (842 copies), *G. laxum* (1380 copies), *G. schwendimanii* (2307 copies) et *G. lobatum* (2915 copies). Les séries d'ADNr 5S des allopolyploïdes naturels (en moyenne 9854 copies) ont probablement augmenté de taille depuis leur formation car le nombre de copies cumulé des deux espèces progénitrices présumées (*G. herbaceum* et *G. raimondii*) est environ 2 fois plus faible (1707 + 2365 = 4072). En revanche, chez les deux seules espèces de cotonniers diploïdes connus pour s'être naturellement hybridés au cours de leur histoire évolutive (*G. gossypioides* et *G. aridum*), le nombre de copies est réduit par rapport aux autres espèces. Cronn *et al.* (1996) interprètent cette situation comme le résultat de phénomènes de contraction des séries qui permettraient d'éliminer les variations de séquences excessives et de contribuer ainsi à la stabilisation des hybrides.

Le nombre d'unités varie fortement au sein des espèces. L'amplitude des variations observées dans deux populations naturelles de *Triticum tauschii* (Lagudah *et al.* 1989) est de 10 à 20 fois. De même, des différences importantes ont été rapportées entre divers génotypes de *Pinus radiata* (1500-9250 copies), 3 cultivars de *Glycine max* (600-4600 copies) et de *Pisum sativum* (3000-6000 copies) et 5 lignées de *Linum usitatissimum* (24800-58500 copies). Il faut souligner que, dans ce dernier cas, les variations ont été très rapides car 4 des lignées (L1, S1, LH et SH) dérivent d'une lignée de référence (P1) après croissance dans des conditions d'environnement différentes (Goldsbrough *et al.* 1981). Une partie des variations observées est probablement liée à la présence de plusieurs loci chez la plupart de ces espèces (**Tableau 12**). Le nombre d'unités évolue en effet de manière indépendante à chaque locus (Lagudah *et al.* 1989) et le brassage chromosomique crée donc constamment de nouvelles combinaisons. Kanazin *et al.* (1993) ont ainsi étudié 4 classes de lignées haploïdes doublées (2 classes parentales et 2 classes recombinantes) obtenues à partir d'un hybride entre deux variétés d'orge (Morex et Steptoe) qui diffèrent par la taille des séries situées sur les paires de chromosomes 2

et 3 (Morex a beaucoup d'unités sur les chromosomes 2 et peu sur les chromosomes 3 ; c'est l'inverse pour Steptoe).

Les deux classes parentales $2^M3^M$ (porteuse des chromosomes 2 et 3 de la variété Morex) et $2^S3^S$ (porteuse des chromosomes 2 et 3 de la lignée Steptoe) ont un nombre d'unités similaires (5500 et 4600) tandis que les deux classes recombinantes ($2^M3^S$ et $2^S3^M$) sont nettement différentes avec 6900 et 1800 copies, respectivement. Ces variations du nombre total de copies entre les lignées parentales et leurs descendants n'a toutefois pas de conséquences phénotypiques (Kanazin *et al.* 1993). Il faut aussi remarquer que ces auteurs n'ont obtenu que 4 classes de lignées haploïdes doublées à partir de ce croisement (correspondant au 4 associations chromosomiques possibles). La taille des séries sur chaque chromosome est donc apparemment stable d'une génération à l'autre chez l'orge (voir aussi Röder *et al.* 1992). Danna *et al.* (1996) ont d'ailleurs démontré sur un autre matériel (*Glycine max*) que la taille des séries n'évolue pas au cours du développement. L'unique série de 600 unités (200 kpb) du cultivars BSR101 a en effet la même taille dans les racines, les tiges et les feuilles d'une cinquantaine de jeunes plants.

En plus du brassage entre chromosomes, le mécanisme moléculaire généralement proposé pour expliquer l'instabilité importante du nombre de copies d'ADNr 5S chez les plantes est le crossing-over inégal intrachromosomique (Scoles *et al.* 1988, Dvorák *et al.* 1989, Reddy et Appels 1989, Sastri *et al.* 1992, Kanazin *et al.* 1993, Danna *et al.* 1996). D'autres mécanismes d'amplification/délétion comme l'amplification en cerceau ont toutefois été suggérés par certains auteurs (Lagudah *et al.* 1989, Reddy et Appels 1989).

**2- Variations de longueur des unités**

La longueur des unités d'ADNr 5S est très variable chez les plantes supérieures (**Tableau 14**). Dans le cas des angiospermes, elle varie de moins de 200 pb (les unités les plus courtes de plusieurs espèces du genre *Oryza*) à plus de 600 pb (certaines unités de *Nicotiana rustica* et *Populus deltoides*), soit un écart de près de 400 pb. L'amplitude des différences est encore plus grande chez les gymnospermes (plus de 1300 pb) bien que les données disponibles soient comparativement bien moins nombreuses. Les unités les plus courtes ont été observées chez *Picea glauca* (221 pb) et les plus longues chez *Cycas revoluta* (1572 pb). Les variations au sein d'un même genre sont souvent plus faibles. L'écart n'est par exemple que de 5 pb entre *Solanum tuberosum* et *S. pinnatisectum*, 6 pb entre *Angophora floribunda* et *A. hispida*, 15 pb entre 20 espèces de *Gossypium*, 18 pb entre 20 espèces d'*Eucalyptus*, et environ 90 pb entre 4 espèces de *Glycine*. De même, et bien que peu étudiées jusqu'à présent, les différences entre individus d'une même espèce paraissent très limitées ou de trop faible amplitude pour être détectées par les techniques RFLP classiques. Ainsi, sur les 415 provenances de *Triticum tauschii* examinées par Lagudah *et al.* (1989), seules deux ont pu être distinguées sur le critère de la longueur des unités d'ADNr 5S (présence d'une unité de 400 pb en plus des unités de 420 et 490 pb).

**Tableau 14.** Taille des unités d'ADNr 5S chez les plantes supérieures.

| Organisme[a] | Taille (pb)[b] | Référence |
|---|---|---|
| Gymnospermes | | |
| *Cycas revoluta* | 1572 | Gottlob-McHugh *et al.* (1990) |
| *Juniperus virginiana* | 378 | Gottlob-McHugh *et al.* (1990) |
| *Picea glauca* | 221 (C) | Brown et Carlson (1997) |
| | ~ 600 (L) | |
| *Pinus radiata* | **502-538 (C)** | Moran *et al.* (1992) |
| | 851 (L) | |
| *Pseudotsuga menziesii* | 871-888 | Amarasinghe et Carlson (1998) |
| Angiospermes | | |
| *Acer rubrum* | 338 | Gottlob-McHugh *et al.* (1990) |
| *Albizia lebbek* | 477 | Gottlob-McHugh *et al.* (1990) |
| *Allium cepa* | 336 | Gottlob-McHugh *et al.* (1990) |
| *Allosyncarpia ternata* | 334 | Udovicic *et al.* (1995) |
| *Alnus incana* | 479-480 | Johnson *et al.* (1992) |
| *Angophora* (2 sp) | 380-386 | Udovicic *et al.* (1995) |
| *Arabidopsis thaliana* | 497 | Campell *et al.* (1992) |
| *Arillastrum gummiferum* | 310-324 | Udovicic *et al.* (1995) |
| *Beta vulgaris* | 349-351 | Schmidt *et al.* (1994) |
| *Betula papyrifera* | 471-490 | Johnson *et al.* (1992) |
| *Brassicaceae* (12 sp, 6 gr) | ~ 500 | Bhatia *et al.* (1993) |
| *Brassica campestris* | 490-495 | Bhatia *et al.* (1993) |
| *Brassica oleracea* | ~ 250 (C) | Bhatia *et al.* (1993) |
| | ~ 500 (L) | |
| *Bromus inermis* | 427 | Sastri *et al.* (1992) |
| *Casuarina equisetifolia* | 346 | Gottlob-McHugh *et al.* (1990) |
| *Ceratonia siliqua* | 330 | Gottlob-McHugh *et al.* (1990) |
| *Cucurbita sativus* | 341 | Gottlob-McHugh *et al.* (1990) |
| *Cynara scolymus* | 320 (C) | Gottlob-McHugh *et al.* (1990) |
| | 330 (I) | |
| | 480 (L) | |
| *Datisca glomerata* | 219 | Gottlob-McHugh *et al.* (1990) |
| *Dracaena margina* | 486 | Gottlob-McHugh *et al.* (1990) |
| *Elaeagnus angustifolia* | 330 | Gottlob-McHugh *et al.* (1990) |
| *Eucalyptopsis papuana* | 345 | Udovicic *et al.* (1995) |
| *Eucalyptus* (19 sp) | 397-415 | Udovicic *et al.* (1995) |
| *Gleditsia triacanthos* | 278 | Gottlob-McHugh *et al.* (1990) |
| *Glehnia littoralis* | 338 | Mizukami (1995) |
| *Glycine arenaria* | 285 | Doyle et Brown (1989) |
| *Glycine max* | 321-333 | Gottlob-McHugh *et al.* (1990) |
| *Glycine soja* | 345 | Danna *et al.* (1996) |
| *Glycine tomentella* (18 pr 2x) | 260-310 | Doyle et Brown (1989) |
| *Glycine tomentella* (1 pr 2x) | 295 (C) | Doyle et Brown (1989) |
| | 310 (L) | |
| *Glycine tomentella* (10 pr 4x) | 295 | Doyle et Brown (1989) |
| *Glycine tomentella* (20 pr 4x) | 295-310 (C) | Doyle et Brown (1989) |
| | 255-295 (L) | |
| *Gossypium* (20 sp) | 295-314 | Cronn *et al.* (1996) |
| *Gymnocladus dioicus* | 215 | Gottlob-McHugh *et al.* (1990) |
| *Hordeum* (4 sp) | 249-512, **489**, **510** (L-gr4) | Baum et Johnson (1998) |
| | 561-607, **602** (L-X1) | |
| | 233-489, **303** (C-gr1) | |
| | 430-454, **452** (L-gr5) | |
| | 471-476, **475** (L-gr6) | |
| | 472-478, **472** (L-gr7) | |
| | 350-383, **382** (C-gr2) | |
| | 546-508, **487** (C-gr3) | |
| *Kengyilia alatavica* | 307-315, **313**, **314** (C-P1) | Baum et Bailey (1997) |
| | 441-447 (L-S1) | |
| | 380-472 (L-P1) | |
| | 478-521 (L-R1) | |
| | 422-428 (L-P2) | |
| *Linum usitatissimum* | 325-362 | Goldsbrough *et al.* (1982) |
| | | Schneeberger et Cullis (1992) |
| *Lupinus luteus* | 342 | Rafalski *et al.* (1982) |
| *Lycopersicon esculentum* | **355** | Venkateswarlu *et al.* (1991) |
| *Matthiola incana* | 510 | Hemleben et Werts (1988) |

**Tableau 14.** Suite.

| Organisme[a] | Taille (pb)[b] | Référence |
|---|---|---|
| *Myrica gale* | 362 (C)<br>413 (L) | Gottlob-McHugh *et al.* (1990) |
| *Myrica pensylvanica* | 404 | Gottlob-McHugh *et al.* (1990) |
| *Neonotonia wightii* | 520 | Danna *et al.* (1996) |
| *Nicotiana rustica* | **430 (C)**<br>521 (L) | Venkateswarlu *et al.* (1991) |
| *Nicotiana tabacum* | 416-430 (C)<br>617-645 (L) | Gottlob-McHugh *et al.* (1990)<br>Fulnecek *et al.* (1998) |
| *Oryza australiensis* | 180-**302** | McIntyre *et al.* (1992) |
| *Oryza brachyantha* | 100-**350 (C)**<br>380-**490 (L)** | McIntyre *et al.* (1992) |
| *Oryza latifolia* | 160-420, **280** | McIntyre *et al.* (1992) |
| *Oryza minuta* | 150-**194 (C)**<br>**237 (L)** | McIntyre *et al.* (1992) |
| *Oryza nivara* | 250-600, **322** | McIntyre *et al.* (1992) |
| *Oryza officinalis* | 220-550, **326** | McIntyre *et al.* (1992) |
| *Oryza punctata* | 200-600, **326** | McIntyre *et al.* (1992) |
| *Oryza rufipogon* | **322**-410 | McIntyre *et al.* (1992) |
| *Oryza sativa* | 280-320, **302** | McIntyre *et al.* (1992) |
| *Petunia hybrida* | 461 | Frasch *et al.* (1989) |
| *Phaseolus vulgaris* | 389 | Gottlob-McHugh *et al.* (1990) |
| *Pisum sativum* | **279**-329 | Ellis *et al.* (1988) |
| *Polygonatum multiflora* | 486 | Gottlob-McHugh *et al.* (1990) |
| *Populus deltoides* | 543 (C)<br>634 (L) | Gottlob-McHugh *et al.* (1990) |
| *Prosopis alba* | 293 | Gottlob-McHugh *et al.* (1990) |
| *Solanum pinnatisectum* | 329 | Zanke *et al.* (1995) |
| *Solanum tuberosum* | 324 | Zanke *et al.* (1995) |
| *Spinacia vulgaris* | 336 | Gottlob-McHugh *et al.* (1990) |
| *Syringa vulgaris* | 353 | Gottlob-McHugh *et al.* (1990) |
| *Teramnus labialis* | 245 (C)<br>350 (I)<br>600 (L) | Danna *et al.* (1996) |
| *Thapsia garganica* | ~ 300<br>~ 310 | Rasmussen et Avato 1998 |
| *Trifolium repens* | 330 | Gottlob-McHugh *et al.* (1990) |
| *Triticeae* (28 sp, 10 gr) | 327-468 (C)<br>469-508 (L) | Baum et Appels (1992)<br>Sastri *et al.* (1992) |
| *Triticum tauschii* (415 pr) | 400-420<br>490 | Lagudah *et al.* (1989) |
| *Vigna radiata* | 215 | Hemleben et Werts (1988) |
| *Zea mays* | 320-321 | Mascia *et al.* (1981)<br>Sastri *et al.* (1992) |

[a] sp : espèce ; gr : genre ; pr : provenance ; 2x : diploïde ; 4x : tetraploïde

[b] Les différentes classes de taille définies par les auteurs sont indiquées entre parenthèses. C : classe courte ; I : classe intermédiaire ; L : classe longue. Les unités les plus représentées sont indiquées en gras.

De même, très peu de variations de longueur ont été détectées par analyse RFLP dans un groupe de 25 individus majoritairement issus d'une population naturelle de *Pinus radiata* (Moran *et al.* 1992). Les variations sont en revanche très fréquentes au sein d'un génome (**Tableau 14**). Dans le genre *Hordeum*, il existe ainsi selon les espèces 2 à 6 groupes d'unités, de longueur comprise entre 249 pb et 607 pb (Baum et Johnson 1998). Plusieurs classes bien distinctes, elles-mêmes souvent polymorphes, ont par ailleurs été clairement définies chez *Pinus radiata* (502-538 pb et 851 pb), *Nicotiana tabacum* (416-430 pb et 617-645 pb), *Oryza brachyantha* (100-350 pb et 380-490 pb), *Populus deltoides* (543 pb et 634 pb), *Teramnus labialis* (~ 245 pb, ~ 350 pb et ~ 600 pb) et plusieurs espèces de *Triticeae* (327-468 pb et 469-508 pb). Des différences de moindre amplitude ont aussi été détectées dans le génome de nombreuses plantes comme, par exemple, *Pseudotsuga menziesii* (871-

888 pb), *Betula papyrifera* (471-490 pb), *Brassica campestris* (490-495 pb), *Glycine max* (321-333 pb), *Oryza australiensis* (180-302 pb) et d'autres espèces du genre, *Pisum sativum* (279-329 pb), *Linum usitatissimum* (325-362 pb) ou bien encore *Thapsia garganica* (300-310 pb).

Toutes ces variations de longueur reposent principalement sur des insertions/délétions (indels) et des duplications qui sont à la base de l'évolution relativement rapide des unités d'ADNr 5S (Scoles *et al.* 1988, Sastri *et al.* 1992). Ces événements sont très fréquents dans les espaceurs et affectent même parfois les gènes.

a- Variations de longueur dans les gènes

Près de 500 séquences de gènes 5S provenant de plus de 140 espèces de plantes supérieures sont actuellement disponibles dans les banques de données (listées par Szymanski *et al.* 1998). La comparaison de ces séquences avec celles des ARNr 5S matures (voir **Partie I-3** de ce **Chapitre II**) montre que la longueur du gène diffère généralement très peu de 120 pb (117-122 pb). Quelques cas de variations plus importantes ont toutefois été rapportés. Un clone d'*Acacia pycnantha* porte ainsi une délétion d'environ 50 pb (Playford *et al.* 1992). Inversement, des duplications en tandem de grande taille ont été observées chez *Triticum aestivum* (15 pb, positions + 62 à + 76, Gerlach et Dyer 1980, 41 pb, positions + 50 à + 90, Appels *et al.* 1992), *Agropyron cristatum* (38 pb, positions + 52 à + 89, Scoles *et al.* 1988), *Psathyrostachys juncea* (43 pb, positions + 49 à + 91, Scoles *et al.* 1988), *Hordeum bulbosum* (48 pb, positions + 47 à + 94, Scoles *et al.* 1988), *Hordeum spontaneum* (42 pb, positions + 52 à + 93, Baum et Johnson 1996) et *Oryza brachyantha* (29 pb, positions + 58 à + 86 McIntyre *et al.* 1992). Il est remarquable de constater qu'elles sont toutes approximativement centrées sur le motif CGAGAG constituant l'élément M de l'ICR (**Figure 13**). La plus grande (48 pb, positions + 47 à + 94) contient pratiquement l'intégralité du promoteur interne (positions + 50 à + 97) qui se trouve ainsi répété en tandem. Des duplications de 27 pb séparées par un motif de 14 pb ont par ailleurs été observées dans la région de la boîte C (positions + 78 à +104) chez *Triticum aestivum* (Van Campenhout *et al.* 1998). Ces "maxigènes" ne sont donc probablement pas transcrits (pseudogènes) ou produisent des ARNr 5S anormaux (voir la **Partie I-1** de ce **Chapitre II**). La présence dans la région de l'ICR de petites répétitions TGGG aux positions + 63, + 84 et + 96 (**Figure 13**), semble indiquer que ces duplications sont provoquées par des glissements de brins au cours de la réplication (Scoles *et al.* 1988, **Partie II-2-c-α** du **Chapitre I**).

b- Variations de longueur dans les espaceurs

L'espaceur peut être schématiquement divisé en 3 zones plus ou moins affectées par les variations de longueur (Scoles *et al.* 1988, Sastri *et al.* 1992) : la zone centrale très instable, la bordure 5' modérément variable et la bordure 3' très conservée.

α- Les bordures 5' et 3'

L'extrémité 3' (30 à 50 pb) de l'espaceur, située en amont du gène et impliquée dans la régulation de l'initiation de la transcription par la Pol III (**Partie I-1-a-β** de ce **Chapitre II**), semble totalement épargnée par les insertions/délétions et les duplications. Aucune mutation de ce type n'a ainsi pu être mise en évidence dans la famille des *Triticeae* qui constitue la base de donnée la plus importante actuellement disponible (des dizaines d'espèces examinées, Baum et Appels 1992, Kellogg et Appels 1995). L'extrémité 5' en aval du gène (10-30 pb) est un peu plus polymorphe. Chez *Agropyron cristatum* (Scoles *et al.* 1988) ou *Oryza brachyantha* (Mc Intyre *et al.* 1992), elle est par exemple le siège de duplications de 24 pb et 58 pb, respectivement. Le signal de terminaison de la transcription (poly T) présent dans cette région (**Partie I-1-b** de ce **Chapitre II**) peut aussi être sujet à des glissements de brins lors de la réplication. Cronn *et al.* (1996) ont ainsi montré que le poly T est à l'origine de différences mineures de la longueur des espaceurs entre 20 espèces de *Gossypium*. De même, deux clones d'*Arachis hypogaea* se distinguent par des poly T de tailles différentes (Sastri *et al.* 1992) et les variations de longueur des unités d'ADNr 5S observées chez *Glycine max* (321-333 pb) sont entièrement imputables à l'évolution du poly T (Gottlob-McHugh *et al.* 1990).

β- La région centrale

Le polymorphisme observé au niveau des bordures n'est toutefois en rien comparable à celui qui prévaut dans la région centrale des espaceurs. Les insertions/délétions et les duplications y sont extrêmement nombreuses. L'analyse de 38 clones d'ADNr 5S provenant de 15 espèces (11 genres) de *Triticeae* a permis d'identifier 35 délétions apparentes de plus de 15 pb (Scoles *et al.* 1988). L'une d'entre elles, de très grande taille (dans un clone de *Psathyrostachys juncea*), élimine presque totalement cette région centrale ne laissant en place que les bordures 5' et 3'. La plupart sont concentrées au niveau de plusieurs "points chauds" (zones fragiles) et cette distribution hétérogène semble indiquer, là aussi, qu'elles sont générées par des glissements de brins au cours de la réplication (Scoles *et al.* 1988).

L'apparition de différentes classes d'unités ou de variants de longueur au sein du génome d'une plante est souvent la conséquence d'un petit nombre d'événements de ce type (souvent un seul) dans cette région centrale de l'espaceur. Les unités courtes et longues de *Pinus radiata* diffèrent ainsi par l'insertion d'une grande séquence de 330 pb localisée 35 pb en aval du gène (Moran *et al.* 1992). De même, une petite insertion/délétion de 17 pb (79 pb en aval du gène) explique la différence de longueur observée entre les unités de 871 pb et 888 pb identifiées chez *Pseudotsuga mensiezii* (Amarasinghe et Carlson 1998). La situation est similaire chez les angiospermes. Les unités longues de *Nicotiana rustica* (521 pb) sont ainsi caractérisées par la présence d'une séquence supplémentaire de 86 pb (111 pb en aval du gène) par rapport aux unités courtes (Venkateswarlu *et al.* 1991), tandis que chez *Triticum aestivum*, ce sont deux insertions/délétions majeures de 46 pb (76 pb en amont du gène) et 27 pb (155 pb en aval du gène) qui sont la source des variations de longueur entre les deux

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

classes de 416-425 pb et 508 pb, respectivement (Cox *et al.* 1992). Les unités courtes de l'orge (localisées sur le chromosome 2) dérivent des unités longues (localisées sur le chromosome 3 et en partie sur le chromosome 2) par une délétion de 130 pb (située 116 pb en aval du gène) qui a pu être identifiée grâce aux similarités de séquences observées dans cette région avec le seigle et *Triticum tauschii* (Kanazin *et al.* 1993). Chez *Secale cereale*, il a été montré que des délétions de 39 pb (unités courtes, 80 pb en amont du gène) et 9 pb (unités longues, 184 pb en aval du gène) et une insertion de 11 pb (unités courtes, 75 pb en aval du gène) expliquent la différence de longueur entre les deux classes d'unités de 460 pb et 480 pb qui sont respectivement localisées sur les chromosomes 1 et 5 (Reddy et Appels 1989). Les délétions de 39 pb et 9 pb sont sans doute anciennes car elles existent aussi chez d'autres espèces de *Triticeae*. En revanche, l'insertion de 11 pb caractéristique de la classe courte est récente car spécifique de cette espèce. La présence d'une répétition directe de 4 pb (TTTT) en bordure de cette région a conduit les auteurs à conclure dans ce cas à un événement de transposition. Enfin, ce sont des duplications en tandem de 50 pb (55 pb en aval du gène) et 21 pb (102 pb en aval du gène) qui sont à l'origine chez le pois (Ellis *et al.* 1988) et le lin (Goldsbrough *et al.* 1982) de l'existence de différents variants de longueur de 279/329 pb et 341/362 pb, respectivement.

La présence de microsatellites peut aussi être à l'origine de variations de l'espaceur. La taille de ces régions est en effet très instable en raison de probables glissements de brins au cours de la réplication. Chez *Pinus radiata* par exemple, un motif TC répété en tandem dans la région centrale des espaceurs explique 93 % des variations de longueur (502-538 pb) observées entre 4 clones d'unités courtes (Moran *et al.* 1992). De même, dans le genre *Hordeum*, le nombre de copies du motif microsatellite TAG situé dans les unités courtes (60 pb en amont du gène), varie de 4 à 53 selon les espèces, les provenances et les clones examinés (Kanazin *et al.* 1993, Baum et Johnson 1994, 1996).

### 3- Variations de séquence

#### a- Variation intragénomique

Les variations nucléotidiques de l'ADNr 5S sont très fréquentes au sein du génome des plantes supérieures. Les unités séquencées au hasard à partir d'un seul individu sont généralement toutes différentes à une ou plusieurs positions. Cela a été vérifié par exemple chez 31 individus (N = 2-10 unités examinées) représentant 28 espèces (13 genres) de *Triticeae* (Reddy et Appels 1989, Kellogg et Appels 1995) et 21 individus (N = 2-10 unités) correspondant à autant d'espèces du genre *Gossypium* (Cronn *et al.* 1996). La même situation a aussi été observée chez 15 autres plantes de 15 espèces et 12 genres différents, et il semble donc qu'elle soit très répandue. Le **Tableau 15** récapitule le pourcentage de positions nucléotidiques variables ($P_n$) calculé à partir d'un échantillon de 2 à 10 unités d'ADNr 5S provenant de chacune de ces 67 plantes (64 espèces, 26 genres). Estimée à environ 8 % en moyenne, cette mesure du polymorphisme intragénomique varie de moins de 1 % chez *Solanum tuberosum* (0,9 %, N = 4), *Alnus incana* (0,8 %, N = 3), *Pisum sativum* (0,7 %, N = 2) et

*Triticum dichasian* (0,2 %, N = 2) à plus de 15 % dans le génome d'un représentant de *Gossypium raimondii* (15,8 %, N = 6), *Crithopsis delileana* (17,3 %, N = 7), *Pinus radiata* (19,1 %, N = 5), *Henrardia persica* (20,5 %, N = 6), *Gossypium gossypioides* (22,6 %, N = 4) et *Brassica campestris* (25,2 %, N = 4). Elle n'est jamais nulle. Ces valeurs semblent remarquablement élevées au regard du faible nombre d'unités séquencées par génome (2-10 unités seulement sur plusieurs milliers, **Partie III-1** de ce **Chapitre II**) et probablement sous-évaluées. L'estimateur $P_n$ utilisé est en effet sensible à la taille de l'échantillon (Nei 1987). La proportion de positions variables est ainsi en moyenne d'environ 4,1 % pour N = 2 (17 mesures), 6,1 % pour N = 3 (13 mesures), 9,7 % pour N = 4 (16 mesures), 8,4 % pour N = 5 (16 mesures) et 12,4 % pour N = 6-10 (12 mesures).

α- Variation "interloci" VS variation "intralocus"

Lorsque, comme chez *Beta vulgaris* et plusieurs espèces diploïdes de *Gossypium* (**Tableau 12**), l'ADNr 5S est localisé à un seul locus, le polymorphisme intragénomique observé correspond exclusivement à la variation de séquence au sein des séries ou entre les séries situées à ce locus sur les chromosomes homologues. En revanche, lorsque l'ADNr 5S occupe plusieurs loci comme chez la plupart des *Triticeae* (séries d'unités courtes et longues sur les chromosomes 1 et 5, respectivement), il peut être divisé en composantes "interloci" et "intralocus". La variation de séquence entre les loci est généralement équivalente ou supérieure à celle existant au sein d'un locus (Lagudah *et al.* 1989, Reddy et Appels 1989, Appels *et al.* 1992, Cox *et al.* 1992, Kellogg et Appels 1995). Par exemple la proportion de positions variables (insertions/délétions et duplications comptabilisées comme un seul événement) entre les séquences consensus des unités courtes et longues d'un individu ("3", **Tableau 15**) de *Secale cereale* (N = 5 dans les deux cas) qui sont situées à deux loci différents a été estimée à environ 17 % par Reddy et Appels (1989). Ce polymorphisme interloci est équivalent au polymorphisme intralocus des unités courtes (17,3 %). En revanche, il est nettement supérieur au polymorphisme intralocus des unités longues (2,7 %). Les résultats de Reddy et Appels (1989) ont été vérifiés chez un deuxième individu par Kellogg et Appels (1995). Les valeurs obtenues (individus 1 et 3, **Tableau 15**) sont remarquablement similaires (17,3 % pour les séries d'unités courtes, 2,7 % et 3,1 % pour les séries d'unités longues, N = 5 dans chaque cas) ce qui semble indiquer que les niveaux de variation intralocus sont conservés pour différents individus de la même espèce. L'examen du **Tableau 15** montre par ailleurs que les séries d'unités courtes sont plus variables que les séries d'unités longues chez 4 autres individus (3 espèces) du genre *Secale*. Toutefois, des situations inverses sont observées chez *Triticum tauschii* et *Australopyrum velutinum* (Lagudah *et al.* 1989; Kellogg et Appels 1995).

**Tableau 15.** Polymorphisme intragénomique ($P_n$) au sein des unités d'ADNr 5S chez quelques plantes*.

| Espèce[a] | N[b] | Unité | Gène | Espaceur | Référence |
|---|---|---|---|---|---|
| Gymnospermes | | | | | |
| *Pinus radiata* | 5 | 19,1 | 3,4 | 23,4 | Moran *et al.* (1992) |
| *Pseudotsuga menziesii* | 4 | 1,6 | 0 | 1,9 | Amarasinghe et Carlson (1998) |
| Angiospermes | | | | | |
| *Acacia albida* | 4 | 8,8 | 10,3 | 8,4 | Playford *et al.* (1992) |
| *A. melanoxylon* | 7 | 13,6 | 15,9 | 12,8 | |
| *A. ulicifolia* | 4 | 10,9 | 12,1 | 10,5 | |
| *Agropyron cristatum* (L) | 4 | 11,6 | 9,2 | 12,3 | Kellogg et Appels (1995) |
| *Alnus incana* | 3 | 0,8 | 0,8 | 0,8 | Johnson *et al.* (1992) |
| *Australopyrum pectinatum* (C) | 5 | 8,9 | 5,8 | 11,2 | Kellogg et Appels (1995) |
| *A. retrofactum* (L) | 7 | 12,3 | 4,2 | 14,9 | |
| *A. velutinum* (C) | 3 | 4,3 | 5,8 | 3,6 | |
| *A. velutinum* (L) | 4 | 9,9 | 5,8 | 11,2 | |
| *Beta vulgaris* | 2 | 5,1 | 1,7 | 6,9 | Schmidt *et al.* (1994) |
| *Betula papyrifera* | 5 | 9,7 | 3,3 | 12,1 | Johnson *et al.* (1992) |
| *Brassica campestris* | 4 | 25,2 | 1,7 | 32,6 | Bhatia *et al.* (1993) |
| *Critesion bogdanii* (L) | 3 | 2,9 | 4,2 | 2,4 | Kellogg et Appels (1995) |
| *Crithopsis delileana* (C) | 7 | 17,3 | 3,3 | 22,3 | |
| *Dasyprum villosum* (L) | 2 | 4,0 | 0 | 5,4 | |
| *Glycine max* | 9 | 4,5 | 4,2 | 4,7 | Gottlob-McHugh *et al.* (1990) |
| *Gossypium arboreum* (gA) | 5 | 10,7 | 14,1 | 9,6 | Cronn *et al.* (1996) |
| *G. aridum* (gD) | 2 | 8,3 | 6,6 | 11,5 | |
| *G. armourianum* (gD) | 2 | 5,2 | 8,3 | 3,3 | |
| *G. bardadense* (gA) | 7 | 13,1 | 14,9 | 11,9 | |
| *G. davidsonii* (gD) | 2 | 6,3 | 5,8 | 6,7 | |
| *G. gossypioides* (gD) | 4 | 22,6 | 13,2 | 28,9 | |
| *G. harknessii* (gD) | 4 | 8,6 | 10,7 | 7,1 | |
| *G. herbaceum* (gA) | 5 | 4,4 | 3,3 | 5,1 | |
| *G. hirsutum* (gA) | 5 | 15,1 | 15,7 | 14,7 | |
| *G. hirsutum* (gD) | 4 | 12,5 | 14,9 | 10,8 | |
| *G. klotzschianum* (gD) | 4 | 7,0 | 6,6 | 7,2 | |
| *G. laxum* (gD) | 3 | 6,6 | 2,5 | 9,3 | |
| *G. lobatum* (gD) | 5 | 8,9 | 11,6 | 7,1 | |
| *G. mustelinum* (gA) | 5 | 10,4 | 12,4 | 9,6 | |
| *G. mustelinum* (gD) | 5 | 15,8 | 15,7 | 15,9 | |
| *G. raimondii* (gD) | 6 | 15,8 | 7,4 | 21,4 | |
| *G. robinsonii* (gC) | 4 | 9,3 | 12,3 | 7,3 | |
| *G. schwendimanii* (gD) | 4 | 10,2 | 8,3 | 11,5 | |
| *G. thurberi* (gD) | 3 | 11,9 | 9,9 | 13,2 | |
| *G. tomentosum* (gA) | 6 | 13,8 | 15,7 | 12,5 | |
| *G. trilobum* (gD) | 3 | 6,6 | 5,0 | 7,7 | |
| *G. turneri* (gD) | 3 | 4,6 | 5,0 | 4,4 | |
| *G. arboreum x G. thurberi* (gA) | 5 | 10,4 | 12,4 | 9,6 | |
| *G. arboreum x G. thurberi* (gD) | 4 | 6,6 | 7,4 | 6,0 | |
| *Henrardia persica* (C) | 6 | 20,5 | 15,0 | 22,5 | Kellogg et Appels (1995) |
| *Hordeum bulbosum* (L) | 2 | 2,5 | 0 | 3,3 | |
| *Linum usitatissimum* | 5 | 3,5 | 0 | 5,4 | Goldsbrough *et al.* (1982) |
| *Lupinus luteus* | 2 | 4,4 | 6,7 | 3,2 | Rafalski *et al.* (1982) |
| *Pisum sativum* | 2 | 0,7 | 0 | 1,2 | Ellis *et al.* (1988) |
| *Psathyrostachys juncea* (C) | 2 | 4,0 | 3,3 | 4,2 | Kellogg et Appels (1995) |
| *Pseudoroegneria libanotica* (C) | 3 | 13,3 | 4,2 | 16,7 | |
| *Secale cereale* 1 (C) | 5 | 17,3 | 7,5 | 20,6 | |
| *S. cereale* 1 (L) | 5 | 3,1 | 2,5 | 3,3 | |
| *S. cereale* 2 (C) | 2 | 0,9 | 0,8 | 0,9 | |
| *S. cereale* 3 (C) | 5 | 17,3 | 5,8 | 21,3 | Reddy et Appels (1989) |
| *S. cereale* 3 (L) | 5 | 2,7 | 3,3 | 2,5 | |
| *S. montanum* (C) | 2 | 6,0 | 0,8 | 7,8 | Kellogg et Appels (1995) |
| *S. montanum* (L) | 2 | 4,7 | 5,8 | 4,4 | |
| *S. sylvestre* (C) | 3 | 9,4 | 6,7 | 10,3 | |
| *S. sylvestre* (L) | 2 | 3,9 | 0,8 | 4,9 | |
| *S. vavilovii* (C) | 3 | 9,2 | 6,7 | 10,1 | |
| *S. vavilovii* (L) | 3 | 2,7 | 4,2 | 2,2 | |
| *Solanum pinnatisectum* | 5 | 7,3 | 3,3 | 9,6 | Zanke *et al.* (1995) |
| *S. tuberosum* | 4 | 0,9 | 1,7 | 0,5 | |
| *Taeniatherum caput-medusae* (C) | 3 | 6,4 | 3,3 | 7,6 | Kellogg et Appels (1995) |

**Tableau 15.** Suite.

| Espèce[a] | N[b] | Unité | Gène | Espaceur | Référence |
|---|---|---|---|---|---|
| *Thinopyrum bessarabicum* (C) | 2 | 1,3 | 2,5 | 0,9 | Kellog et Appels (1995) |
| *T. junceiforme* (C) | 2 | 1,3 | 1,7 | 1,2 | |
| *Triticum bicorne* (C) | 4 | 5,8 | 7,5 | 5,1 | |
| *T. comosum* (C) | 6 | 2,2 | 0,8 | 2,8 | |
| *T. dichasian* (C) | 3 | 0,2 | 0,8 | 0 | |
| *T. longissimum* (C) | 2 | 6,4 | 12,5 | 3,9 | |
| *T. monococcum* (C) | 5 | 6,7 | 3,3 | 8,7 | |
| *T. sharonense* (C) | 4 | 4,3 | 6,7 | 3,4 | |
| *T. speltoides* 1 (L) | 4 | 8,2 | 2,5 | 10,1 | |
| *T. speltoides* 2 (L) | 2 | 4,3 | 2,5 | 4,9 | |
| *T. tauschii* (C) | 5 | 5,4 | 3,3 | 6,2 | |
| *T. tauschii* (L) | 4 | 10,7 | 17,5 | 8,6 | |
| *T. tripsacoides* (C) | 5 | 3,6 | 0 | 5,1 | |
| *T. umbellulatum* (L) | 5 | 5,4 | 7,5 | 4,8 | |
| **Moyenne** | 4,0 | 8,1 | 6,3 | 8,8 | |

* Exprimé en pourcentage de positions nucléotidiques variables dans l'alignement. Les insertions/délétions ou duplications sont comptabilisées comme un seul évènement.
[a] C : unités courtes ; L : unités longues ; 1, 2, 3 : désignation des individus ; gA, gC, gD : génome diploïde de type A, C ou D ; gAD : génome allopolyploïde.
[b] Nombre de clones analysés. Les gènes et les espaceurs proviennent du même clone.

β- Variation dans les gènes VS variation dans les espaceurs

Le polymorphisme de séquence intragénomique ($P_n$) ne semble pas très différent en moyenne entre les espaceurs (8,8 %) et les gènes (6,3 %) des plantes examinées (**Tableau 15**). Le polymorphisme des espaceurs est généralement plus élevé que celui des gènes mais c'est la situation inverse qui prévaut dans plus de 37 % des cas analysés. Les calculs de diversité nucléotidique[7] ($\pi$) effectués chez les *Triticeae* (Kellogg et Appels 1995) et dans le genre *Gossypium* (Cronn *et al.* 1996) ont montré que les différences observées entre gène et espaceur au sein d'un même individu sont rarement significatives. Seules des différences calculées globalement sur un ensemble d'individus ont pu être mises en évidence. Concernant la série d'unités courtes des *Triticeae* (23 plantes examinées), la diversité nucléotidique est ainsi en moyenne plus importante (test t, $p < 0{,}04$) dans les espaceurs ($\pi = 0{,}040$) que dans les gènes ($\pi = 0{,}023$). La différence n'est cependant plus significative si les délétions/insertions et duplications sont exclues de l'analyse ($\pi = 0{,}028$ pour les espaceurs, $\pi = 0{,}019$ pour les gènes). Elle ne l'est pas non plus dans le cas de la série d'unités longues ($\pi = 0{,}042$ pour les espaceurs, $\pi = 0{,}028$ pour les gènes ; 13 plantes examinées). Au sein du genre *Gossypium*, des situations inverses ou similaires ont été décrites. Dans les génomes de type A (7 plantes concernées, 2 diploïdes et 5 polyploïdes), la diversité nucléotidique des gènes ($\pi = 0{,}057$) est ainsi supérieure (test t, $p < 0{,}02$) à celle des espaceurs ($\pi = 0{,}043$), tandis que dans les génomes de type D, la différence est

[7]Mesure du polymorphisme dans un groupe de séquences numériquement équivalente au nombre moyen de différences nucléotidiques par site entre deux séquences. Cet estimateur est nettement moins biaisé que $P_n$ par rapport à la taille de l'échantillon (Nei 1987).

très faible ($\pi$ = 0,064 pour les gènes, $\pi$ = 0,067 pour les espaceurs). Il est donc difficile de dégager une règle générale quant à la distribution du polymorphisme nucléotidique intragénomique au sein des unités et cette situation suggère que les mutations s'accumulent au hasard dans les gènes et les espaceurs des séries d'ADNr 5S (Kellogg et Appels 1995, Cronn *et al.* 1996).

$\chi$- Pseudogènes

L'existence d'un important polymorphisme de séquence dans les gènes est surprenante en raison du rôle important des ARNr 5S dans la synthèse protéique. La question a été bien étudiée chez les *Triticeae* par Kellogg et Appels (1995). Plus de 70 % des clones étudiés (107/152) par ces auteurs diffèrent de la séquence consensus (73 positions variables sur 120). L'analyse de la distribution des variations montre que les boîtes A et/ou C de l'ICR impliquées dans l'initiation de la transcription sont modifiées dans plus de 17 % des gènes qui seraient donc non fonctionnels (pseudogènes). Kellogg et Appels (1995) indiquent de plus que ce chiffre est probablement sous-estimé car les nombreuses mutations localisées à l'extérieur de l'ICR peuvent altérer l'intégrité des structures secondaires et tertiaires des ARNr 5S. Schneeberger et Cullis (1992) ont d'ailleurs aussi été amenés à conclure à la présence de nombreux pseudogènes dans les séries d'ADNr 5S du lin (*Linum usitatissimum*) après avoir observé de multiples mésappariements dans les régions bicaténaires (I-V, **Figure 17**) des ARNr 5S déduits de la séquence des gènes de plusieurs groupes d'unités (3-5). Enfin, les mutations sont susceptibles d'affecter le pool des transcrits (Kellogg et Appels 1995) en perturbant les sites d'interaction entre les ARNr 5S et le facteur de transcription TFIIIA (consulter la **Figure 17** et la **Partie I-2** de ce **Chapitre II**). Les pseudogènes sont donc probablement nombreux dans les séries de l'ADNr 5S. Toutefois, aucune donnée de séquence n'est actuellement disponible pour préciser si les ARNr 5S sont aussi polymorphes que les gènes ou s'ils dérivent d'une fraction réduite des membres de la famille.

b- Variation intraspécifique (entre les individus, populations, *etc.*)

Curieusement, la plupart des auteurs ne font pas la distinction entre les variations intragénomique et intraspécifique de l'ADNr 5S (Appels *et al.* 1992, Playford *et al.* 1992, Kellogg et Appels 1995, Udovicic *et al.* 1995, Cronn *et al.* 1996). Les informations relatives aux variations de séquence entre individus d'une même population ou entre les populations d'une même espèce sont donc peu nombreuses et souvent très limitées. Dans la vaste étude des variations intraspécifiques et interspécifiques de l'ADNr 5S chez les *Triticeae* de Kellogg et Appels (1995) par exemple, on ne trouve des données que pour 4 espèces (*Secale cereale*, *Triticum monococcum*, *T. tripsacoides*, *T. speltoides*) chacune représentée par seulement deux individus pour lesquels 1 à 5 séquences ont été établies. De même, pour évaluer le polymorphisme intraspécifique au sein de 2 espèces d'*Eucalyptus* (*E. obliqua* et *E. calophylla*), Udovicic *et al.* (1995) n'ont séquencé pour la première que 4 unités issues de 4 individus et, pour la seconde, seulement deux unités issues de deux individus. Dans ces trois cas, les différences observées entre les quelques individus examinés se sont avérées très limitées

(quelques mutations ponctuelles) et les auteurs en ont alors conclu qu'un individu pouvait être considéré comme représentatif de l'espèce.

A notre connaissance, l'évaluation de la diversité moléculaire intraspécifique des unités d'ADNr 5S n'a été sérieusement entreprise que pour 4 espèces du genre *Hordeum* (Baum et Johnson 1994, 1996, 1998) et une espèce apparentée, *Kengyilia alatavica* (Baum et Bailey 1997). Cinq groupes de séquences (C-gr1, C-gr2, C-gr3, L-gr4, L-gr5, **Tableau 14**) ont ainsi été caractérisés chez *H. vulgare* (10 provenances, 26 clones, Baum et Johnson 1994), 5 groupes (C-gr1, C-gr2, L-gr4, L-gr6, L-gr7) chez *H. bulbosum* (11 provenances, 15 clones, Baum et Johnson 1996), 5 groupes (C-P1, L-S1, L-P1, L-R1, L-P2) chez *K. alatavica* (3 provenances, 30 clones, Baum et Bailey 1997), 6 groupes (C-gr1, C-gr2, C-gr3, L-gr4, L-gr5, L-gr6) chez *H. spontaneum* (8 provenances, 15 clones, Baum et Johnson 1996) et enfin, seulement 2 groupes (L-gr4 et L-X1) chez *H. marinum* (11 provenances, 42 clones, Baum et Johnson 1998). La variation de séquence au sein de chaque groupe est faible, même entre les provenances. En revanche, le nombre et/ou le type de séquence détectés semblent variables d'une provenance à l'autre (1-3 groupes représentés par provenance). La taille de l'échantillon est cependant encore trop limitée (1-10 clones examinés par provenance) pour conclure avec certitude à d'éventuelles différences entre provenances au niveau de l'ADNr 5S.

c- Variations interspécifiques

Les fortes variations de séquence des espaceurs observées au sein d'un individu se traduisent, comme on peut logiquement s'y attendre, par un niveau de divergence interspécifique élevé. Ce n'est pas le cas en revanche pour les gènes. Malgré un niveau de polymorphisme intragénomique comparable à celui des espaceurs (**Partie III-3-a-β** de ce **Chapitre II**), la séquence des gènes est extrêmement conservée entre les espèces. Le polymorphisme intragénomique et la divergence interspécifique sont donc corrélés pour les espaceurs et découplés pour les gènes (Kellogg et Appels 1995, Cronn *et al.* 1996).

α- Variations du gène

Comme dans les ARNr 5S correspondants (**Partie I-3** de ce **Chapitre II**), il y a très peu de variations interspécifiques au sein de la région transcrite de l'ADNr 5S chez les plantes supérieures (90-100 % d'homologie, Szymanski *et al.* 1998). Elles ont néanmoins été exploitées pour étudier l'évolution des grands groupes de plantes (Hori *et al.* 1985, Devereux *et al.* 1990, Van de Peer *et al.* 1990) mais les résultats obtenus sont très discutés en raison de l'insuffisance du polymorphisme (lié en particulier à la faible taille du gène), de l'existence de substitutions compensatrices (les positions nucléotidiques impliquées dans les régions bicaténaires I-V des ARNr 5S ne varient pas indépendamment), et de mutations répétées (homoplasie) à certains sites spécifiques (Steele *et al.* 1991, Halanich 1991). Sur les 60 positions variables entre 26 ARNr 5S de plantes supérieures (6 algues vertes, 4 bryophytes, 4 ptéridophytes, 3 gymnospermes et 9 angiospermes) analysés par Steele

*et al.* (1991), seules 36 sont réellement indépendantes et informatives[8]). De même, malgré le grand nombre de sites polymorphes repérés dans les gènes des unités courtes (74) et longues (67) des 34 *Triticeae* étudiées par Kellogg et Appels (1995), il n'y a au total que 3 différences nucléotidiques fixées pour une espèce dans les unités courtes [C en position + 6 chez *Triticum longissimum*, A (+ 8) chez *Psathyrostachys juncea*, G (+ 75) chez *Henrardia persica*], et seulement 4 dans les unités longues [C (+ 29) chez *Triticum umbellulatum*, T (+ 29) chez *Australopyrum velutinum*, G (+ 35) chez *Secale sylvestre*, A (+ 52) chez *Secale vavilovii*, T (+ 103) chez *Hordeum bulbosum*]. Enfin, aucune des 94 positions variables observées dans le genre *Gossypium* (20 espèces) par Cronn *et al.* (1996) ne s'est avérée informative. Seuls quelques polymorphismes (indels et substitutions) communs à plusieurs espèces de ce genre (et donc conservés au cours de l'évolution) ont pu être détectés. Néanmoins, les séquences consensus établies à partir des 99 clones de *Gossypium* ou des 152 clones de *Triticeae* sont identiques, ce qui indique (Cronn *et al.* 1996) que le gène n'a donc pas significativement évolué depuis la séparation des *Malvaceae* et des *Poaceae*, il y a plus de 120 millions d'années (Ma).

β- Variations de l'espaceur

Contrairement aux gènes, la divergence interspécifique des espaceurs est généralement importante. Les bordures de l'espaceur présentent cependant un polymorphisme de séquence plutôt modéré en raison de leur rôle probable dans la régulation de la transcription (Scoles *et al.* 1988). En revanche, la région centrale de l'espaceur qui ne contient normalement pas de séquences fonctionnelles[9], est souvent perturbée par des insertions/délétions et des duplications (**Partie III-2-b** de ce **Chapitre II**). Gottlob-McHugh *et al.* (1990) n'ont ainsi obtenu aucun signal positif en hybridant une sonde constituée d'un espaceur de *Glycine max* sur des digestions de l'ADN génomique de plantes appartenant à 4 classes et 30 familles de gymnospermes et d'angiospermes monocotylédones ou dicotylédones. Les espaceurs sont toutefois suffisamment conservés pour permettre des comparaisons par alignement de séquences ou par les techniques RFLP au sein des familles (*Triticeae*, Scoles *et al.* 1988, Baum et Appels 1992, Kellogg et Appels 1995, *Myrtaceae*, Udovicic *et al.* 1995, *Solanaceae*, Zanke *et al.* 1995) ou des genres (*Acacia*, Playford *et al.* 1992, *Pinus*, Moran *et al.* 1992, *Oryza*, McIntyre *et al.* 1992, *Triticum*, Appels *et al.* 1992, *Gossypium*, Cronn *et al.* 1996, *Hordeum*, Baum et Johnson 1998).

---

[8] Dans un alignement de séquences de divers taxa, une position informative est un site qui comporte au moins deux types de nucléotides, chacun représenté dans au moins deux taxa.

[9] La seule exception connue concerne la présence d'un gène d'ARNt dans les espaceurs des unités d'ADNr 5S du riz (Hariharan *et al.* 1987). Toutefois, ces gènes ont une structure secondaire anormale et ne semblent pas transcrits (pseudogènes).

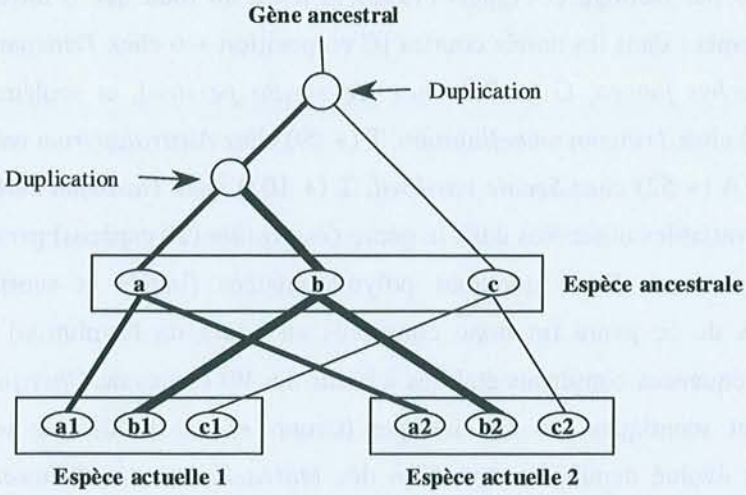



**Figure 19.** Relations alléliques hypothétiques au sein d'une famille multigénique. Adapté de Avise (1994). Les duplications successives d'un gène ancestral produisent 3 gènes contemporains a, b et c éventuellement localisés à 3 loci différents d'une espèce ancestrale. A la suite d'un évènement de spéciation, les espèces actuelles 1 et 2 reçoivent respectivement les allèles a1, b1, c1 et a2, b2, c2. Les comparaisons génétiques a1/a2, b1/b2 et c1/c2 sont dites orthologues. Les autres sont dites paralogues (par exemple, a1/b2, a1/c2 ou a1/b1). En l'absence d'évolution concertée, la similarité de séquence de deux gènes orthologues est corrélée au temps de divergence depuis la spéciation tandis que celle de deux gènes paralogues remonte à l'évènement de duplication qui peut précéder (comme indiqué sur le schéma), coincider ou suivre l'évènement de spéciation. Cependant, l'homogénéisation de tout ou partie des membres d'une famille par les mécanismes de l'évolution concertée peut réduire la pertinence des comparaisons orthologues et paralogues. A l'extrême, elle peut théoriquement conduire à la perte de l'orthologie de séquence entre espèces, par exemple dans le cas du transfert horizontal de l'allèle a1 à tous les membres de la famille (fixation) dans l'espèce 1 et de l'allèle b2 dans l'espèce 2.

Les similarités de séquence sont par exemple de plus de 70 % entre 5 espèces de *Solanaceae* (Zanke *et al.* 1995) et, supérieure à 67 % (unités courtes, 21 espèces, 9 genres) ou 69 % (unités longues, 13 espèces, 7 genres) dans la famille des *Triticeae* (Kellogg et Appels 1995). Au sein des genres, elles dépassent 70 % entre 20 espèces de *Gossypium* (Cronn *et al.* 1996), 70 % (unités courtes) ou 75 % (unités longues) entre 11 espèces de *Triticum* (Appels *et al.* 1992), 53 % entre 9 espèces d'*Oryza* (McIntyre *et al.* 1992) et 33 % entre 7 espèces d'*Acacia* (Playford *et al.* 1992). Les variations de l'espaceur se sont ainsi révélées particulièrement adaptées à l'étude des relations phylogénétiques dans la plupart de ces groupes. Toutefois, la méconnaissance des relations de paralogie et d'orthologie (**Figure 19**) entre les unités d'ADNr 5S limite fortement l'interprétation des résultats dans certains cas. Par exemple, l'analyse de l'importante variation de séquence observée dans les genres *Acacia* (Playford *et al.* 1992) et dans une moindre mesure *Oryza* (McIntyre *et al.* 1992) conduit à des arbres phylogénétiques confus, les séquences d'une même espèce n'apparaissant pas toujours groupées. Cette situation semble liée à la présence dans ces genres de plusieurs lignées encore mal identifiées de séquences paralogues très divergentes issues d'événements de duplication antérieurs à la différenciation des espèces.

La caractérisation des différentes classes de séquences paralogues présentes dans un génome (analyse de la diversité moléculaire intraspécifique) est un préalable indispensable à l'étude des variations interspécifiques. Elle permet d'établir l'orthologie des classes entre les espèces et d'évaluer l'impact de l'évolution concertée sur ces relations (**Figure 19**). Seules les comparaisons de séquences orthologues (c'est-à-dire qui évoluent parallèlement depuis un événement de spéciation) sont en effet susceptibles de fournir des informations phylogénétiques fiables (Sanderson et Doyle 1992, Buckler *et al.* 1997). La comparaison de séquences paralogues conduit généralement à des conclusions erronées car les duplications coïncident rarement avec les événements de spéciation. L'importante base de données disponible chez les *Triticeae* a ainsi permis d'identifier deux grands groupes de séquences paralogues localisés sur les chromosomes 1 (généralement des unités courtes) et/ou 5 (généralement des unités longues) de la plupart des espèces (Scoles *et al.* 1988, Dvorák *et al.* 1989, Lagudah *et al.* 1989, Reddy et Appels 1989, Cox *et al.* 1992, Sastri *et al.* 1992). Les seules exceptions connues concernent le genre *Hordeum* et probablement aussi le genre apparenté *Critesion*. Chez *H. vulgare*, les unités courtes occupent ainsi au moins deux loci distincts sur les chromosomes 2 et 3 (Kanazin *et al.* 1993, Baum et Johnson 1994) tandis que la localisation des unités longues n'est pas encore précisément connue (il existe d'autres loci sur les chromosomes 1, 4, 6 et 7, **Tableau 12**). Les relations phylogénétiques établies à partir des variations de séquences *a priori* orthologues des unités courtes ou des unités longues (Appels *et al.* 1992, Baum et Appels 1992, Kellogg et Appels 1995) confortent globalement la classification usuelle des *Triticeae* (les genres sont bien séparés) mais ne sont pas totalement concordantes au niveau des espèces (les séquences d'une espèce au sein d'un genre ne sont pas toujours groupées). Ces discordances sont parfois révélatrices d'histoires évolutives complexes (les hybridations introgressives sont par exemple évidentes dans le genre *Secale*) mais elles restent le plus souvent inexpliquées, probablement en raison de problèmes de paralogie au sein des classes d'espaceurs. Baum et Johnson (1998) ont ainsi récemment identifié plusieurs groupes de séquences paralogues aussi bien dans la classe d'unités courtes (C-gr1, C-gr2 et C-gr3) que dans la classe d'unités longues (L-X1, L-gr4, L-gr5, L-gr6 et L-gr7) de 4 espèces du genre *Hordeum* (consulter la **Partie III-3-b** de ce **Chapitre II** et le **Tableau 14**). Une situation similaire existerait aussi dans le genre *Critesion* (Kellogg et Appels 1995). La similitude de taille des unités n'est donc pas un critère suffisant pour établir l'orthologie entre les espèces. Certaines unités courtes peuvent ainsi se révéler plus apparentées en séquence aux unités longues et vice versa du fait des insertions/délétions et duplications. Dans l'état actuel des connaissances, seul le groupe d'unités longues L-gr4 fournit l'orthologie nécessaire pour l'étude des relations phylogénétiques entre les espèces d'*Hordeum* et les résultats obtenus sont en accord avec les données basées sur la morphologie et la cytologie (Baum et Johnson 1998).

L'existence de plusieurs classes d'espaceurs paralogues a aussi été mise en évidence par des techniques RFLP dans le genre *Pinus* (Moran *et al.* 1992). Les espèces américaines du sous-genre *Pinus* se caractérisent par la présence de deux classes polymorphes d'unités, courtes et longues, qui diffèrent essentiellement par une insertion de 330 pb dans l'espaceur des unités longues. En revanche,

les espèces eurasiatiques du même sous-genre ne possèdent que des unités longues. Dans le sous-genre *Strobus*, trois classes d'unités (courtes, intermédiaires et longues) ont été détectées mais aucune d'entre elles n'est cependant orthologues de la classe d'unités longues du sous-genre *Pinus* (une sonde constituée de l'insertion de 330 pb ne révèle aucun signal).

L'orthologie (et donc les relations phylogénétiques) semble plus facile à établir dans d'autres groupes de plantes. Les 3 classes d'espaceurs détectées dans le genre *Gossypium* (Cronn *et al.* 1996) coïncident ainsi avec les 3 groupes génomiques connus (A, C et D). Les espèces diploïdes (génomes AA, CC et DD) sont chacune caractérisées par une classe tandis que les génomes des espèces allopolyploïdes (AD) contiennent deux classes paralogues. Les comparaisons orthologues au sein de la classe génomique D (15 espèces) sont très résolutives et confortent les études antérieures. En revanche, les espaceurs de la classe génomique A (7 espèces comparées) ne permettent pas d'établir les relations entre les espèces en raison de leur évolution beaucoup plus lente. Une situation similaire a été rencontrée dans le genre *Eucalyptus* qui apparaît paraphylétique sur la base des variations de l'unique classe d'espaceurs (277-295 pb, Udovivic *et al.* 1995). La résolution est en effet très bonne dans le groupe des eucalyptus "bloodwood" (sous-genres *Blakella* et *Corymbia*) mais le nombre de caractères informatifs est nettement insuffisant dans le cas du groupe des "non bloodwood" (les autres sous-genres). Les espaceurs évoluent donc à des vitesses très différentes selon les groupes.

### 4- Evolution concertée et sélection

#### a- Interactions entre les loci

Aucun échange entre séries d'ADNr 5S situées à différents loci sur des chromosomes non homologues n'a pu être mis en évidence jusqu'à présent chez les plantes supérieures. Les séries d'unités courtes (chromosomes 1) et longues (chromosomes 5) de la plupart des *Triticeae* évoluent apparemment de manière totalement indépendante (Scoles *et al.* 1988, Reddy et Appels 1989, Appels *et al.* 1992, Sastri *et al.* 1992, Kellogg et Appels 1995, voir aussi la **Partie III-3-a** de ce **Chapitre II**). Par exemple, la petite insertion de 11 pb caractéristique des unités courtes de *Secale cereale* n'a jamais été observée dans les unités longues (Reddy et Appels 1989). Même entre les séries des chomosomes homéologues des polyploïdes, les échanges paraissent très limités. Van Campenhout *et al.* (1998) ont en effet récemment montré que les unités courtes situées sur les chromosomes 1B ou 1D de *Triticum aestivum* ont des séquences tout à fait spécifiques. De même, les unités d'ADNr 5S des génomes A et D de plusieurs espèces allopolyploïdes du genre *Gossypium* apparues depuis 1-2 Ma ont conservé leur homologie de séquence avec les unités des espèces diploïdes parentales qui ont divergé d'un ancêtre commun il y a 6-11 Ma (Cronn *et al.* 1996). Les phénomènes d'homogénéisation entre les deux loci d'ADNr 5S des génomes A et D sont donc inexistants ou trop peu nombreux pour être détectés. Comme les échanges interchromosomiques entre les loci de l'ADNr situés au niveau des télomères sont au contraire très nombreux (Wendel *et al.* 1995), Cronn *et al.* (1996) concluent que l'apparente absence d'évolution concertée entre les loci d'ADNr 5S (et donc la stabilité de ces loci) est

peut-être liée à leur localisation proche du centromère (**Tableau 12**). Si l'existence de crossing-over interchromosomiques est en effet envisageable dans les régions télomériques (**Partie II-3-c-α** du **Chapitre I**), ils sont *a priori* très délétères au niveau des centromères ou des régions intercalaires (gamètes non viables). L'importante hétérogénéité de séquence des unités d'ADNr 5S observée chez le lin (Schneeberger *et al.* 1989, Schneeberger et Cullis 1992) pourrait aussi être interprétée comme résultant de l'absence ou de la faible fréquence des mécanismes d'homogénéisation entre les nombreux loci dispersés dans ce génome.

b- Homogénéisation intralocus

L'évolution concertée des séries de l'ADNr 5S est en revanche tout à fait évidente au sein d'un locus (Scoles *et al.* 1988, Dvorák *et al.* 1989, Sastri *et al.* 1992, Kellogg et Appels 1995, Cronn *et al.* 1996). Dans le genre *Gossypium* par exemple (Cronn *et al.* 1996), la longueur des unités est très conservée au sein de chaque classe correspondant à l'unique locus des génomes des groupes A (295-298 pb), C (310-314 pb) et D (301-304 pb). De plus, dans un échantillon de 21 espèces examinées, la variation intraspécifique de séquence est inférieure à la divergence interspécifique dans plus de 93 % des comparaisons d'unités (259/276) impliquant les génomes A, C et/ou D. Dans le cas des *Triticeae* (Kellogg et Appels 1995), les chiffres sont de 93 % (196/210 comparaisons) pour le locus des unités courtes (21 espèces) et 89 % (70/78 comparaisons) pour le locus des unités longues (13 espèces). L'important polymorphisme intragénomique observé chez la plupart des plantes (**Tableau 15**) montre toutefois que l'intensité de l'évolution concertée n'est pas suffisante pour homogénéiser complètement les séries d'ADNr 5S. Si le glissement de brins, le crossing-over inégal ou d'autres mécanismes d'amplification ou de contraction des séries interviennent réellement, leur effet est plutôt local (ils impliquent peu d'unités à la fois) et/ou leur fréquence faible. L'existence de quelques polymorphismes communs (aussi bien dans les gènes que dans les espaceurs) entre des espèces proches au sein des genres (*Triticeae*, *Gossypium*) indique que les séries d'unités d'ADNr 5S n'ont pas été complètement homogénéisées par les mécanismes de l'évolution concertée. Ces mêmes études montrent par ailleurs que les unités de la plupart de ces espèces ont apparemment un ancêtre commun plus récent que l'événement de spéciation (les unités d'une espèce sont généralement plus apparentées entre elles qu'aux unités d'autres espèces). L'homogénéisation des séries se réalise donc approximativement à la même vitesse ou un peu plus rapidement que la spéciation dans ces deux groupes (Kellogg et Appels 1995, Cronn *et al.* 1996). C'est un processus très lent.

c- Sélection naturelle

Le polymorphisme intragénomique des gènes est souvent aussi élevé ou même parfois plus élevé que le polymorphisme des espaceurs (**Partie III-3-a-β** de ce **Chapitre II, Tableau 15**). Or, entre les espèces, les gènes sont remarquablement conservés tandis que les espaceurs sont très divergents (**Partie III-3-c** de ce **Chapitre II**). Des mécanismes sélectifs s'exercent donc sur les gènes sans affecter (ou très peu) les espaceurs. Le glissement de brins, les crossing-over inégaux et autres

mécanismes moléculaires plus globaux d'homogénéisation ne peuvent rendre compte de cette différence car, théoriquement, ils agissent indifféremment à la fois sur les gènes et les espaceurs [mais consulter les travaux récents de Polanco *et al.* (1998), voir aussi la **Partie II-3-c-χ** du **Chapitre I**]. Le seul mécanisme concevable dans l'état actuel des connaissances pourrait être la conversion génique biaisée (**Partie II-3-c-β** du **Chapitre I**). Cependant, Kellogg et Appels (1995) ainsi que Cronn *et al.* (1996) invoquent plutôt un modèle de sélection naturelle (proposé par Schlötterer et Tautz 1994 pour l'ADNr de drosophile) opérant non pas individuellement sur chaque gène d'une série (le polymorphisme intragénomique des gènes serait alors plus faible que celui des espaceurs) mais sur la série dans son ensemble (**Figure 20**). Les mutations s'accumuleraient au hasard dans les gènes et les espaceurs jusqu'à un certain seuil de tolérance de la série vis à vis des pseudogènes (effet tampon déterminé par le nombre d'unités) au-delà duquel la sélection naturelle peut provoquer l'élimination de la série dans une population. Dans ce modèle, les événements de contraction des séries par les mécanismes de l'évolution concertée pourraient abaisser fortement le seuil critique.

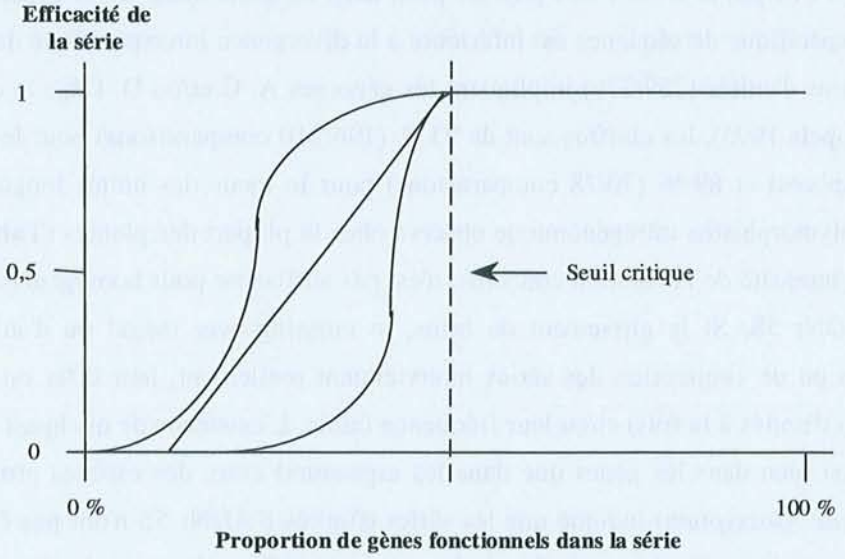



**Figure 20.** Représentation schématique de l'efficacité d'une série en fonction de la proportion de gènes fonctionnels qu'elle contient. D'après Cronn *et al.* (1996). Plusieurs possibilités sont illustrées bien que ni l'allure réelle de la courbe ni la valeur du seuil critique à partir duquel l'efficacité diminue ne soient connus. Quand le nombre de gènes fonctionnels n'est plus suffisant pour satisfaire les besoins de la cellule, la série est éliminée de la population par la sélection naturelle.

## IV- Conclusion

Les séries parfaites de l'ADNr 5S occupent généralement un faible nombre de sites chromosomiques localisés près des centromères ou des télomères chez les plantes supérieures. Le nombre d'unités évolue rapidement entre les espèces et au sein des espèces mais il semble stable au cours du développement d'un individu et d'une génération à l'autre. La méthylation affecte la plupart des résidus cytosine sur les deux brins de chaque unité.

Les variations de longueur des unités sont particulièrement fréquentes au sein des génomes (présence de différents variants ou de différentes classes) et entre les espèces. Elles sont essentiellement provoquées par des insertions et/ou délétions localisées dans la zone centrale des espaceurs, même si les gènes et les régions flanquantes (bordures des espaceurs), beaucoup plus stables, sont parfois la cible de duplications engendrées par des phénomènes de glissements de brins. La grande conservation de la taille des gènes est non seulement liée au rôle majeur des ARNr 5S, en particulier dans la biogenèse des ribosomes, mais aussi à la présence d'un promoteur intragénique. Les bordures des espaceurs sont quant à elles probablement impliquées dans l'initiation et la terminaison de la transcription.

Comme les variations de longueur, le polymorphisme de séquence intragénomique est important et se manifeste aussi bien entre les loci qu'au sein d'un locus. Les gènes sont apparemment aussi variables que les espaceurs et un grand nombre d'entre eux ont des caractéristiques de séquence typiques des pseudogènes. La sélection qui s'exerce sur chaque gène d'une série est donc probablement très faible. Les informations disponibles concernant les variations de séquence intraspécifiques (entre les individus, les populations, *etc.*) sont encore très limitées. En revanche, les variations interspécifiques sont très étudiées. Les gènes sont remarquablement conservés entre les espèces, même très éloignées, et donc leur valeur phylogénétique est très réduite. Les espaceurs sont au contraire très divergents. Ils fournissent de nombreux caractères informatifs pour l'analyse de l'évolution moléculaire au sein des familles et des genres pour peu que l'orthologie puisse être formellement établie, ce qui requiert généralement une analyse préalable de la diversité moléculaire intraspécifique et/ou intragénomique.

Les unités d'ADNr 5S situées à différents loci n'évoluent apparemment pas de manière concertée. En revanche, les échanges sont évidents au sein d'un locus même si leur fréquence n'est semble-t-il pas suffisante pour homogénéiser complètement les séries. L'évolution concertée se produirait à une vitesse approximativement équivalente aux événements de spéciation et serait accompagnée par des phénomènes de sélection naturelle s'exerçant sur l'ensemble d'une série (plutôt que sur chaque gène individuellement).

## Bibliographie

**Amarasinghe V, Carlson JE (1998).** Physical mapping and characterization of 5S rRNA genes in Douglas-fir. J Heredity 89 : 495-500

**Ansari HA, Ellison NW, Reader SM, Badaeva ED, Friebe B, Miller TE, Williams WM (1999).** Molecular cytogenetic organization of 5S and 18S-26S rDNA loci in white clover (*Trifolium repens* L.) and related species. Ann Bot 83 : 199-206

**Appels R, Baum BR, Clarke BC (1992).** The 5S DNA units of bread wheat (*Triticum aestivum*). Plant Syst Evol 183 : 183-194

**Appels R, Gerlach WL, Dennis ES, Swift H, Peacock WJ (1980).** Molecular and chromosomal organisation of DNA sequences coding for the ribosomal RNAs in cereals. Chromosoma 78 : 293-311

**Avise JC (1994).** *Molecular markers, natural history and evolution*. Chapman & Hall, New York

**Badaeva ED, Friebe B, Gill BS (1996).** Genome differentiation in *Aegilops*. 2. Physical mapping of 5S and 18S-26S ribosomal RNA gene families in diploid species. Genome 39 : 1150-1158

**Barciszewska MZ, Erdmann VA, Barciszewska J (1994).** A new model for the tertiary structure of 5S ribonucleic acid in plants. Plant Mol Biol Rep 12 : 116-131

**Baum BR, Appels R (1992).** Evolutionary change at the 5S DNA loci of species in the *Triticeae*. Plant Syst Evol 183 : 195-208

**Baum BR, Bailey LG (1997).** The molecular diversity of the 5S rRNA gene in *Kengyilia alatavica* (Drobov) JL Yang Yen & Baum (*Poaceae: Triticeae*): Potential genomic assignment of different rDNA units. Genome 40 : 215-228

**Baum BR, Johnson DA (1994).** The molecular diversity of the 5S rRNA gene in barley (Hordeum vulgare). Genome 37 : 992-998

**Baum BR, Johnson DA (1996).** The 5S rRNA gene units in ancestral two-rowed barley (*Hordeum spontaneum* C. Koch) and bulbous barley (*H. bulbosum* L.): sequence analysis and phylogenetic relationships with the 5S rDNA units of cultivated barley (*H. vulgare* L.). Genome 39 : 140-149

**Baum BR, Johnson DA (1998).** The 5S rRNA gene in sea barley (*Hordeum marinum Hudson sensu lato*): sequence variation among repeat units and relationship to the X haplome in barley (*Hordeum*). Genome 41 : 652-661

**Bezdek M, Koukalová B, Kuhrovà V, Vyskot B (1992).** Differential sensitivity of CG and CCG DNA sequences to methionine-induced hypomethylation of the *Nicotiana tabacum* genome. FEBS Letters 300 : 268-270

**Bhatia S, Singh K, Jagannathan V, Lakshmikumaran M (1993).** Organization and sequence analysis of the 5S rRNA genes in *Brassica campestris*. Plant Sci 92 : 47-55

**Bogenhagen DF, Brown DD (1981).** Nucleotide sequences in *Xenopus* 5S DNA required for transcription termination. Cell 24 : 261-270

**Bogenhagen DF, Sakonju S, Brown DD (1980).** A control region in the center of the 5S RNA gene directs specific initiation of transcription. II. The 3' border of the region. Cell 19 : 27-35

**Brown GR, Carlson JE (1997).** Molecular cytogenetics of the genes encoding 18S-5.8S-26S rRNA and 5S rRNA in two species of spruce (*Picea*). Theor Appl Genet 95 : 1-9

**Buckler ES, Ippolito A, Holtsford TP (1997).** The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. Genetics 145 : 821-832

**Campell BR, Song Y, Posch TE, Cullis CA, Town CD (1992).** Sequence and organization of 5S ribosomal RNA-encoding genes of *Arabidopsis thaliana*. Gene 112 : 225-228

**Capesius I (1991).** Sequence of the 5S ribosomal gene from *Sinapis alba*. Plant Mol Biol 17 : 169-170

**Cerbah M, Coulaud J, Siljak-Yakovlev S (1998).** rDNA organization and evolutionary relationships in the genus *Hypochaeris* (*Asteraceae*). J Heredity 89 : 312-318

**Charlesworth B, Sniegowski P, Stephan W (1994).** The evolutionary dynamics of repetitive DNA in Eukaryotes. Nature 371 : 215-220

**Challice JL, Segall J (1989).** Transcription of the 5S rRNA gene of *Saccharomyces cerevisiae* requires a promoter element at +1 and a 14-base pairs internal control region. J Biol Chem 264 : 20060-20067

**Christiansen J, Brown RS, Sproat BS, Garrett RA (1987).** *Xenopus* transcription factor IIIA binds primarily at junctions between double helical stems and internal loops in oocyte 5S RNA. EMBO J 6 : 453-460

**Ciliberto G, Rangei G, Constanzo F, Dente L, Cortese R (1983).** Common and interchangeable elements in the promoters of genes transcribed by RNA polymerase III. Cell 32 : 725-733

**Cox VA, Bennett MD, Dyer TA (1992).** Use of the polymerase chain reaction to detect spacer size heterogeneity in plant 5S-rRNA gene clusters and to locate such clusters in wheat (*Triticum aestivum* L). Theor Appl Genet 83 : 684-690

**Cronn RC, Zhao X, Paterson AH, Wendel JF (1996).** Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. J Mol Evol 42 : 685-705

**Cuadrado A, Jouve N (1997).** Distribution of highly repeated DNA sequences in species of the genus *Secale*. Genome 40 : 309-317

**Danna KJ, Workman R, Coryell V, Keim P (1996).** 5S rRNA genes in tribe *Phaseoleae*: array size, number, and dynamics. Genome 39 : 445-455

**Devereux R, Loehlich III AR, Pox GE (1990).** Higher plant origins and the phylogeny of green algae. J Mol Biol 31 : 18-24

**D'Hont A, Ison D, Alix K, Roux C, Glaszmann JC (1998).** Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. Genome 41 : 221-225

**Dong F, Hansen JC, Van Holde KE (1990).** DNA and protein determinants of nucleosome positioning on sea urchin 5S rRNA genes sequences in vitro. Proc Natl Acad Sci USA 87 : 5724-5728

**Doudrick RL, Heslop-Harrison JS, Nelson CD, Schmidt T, Nance WL, Scharzacher T (1995).** Karyotype of slash pine (*Pinus elliotti* var. *elliottii*) using patterns of fluorescence *in situ* hybridization and fluorochrome banding. J Heredity 86 : 289-296

**Dover GA (1982).** Molecular drive: a cohesive mode of species evolution. Nature 299 : 111-117

**Doyle JJ, Brown AHD (1989).** 5S ribosomal gene variation in the *Glycine tomentalla* polyploid complex (*Leguminosae*). Syst Bot 14 : 398-407

**Dvorák J, Zhang HB, Kota RS, Lassner M (1989).** Organization and evolution of the 5S ribosomal RNA gene family in wheat and related species. Genome 32 : 1003-1016

**Ellis THN, Lee D, Thomas CM, Simpson PR, Clealy WG, Newman MA, Burdam KWG (1988).** 5S rRNA genes in *Pisum*: sequence, long range and chromosomal organization. Mol Gen Genet 214 : 333-342

**Frasch M, Wenzel W, Hess D (1989).** The nucleotide sequence of nuclear 5S rRNA genes and spacer regions of *Petunia hybrida*. Nucleic Acids Res 17 : 2857

**Fukui K, Kamisugi Y, Sakai F (1994).** Physical mapping of 5S rDNA loci by direct-cloned biotinylated probes in barley chromosomes. Genome 37 : 105-111

**Fulnecek J, Matyàsek R, Kovarik A, Bezdek M (1998).** Mapping of 5-methylcytosine residues in *Nicotiana tabacum* 5S rRNA genes by genomic sequencing. Mol Gen Genet 259 : 133-141

**Gerlach WL, Dyer TA (1980).** Sequence organization of the repeating units in the nucleus of wheat which contain 5S rRNA genes. Nucleic Acids Res 8 : 4851-4865

**Goldsbrough PB, Ellis THN, Cullis CA (1981).** Organization of the 5S RNA genes in flax. Nucleic Acids Res 9 : 5895-5904

**Goldsbrough PB, Ellis THN, Lomonossoff GP (1982).** Sequence variation and methylation of the flax 5S RNA genes. Nucleic Acids Res 10 : 4501-4514

**Gorman SW, Teasdale RD, Cullis CA (1992).** Structure and organization of the 5S rRNA genes (5S DNA) in *Pinus radiata* (*Pinaceae*). Plant Syst Evol 183 : 223-234

**Gottlob-McHugh SG, Levesque M, MacKenzie K, Olson M, Yarosh O, Johnson DA (1990).** Organization of the 5S rRNA genes in the soybean *Glycine max* (L) Merrill and conservation of the 5S rDNA repeat structure in higher plants. Genome 33 : 486-494

**Gruenbaum Y, Naveh-Many T, Cedar H, Razin A (1981).** Sequence specificity of methylation in higher plant DNA. Nature 292 : 860-862

**Halanych KM (1991).** 5S ribosomal RNA sequences inappropriate for phylogenetic reconstruction. Mol Biol Evol 8 : 249-253

**Hanson RE, Islam-Faridi NM, Percival EA, Crane CF, Ji Y, McKnight TD, Stelly DM, Price HJ (1996).** Distribution of 5S and 18S-28S rDNA loci in a tetraploid cotton (*Gossypium hirsutum* L.) and its putative diploid ancestors. Chromosoma 105 : 55-61

**Hariharan N, Reddy PS, Padayatty JD (1987).** 5S rRNA genes in rice embryos. Plant Mol Biol 9 : 443-451

**Hayes JJ, Tullius TD (1992).** Structure of the TFIIIA-5S DNA complex. J Mol Biol 227 : 407-417

**Hemleben V, Werts D (1988).** Sequence organization and putative regulatory elements in the 5S rRNA genes of two higher plants (*Vigna radiata* and *Matthiola incana*). Gene 62 : 165-169

**Henry RJ (1997).** *Practical applications of plant molecular biology*. Chapman & Hall, Londre

**Hodges-Garcia Y, Hagerman PJ (1992).** Cytosine methylation can induce local distorsions in the structure of duplex DNA. Biochem 31 : 7595-7599

**Hori H, Osawa S (1987).** Origin and evolution of organisms as deduced from 5S ribosomal RNA sequences. Mol Biol Evol 4 : 445-472

**Hori H, Lim BL, Osawa S (1985).** Evolution of green plants as deduced from 5S rRNA sequences. Proc Nat Acad Sci USA 82 : 820-823

**Jeddoloh JA, Richard EJ (1996).** mCCG methylation in angiosperms. Plant J 9 : 579-586

**Johnson DA, Chan CCY, Gottlob-McHugh SG, Mackenzie K, Marengère L, Prud'homme MC (1992).** Structure of the 5S rRNA genes in birch (*Betula papyrifera*) and Alder (*Alnus incana*). Genome 35 : 337-341

**Kamisugi Y, Nakayama S, Nakajima R, Ohtsubo H, Ohtsubo E, Fukui K (1994).** Physical mapping of the 5S ribosomal RNA genes on rice chromosome 11. Mol Gen Genet 245 : 133-138

**Kanazin V, Ananiev E, Blake T (1993).** The genetics of 5S rRNA encoding multigene families in barley. Genome 36 : 1023-1028

**Katsiotis A, Hagidimitriou M, Heslop-Harrison JS (1997).** The close relationship between the A and B genomes in *Avena L.* (*Poaceae*) determined by molecular cytogenetic analysis of total genomic, tandemly and dispersed repetitive DNA sequences. Ann Bot 79 : 103-109

**Kellogg EA, Appels R (1995).** Intraspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. Genetics 140 : 325-343

**Knälmann M, Burger EC (1977).** Cytologische lokalisation von 5S und 18S/25S RNA genorten in mitose-chromosomen von *Vicia faba*. Chromosoma 61 : 177-192

**Kolchinsky A, Gresshoff PM (1992).** Nucleotide sequence of the 5S rRNA gene from *Glycine soja*. Plant Mol Biol 19 : 1045-1047

**Korn LJ (1982).** Transcription of *Xenopus* 5S ribosomal RNA genes. Nature 295 : 101-105

**Lagudah ES, Clarke BC, Appels R (1989).** Phylogenetic relationships of *Triticum tauschii*, the D-genome donor to hexaploid wheat. 4. Variation and chromosomal location of 5S DNA. Genome 32 : 1017-1025

**Lapitan NLV, Ganal MW, Tanksley SD (1991).** Organization of the 5S ribosomal RNA genes in the genome of tomato. Genome 34 : 509-514

**Laurie DA, Pratchett N, Devos KM, Leitch IJ, Gale MD (1993).** The distribution of RFLP markers on chromosome 2(2H) of barley in relation to the physical and genetic location of 5S rDNA. Theor Appl Genet 87 : 177-183

**Lee Y, Erkine AM, Van Ryk DI, Nazar RN (1995).** *In vivo* analyses of the internal control region in the 5S rRNA gene from *Saccharomyces cerevisiae*. Nucleic Acids Res 23 : 634-640

**Leitch IJ, Heslop-Harrison JS (1993).** Physical mapping of four sites of 5S rDNA sequences and one site of the α-amylase-2 gene in barley (*Hordeum vulgare*). Genome 36 : 517-523

**Lubaretz O, Fuchs J, Ahne R, Meister A, Schubert I (1996).** Karyotyping of three *Pinaceae* species via fluorescent *in situ* hybridization and computer-aided chromosome analysis. Theor Appl Genet 92 : 411-416

**Mascia PN, Rubenstein I, Philliph RI, Wang AS, Xiang LZ (1981).** Localization of the 5S RNA genes and evidence for diversity in the 5S rDNA region of maize. Gene 15 : 7-20

**McBryant J, Veldhoen N, Geldulin B, A, Foster MP, Wright PE, Romaniuk PJ, Gottesfeld JM (1995).** Interaction of the RNA binding fingers of *Xenopus* transcription factor IIIA with specific regions of 5S ribosomal RNA. J Mol Biol 248 : 44-57

**McIntyre CL, Winberg B, Houchins K, Appels R, Baum BR (1992).** Relationships between *Oryza* species (*Poaceae*) based on 5S DNA sequences. Plant Syst Evol 183 : 249-264

**McMahon ME, Stamenkovich D, Petes TD (1984).** Tandemly arranged variant 5S ribosomal RNA genes in the yeast *Saccharomyces cerevisiae*. Nucleic Acids Res 12 : 8001-8017

**Mizukami H (1995).** Amplification and sequence of a 5S rRNA gene spacer region from the crude drug "Angelica root". Biol Pharm Bull 18 : 1299-1301

**Moran GF, Smith D, Bell JC, Appels R. (1992).** The 5S RNA genes in *Pinus radiata* and the spacer region as a probe for relationships between *Pinus* species. Plant Syst Evol 183 : 209-221

**Mukai Y, Endo TR, Gill BS (1990).** Physical mapping of the 5S rRNA multigene family in common wheat. J Heredity 81 : 290-295

**Murata M, Heslop-Harrison JS, Motoyoshi F (1997).** Physical mapping of the 5S ribosomal RNA genes in *Arabidopsis thaliana* by multi-color fluorescence *in situ* hybridization with cosmid clones. Plant J 12 : 31-37

**Nei M (1987).** *Molecular evolutionary biology.* Colombia University Press, New York

**Nelson DW, Linning RM, Davison PJ, Honda BM (1998).** 5'-flanking sequences required for efficient transcription *in vitro* of 5S RNA genes, in the related nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*. Gene 218 : 9-16

**Nielsen JN, Hallenberg C, Frederiksen S, Sorensen PD, Lomholt B (1993).** Transcription of human 5S rRNA genes is influenced by an upstrean DNA sequence. Nucleic Acids Res 21 : 3631-3636

**Nygard O, Nilson L (1987).** The ribosomal binding site for eukaryotic elongation factor EF2 contains 5S rRNA. Biochim Biophys Acta 908 : 46-51

**Pâques F, Samson ML, Jordan P, Wegnez M (1995).** Structural evolution of the *Drosophila* 5S ribosomal genes. J Mol Evol 41 : 615-621

**Pieler T, Hamm J, Roeder RG (1987).** The 5S gene internal control region is composed of three distinct sequence elements, organized as two functional domains with variable spacing. Cell 48 : 91-100

**Pieler T, Oei SL, Hamm J, Engelke U, Erdmann VA (1985).** Functional domains of the *Xenopus laevis* 5S gene promoter. EMBO J 4 : 3751-3756

**Playford J, Appels R, Baum BR (1992).** The 5S DNA units of *Acacia* species (*Mimosaceae*). Plant Syst Evol 183 : 235-247

**Polanco C, González AI, de la Fuente A, Dover GA (1998).** Multigene family of ribosomal DNA in *Drosophila melanogaster* reveals contrasting patterns of homogenization for IGS and ITS spacer regions : a possible mechanism to resolve this paradox. Genetics 149 : 243-256

**Prado EA, Faivre-Rampant P, Schneider C, Darmency MA (1996).** Detection of a variable number of ribosomal DNA loci by fluorescent *in situ* hybridization in *Populus* species. Genome 39 : 1020-1026

**Rafalski JA, Wiewiorowski M, Söll D (1982).** Organization and nucleotide sequence of nuclear 5S rRNA genes in yellow lupin (*Lupinus luteus*). Nucleic Acids Res 23 : 7635-7642

**Rasmussen SK, Avato P (1998).** Characterization of chromosomes and genome organization of *Thapsia garganica* L. by localization of rRNA genes using fluorescent *in situ* hybridization. Hereditas 129 : 231-239

**Reddy P, Appels R (1989).** A second locus for the 5S multigene family in *Secale* L.: sequence divergence in two lineages of the family. Genome 32 : 456-467

**Reynolds WF, Azer K (1988).** Sequence differences upstream of the promoters are involved in the differential expression of the *Xenopus* somatic and oocyte 5S RNA genes. Nucleic Acids Res 16 : 3391-3403

**Reznikoff WS, Siegele DA, Cowing DW, Gross CA (1985).** The regulation of transcription initiation in bacteria. Ann Rev Genet 19 : 355-387

**Röder MS, Sorrells ME, Tanksley SD (1992).** 5S ribosomal gene clusters in wheat : pulsed field gel electrophoresis reveals a high degree of polymorphism. Mol Gen Genet 232 : 215-220

**Sastri DC, Hilu K, Appels R, Lagudah ES, Playford J, Baum BR (1992).** An overview of evolution in plant 5S DNA. Plant Syst Evol 183 : 169-181

**Sanderson MJ, Doyle JJ (1992).** Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy, and confidence. Syst Biol 41 : 4-17

**Schlötterer C, Tautz D (1994)**. Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. Curr Biol 4 : 777-783

**Schneeberger RG, Cullis CA (1992)**. Intraspecific 5S rRNA gene variation in flax, *Linum usitatissimum* (*Linaceae*). Plant Syst Evol 183 : 265-280

**Schneeberger RG, Creissen GP, Cullis CA (1989)**. Chromosomal and molecular analysis of 5S RNA gene organization in flax. Gene 83 : 75-84

**Schondelmaier J, Schmidt T, Heslop-Harrison JS, Jung C (1997)**. Genetic and chromosomal localization of the 5S rDNA locus in sugar beet (*Beta vulgaris* L.). Genome 40 : 171-175

**Schuster M, Fuchs J, Schubert I (1997)**. Cytogenetics in fruit breeding. Localization of ribosomal RNA genes on chromosomes of apple (*Malus* x *domestica* Borkh.). Theor Appl Genet 94 : 322-324

**Scoles GJ, Gill BS, Xin ZY, Clarke BC, McIntyre CL, Chapman C, Appels R (1988)**. Frequent duplication and deletion events in the 5S RNA genes and the associated spacer regions of the *Triticeae*. Plant Syst Evol 160 : 105-122

**Sharp SJ, Garcia AD (1988)**. Transcription of the *Drosophila melanogaster* 5S rRNA gene requires an upstream promoter and four intragenic sequence elements. Mol Cell Biol 8 : 1266-1274

**Singer M, Berg P (1992)**. *Genes & Genomes*. Editions Vigot, Paris

**Smith TPL, Young LS, Bender LB, Sprague KU (1995)**. Silkworm TFIIIA requires additional class III factors for commitment to transcription complex assembly on a 5S RNA gene. Nucleic Acids Res 23 : 1244-1251

**Song YC, Gustafson JP (1993)**. Physical mapping of the 5S rDNA gene complex in rice (*Oryza sativa*). Genome 36 : 658-661

**Sorensen PD, Frederiksen S (1991)**. Characterization of human 5S RNA genes. Nucleic Acids Res 19 : 4147- 4151

**Steele KP, Holsinger KE, Jansen RK, Taylor DW (1991)**. Assessing the reliability of 5S rRNA sequence data for phylogenetic analysis in green plants. Mol Biol Evol 8 : 240-248

**Sturges MR, Bartilson M, Peck LJ (1995)**. Interaction of *Xenopus* TFIIIC with a 5S RNA gene. Nucleic Acids Res 23 : 1551-1556

**Szymanski M, Barciszewska M, Erdmann VA, Barciszewski J (1995)**. Are 5S ribosomal RNA primary transcripts edited ? Biochem Mol Biol Int 37 : 413-422

**Szymanski M, Specht T, Barciszewska MZ, Barciszewski J, Erdmann VA (1998)**. 5S rRNA data bank. Nucleic Acids Res 26 : 156-159

**Taketa S, Harrison GE, Heslop-Harrison JS (1999)**. Comparative physical mapping of the 5S and 18S-25S rDNA in nine wild *hordeum* species and cytotypes. Theor Appl Genet 98 : 1-9

**Tschudi C, Pirotta V (1980)**. Sequence and heterogeneity in the 5S RNA gene cluster of *Drosophila melanogaster*. Nucleic Acids Res 8 : 441-451

**Tyler BM (1987)**. Transcription of *Neurospora crassa* 5S rRNA genes requires a TATA box and three internal elements. J Mol Biol 196 : 801-811

**Udovicic F, McFadden G, Ladiges PY (1995)**. Phylogeny of *Eucalyptus* and *Angophora* based on 5S rDNA spacer sequence data. Mol Phyl Evol 4 : 247-256

**Van de Peer YR, De Baere R, Cauwenberghs J, De Wachter R (1990)**. Evolution of green plants and their relationship with other photosynthetic eukaryotes as deduced from 5S ribosomal RNA sequences. Plant Syst Evol 170 : 85-96

**VanCampenhout S, Aert R, Volckaert G (1998)**. Orthologous DNA sequence variation among 5S ribosomal RNA gene spacer sequences on homologous chromosomes 1B, 1D and 1R of wheat and rye. Genome 41 : 244-255

**Venkateswarlu K, Lee SW, Nazar RN (1991)**. Conserved upstream sequence elements in plant 5S ribosomal RNA-encoding genes. Gene 105 : 249-253

**Wendel JF, Schnabel A, Seelanan T (1995)**. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc Natl Acad Sci USA 92 : 280-284

**Wolffe AP, Brown DD (1988)**. Developmental regulation of two 5S ribosomal RNA genes. Science 241 : 1626-1632

**Zanke C, Borisjuk N, Ruoss B, Schilde-Rentschler L, Ninnemann H, Hemleben V (1995)**. A specific oligonucleotide of the 5S rDNA spacer and species-specific elements identify symmetric somatic hybrids between *Solanum tuberosum* and *S. pinnatisectum*. Theor Appl Genet 90 : 720-726

# *PARTIE EXPERIMENTALE*

# PRESENTATION GENERALE

## I- Espèces étudiées

Le bilan bibliographique précédent montre que la connaissance de la structure, de l'organisation, du polymorphisme et de l'évolution de l'ADNr 5S chez les plantes supérieures a essentiellement été acquise à partir d'espèces herbacées à cycle court, annuelles ou pérennes, le plus souvent domestiquées. La base de données obtenue dans le groupe des *Triticeae* et plus largement dans la famille des *Poaceae* est ainsi particulièrement imposante (Sastri *et al.* 1992). Les informations disponibles concernant les espèces ligneuses à temps de génération long et à maturité sexuelle tardive comme les arbres forestiers sont comparativement peu nombreuses et fragmentaires (**Tableau 16**). Elles se limitent souvent à l'estimation de la taille approximative des unités, à l'établissement de la séquence d'un ARNr 5S ou à la détermination du nombre de loci et/ou de leur localisation chromosomique. L'étude des variations intraspécifiques des unités d'ADNr 5S n'a ainsi été entreprise que pour 11 arbres angiospermes (6 espèces d'*Acacia*, Playford *et al.* 1992, *Alnus incana* et *Betula papyrifera*, Johnson *et al.* 1992, *Arillastrum gummiferum*, *Eucalyptus calophylla* et *Eucalyptus obliqu*a, Udovicic *et al.* 1995) et 3 gymnospermes (*Picea glauca*, Brown et Carlson 1997, *Pinus radiata*, Moran *et al.* 1992, *Pseudotsuga menziesii*, Amarasinghe et Carlson 1998). Ces travaux portent surtout sur l'évaluation du polymorphisme intragénomique (souvent à partir d'un petit nombre de données) mais très peu sur les variations entre les individus (polymorphisme RFLP entre 25 individus de *P. radiata*, variations de séquence entre 2 individus d'*E. calophylla* ou 4 individus d'une population d'*E. obliqua*). L'analyse de l'évolution interspécifique des unités d'ADNr 5S a seulement été initiée dans la famille des *Myrtaceae* (essentiellement le genre *Eucalyptus*, Udovicic *et al.* 1995), le genre *Acacia* (Playford *et al.* 1992) et le genre *Pinus* (polymorphisme RFLP, Moran *et al.* 1992).

Notre contribution à l'étude du polymorphisme et de l'évolution de l'ADNr 5S chez les plantes angiospermes et gymnospermes à long cycle porte sur les deux espèces sympatriques *Quercus robur* L. (chêne pédonculé) et *Quercus petraea* (Matt.) Liebl. (chêne sessile) ainsi que sur les deux espèces allopatriques *Larix decidua* Mill. (mélèze d'Europe) et *Larix kaempferi* (Lamb.) Carr. (mélèze du Japon). Ces deux couples d'espèces apparentées et compatibles (les barrières reproductives étant faibles) de chênes (présentation dans la **Partie I** du **Chapitre III**) et de mélèzes (présentation dans la **Partie I** du **Chapitre IV**) soulèvent différents problèmes d'identification moléculaire, en particulier des hybrides interspécifiques, auxquels l'ADNr 5S est susceptible de répondre.

**Tableau 16.** Informations disponibles concernant l'ADNr 5S chez les espèces ligneuses à long cycle*.

| Espèce[a] | LC | NU | TU | SA | SG | SE | VIG | VEI | VIS |
|---|---|---|---|---|---|---|---|---|---|
| Gymnospermes | | | | | | | | | |
| *Cycas revoluta* | | | * | * | | | | | |
| *Ephedra kokanica* | | | | * | | | | | |
| *Ginkgo biloba* | | | | * | | | | | |
| *Gnetum gnemon* | | | | * | | | | | |
| *Juniperus media* | | | | * | | | | | |
| *Juniperus virginiana* | | | * | | | | | | |
| *Larix decidua* | * | | | | | | | | |
| *Metasequoia glyptostroboides* | | | | * | | | | | |
| *Picea abies* | * | | | | | | | | |
| *Picea glauca* | * | * | * | | * | * | | | |
| *Picea sitchensis* | * | | | | | | | | |
| *Pinus* (30 sp) | | | | | | | | | * |
| *Pinus contorta* | | | | * | | | | | |
| *Pinus elliottii* | * | | | | | | | | |
| *Pinus radiata* (25 ind) | * | * | * | | * | * | * | * | |
| *Pinus sylvestris* | * | | | * | | | | | |
| *Pseudotsuga menziesii* | * | | * | | * | * | * | | |
| *Taxus baccata* | | | | * | | | | | |
| Angiospermes | | | | | | | | | |
| *Acer rubrum* | | | * | | | | | | |
| *Acacia* (7 sp) | | | * | | * | * | * | | * |
| *Albizia lebbek* | | | * | | | | | | |
| *Alnus incana* | | | * | | * | * | * | | |
| *Arillastrum gummiferum* | | | * | | * | * | * | | |
| *Betula papyrifera* | | | * | | * | * | * | | |
| *Casuarina equisetifolia* | | | * | | | | | | |
| *Ceratonia siliqua* | | | * | | | | | | |
| *Elaeagnus angustifolia* | | | * | | | | | | |
| *Eucalyptus calophylla* (2 ind) | | | * | | * | * | | * | |
| *Eucalyptus obliqua* (4 ind) | | | * | | * | * | | * | |
| *Gleditsia triacanthos* | | | * | | | | | | |
| *Gymnocladus dioicus* | | | * | | | | | | |
| *Magnolia cobus* | | | | * | | | | | |
| *Magnolia stellata* | | | | * | | | | | |
| *Malus x domestica* | * | | | | | | | | |
| *Myrtaceae : Allosyncarpia* (1 sp) | | | * | | * | * | | | * |
| *Angophora* (2 sp) | | | | | | | | | |
| *Arillastrum* (1 sp) | | | | | | | | | |
| *Eucalyptopsis* (1 sp) | | | | | | | | | |
| *Eucalyptus* (19 sp) | | | | | | | | | |
| *Populus nigra* | * | | | | | | | | |
| *Populus deltoides* | * | | * | | | | | | |
| *Populus euroamericana* | * | | | | | | | | |
| *Populus balsamifera* | * | | | | | | | | |
| *Populus alba* | * | | | | | | | | |
| *Syringa vulgaris* | | | * | | | | | | |

* LC : nombre de loci et/ou localisation chromosomique (**Tableau 12**) ; NU : nombre d'unités (**Tableau 13**) ; TU : taille des unités (**Tableau 14**) ; SG : séquence du gène ; SA : séquence de l'ARNr 5S ; SE : séquence de l'espaceur ; VIG : variation intragénomique (**Tableau 15**) ; VEI : variation entre les individus ; VIS : variations interspécifiques.

[a] ind : individu ; sp : espèce.

## II- Objectifs

La structure et l'organisation de l'ADNr 5S est encore quasiment inconnue dans les genres *Quercus* et *Larix*. A notre connaissance, seule la localisation chromosomique obtenue par hybridation *in situ* d'une sonde hétérologue a été publiée récemment pour le mélèze d'Europe (un locus en position subterminale sur le bras court du chromosome 3, Lubaretz *et al.* 1996). Plusieurs études préliminaires ont été réalisées dans notre équipe et concernent essentiellement les chênes pédonculé et sessile (Senequier 1992, Favre et Arnould 1996) ainsi que le mélèze d'Europe, le mélèze du Japon et leurs hybrides interspécifiques (Trontin 1993, Grandemange 1994). Ces travaux ont permis d'obtenir quelques éléments de structure et de polymorphisme des unités qui suggèrent que cette famille multigénique présente un intérêt à différents niveaux taxinomiques, de la détection de variations somaclonales au sein de cultures embryogènes (Trontin 1993) à la différenciation interspécifique (Grandemange 1994, Favre et Arnould 1996). Nous avons donc articulé notre investigation autour des grands thèmes suivants : structure, organisation, polymorphisme et évolution des unités d'ADNr 5S.

### 1- Structure des unités

Nous chercherons tout d'abord à déterminer la taille et la séquence nucléotidique des gènes et des espaceurs pour les différentes classes d'unités détectées dans le génome des 4 espèces modèles. Ces informations seront ensuite exploitées pour accéder à la nature et à la topographie des variations et aux mécanismes de l'évolution structurale des unités. Enfin, l'organisation fonctionnelle de cette famille de gènes sera précisée par comparaison avec les données disponibles chez les plantes supérieures (voir la **Partie I** du **Chapitre II**).

### 2- Organisation des unités

Le nombre d'unités est un paramètre très variable au sein des espèces (**Partie III-1** du **Chapitre II**) et difficile à déterminer avec précision. De plus, son intérêt paraît limité. Nous ne proposerons ici qu'une estimation préliminaire pour un chêne pédonculé et un mélèze d'Europe. Notre effort portera davantage sur l'analyse de la structure et de la composition des séries de l'ADNr 5S. Au travers de l'étude de quelques croisements, nous tenterons en particulier de formuler quelques génotypes en étudiant l'héritabilité des variants alléliques afin d'accéder au nombre de loci et éventuellement à leur localisation génomique par cartographie génétique. Malgré son intérêt, ce dernier objectif a rarement été atteint chez les plantes supérieures (**Partie II-1-b** du **Chapitre II**).

### 3- Polymorphisme et évolution des unités

L'évaluation de la diversité moléculaire intraspécifique des unités d'ADNr 5S ne se limitera pas aux seules variations intragénomiques dont l'importance est directement liée à l'intensité de l'évolution concertée. Elle sera plus largement abordée à partir d'un ensemble d'individus de différentes provenances issues de l'ensemble de l'aire de répartition de chaque espèce. A notre connaissance, très

peu d'études concernant l'ADNr 5S ont été entreprises sur un tel dispositif géographique (voir les **Parties III-2** et **III-3-b** du **Chapitre II**). Pour ce faire, nous estimerons le polymorphisme des gènes et des espaceurs en poussant l'analyse jusqu'au stade ultime du séquençage. Nous nous efforcerons alors de dégager l'éventuelle structuration géographique des variations en liaison avec la phylogenèse des variants alléliques qui permettra par ailleurs d'établir les relations de paralogie et d'orthologie au sein des deux systèmes modèles. Ces données doivent nous permettre de contribuer à l'évaluation du niveau d'information taxinomique fourni par cette famille multigénique chez les espèces à long cycle, de la génétique des populations à l'analyse phylogénétique (**Tableau 1**).

```
                          10        20        30        40        50
Cons. plastidiale         TATTCTGGTGTCCTAGGC-GTAGAGGA-ACCACACCAAT-CCATCCCGAACTTGGT-GGTTAAAC
                                :  :   :     :         :  ::: :  :: :::: : :::::  :   :::::  :
Cons. nucléaire           --G-GATGCGATCATACC-AGCACTAA-TGCAC-CGGATCCCAT-CAGAACTCCGA-AGTTAAGC
                                :: :  :      ::    : :   ::: : :::::::: : :: :::      ::   :
Cons. mitochondriale      AAACCGGGC-ACTACGGTGAGACGTGAAAACAC-CCGATCCCATTCCGACCTCGATATGTGGAAT
                                  ←──────────────────────────
                                                B

                          60        70        80        90       100       110       120
Cons. plastidiale         TCTACTGCGGT-GAC-GATACTGTAG-GGGAGGTCCTGCGGAAAAATAGCTCGACGCCAGGAT-
                               :: :   ::    ::::     ::: :  :    :: ::     :  :  ::
Cons. nucléaire           GTGCTTGGGCGAGAGTAGTACTAGGATGGGTGACCTCCTGGGAAGTCCTCGTGTTGCATC-CCC
                           :::: ::      :::::        :        ::::                  ::
Cons. mitochondriale      CGTCTTGCGCC-ATA-TGTACTGAA----ATTG---TTCGGGAGACATGGTCAAAGCTCGGAAA
                                                                  ──────────────────────→
                                                                              F
```

**Figure 21.** Comparaison des séquences consensus (Cons.) établies à partir de gènes 5S et/ou d'ARNr 5S d'origine plastidiale, nucléaire ou mitochondriale chez les plantes supérieures (Szymanski *et al.* 1998). Les positions nucléotidiques sont numérotées par rapport à la séquence consensus nucléaire. Les régions encadrées B (positions 24-5) et F (positions 96-115) ont été utilisées pour développer des couples d'amorces (avec ou sans introduction de sites de restriction, voir **Matériel et Méthodes** des **Chapitres II** et **III**) dites universelles chez les plantes supérieures qui permettent d'amplifier spécifiquement les espaceurs des unités d'ADNr 5S nucléaire (leur orientation est indiquée par des flèches). Ces amorces ne s'hybrident pas sur les gènes 5S plastidiaux ou mitochondriaux.

## III- Stratégie d'accès à l'ADNr 5S nucléaire des plantes supérieures

Nos travaux s'appuient largement sur la technique de PCR (Saiki *et al.* 1988) qui est particulièrement bien adaptée à l'étude des familles de séquences répétées en tandem. Elle permet d'accéder rapidement, sélectivement et avec une grande sensibilité de détection à l'ensemble des gènes et des espaceurs d'une série. De nombreuses séquences de gènes 5S et d'ARNr 5S établies depuis plus de 30 ans chez divers organismes procaryotes (485 séquences, 381 espèces) et eucaryotes (1094 séquences, 464 espèces) ainsi que chez les organelles (34 séquences plastidiales et 9 séquences mitochondriales dont 6 de plantes supérieures) sont maintenant accessibles dans la "5S rRNA Data Bank" (http://www.chemie.fu-berlin.de/fb_chemie/agerdmann/5S_rRNA.html.) développée par

Szymanski *et al.* (1998). Cette base de données comporte en particulier 490 séquences de gènes 5S ou d'ARNr 5S nucléaires isolés à partir de 143 espèces de plantes supérieures (dont 134 spermaphytes). Leur grande conservation autorise la construction de séquences consensus à partir desquelles des amorces de PCR dites universelles chez les plantes supérieures peuvent être sélectionnées. Nous avons utilisé des couples d'amorces initialement développées par Kolchinsky *et al.* (1991), Cox *et al.* (1992) et Ko et Henry (1996) qui s'hybrident sur les bordures des gènes 5S nucléaires au niveau de régions de très faible similarité avec les gènes 5S d'origine plastidiale ou mitochondriale (**Figure 21**). Ces amorces sont donc spécifiques de l'ADNr 5S nucléaire. Leur orientation a été choisie de manière à amplifier sélectivement les espaceurs variables qui alternent avec les gènes. Des couples d'amorces permettant d'accéder au gène dans son intégralité peuvent alors être recherchés à partir des séquences des espaceurs. En raison du polymorphisme intragénomique de l'ADNr 5S, le séquençage direct des produits de PCR n'est pas approprié (voir Baum et Johnson 1994) et nous avons donc opté pour un clonage préalable systématique. Des sites de restriction ont pu être introduit dans les amorces sans effet apparent sur les profils d'amplification (voir **Matériel et Méthodes** des **Chapitres III et IV**).

## Bibliographie


**Amarasinghe V, Carlson JE (1998).** Physical mapping and characterization of 5S rRNA genes in Douglas-fir. J Heredity 89 : 495-500

**Baum BR, Johnson DA (1994).** The molecular diversity of the 5S rRNA gene in barley (*Hordeum vulgare*). Genome 37 : 992-998

**Brown GR, Carlson JE (1997).** Molecular cytogenetics of the genes encoding 18S-5.8S-26S rRNA and 5S rRNA in two species of spruce (*Picea*). Theor Appl Genet 95 : 1-9

**Cox VA, Bennett MD, Dyer TA (1992).** Use of the polymerase chain reaction to detect spacer size heterogeneity in plant 5S-rRNA gene clusters and to locate such clusters in wheat (*Triticum aestivum* L). Theor Appl Genet 83 : 684-690

**Favre JM, Arnould MF (1996).** Nuclear DNA content and inter-specific ribosomal variation between *Q. robur* and *Q. petraea*. Dans *inter- and intra-specific variation in European oaks: evolutionary implications and practical consequences* pp 41-51. Edité par Kremer A, Mühs H. European Commission, EUR 16717EN, Luxembourg

**Grandemange C (1994).** L'ADN ribosomique 5S chez le mélèze : clonage, séquençage et essai d'utilisation pour le typage de différentes espèces. DEA en Biologie Forestière, 18 pp. Université Henri Poincaré, Nancy I

**Johnson DA, Chan CCY, Gottlob-McHugh SG, Mackenzie K, Marengère L, Prud'homme MC (1992).** Structure of the 5S rRNA genes in birch (*Betula papyrifera*) and Alder (*Alnus incana*). Genome 35 : 337-341

**Ko HL, Henry RJ (1996).** Specific 5S ribosomal RNA primers for plant species identification in admixtures. Plant Mol Biol Rep 14 : 33-43

**Kolchinsky A, Kolesnikova M, Ananiev E (1991).** "Portraying" of plant genomes using polymerase chain reaction amplification of ribosomal 5S genes. Genome 34 : 1028-1031

**Lubaretz O, Fuchs J, Ahne R, Meister A, Schubert I (1996).** Karyotyping of three *Pinaceae* species via fluorescent *in situ* hybridization and computer-aided chromosome analysis. Theor Appl Genet 92 : 411-416

**Moran GF, Smith D, Bell JC, Appels R. (1992).** The 5S RNA genes in *Pinus radiata* and the spacer region as a probe for relationships between *Pinus* species. Plant Syst Evol 183 : 209-221

**Playford J, Appels R, Baum BR (1992).** The 5S DNA units of *Acacia* species (*Mimosaceae*). Plant Syst Evol 183 : 235-247

**Saiki RK, Gelfand DH, Stoffel S, Starf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988).** Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239 : 487-491

**Sastri DC, Hilu K, Appels R, Lagudah ES, Playford J, Baum BR (1992).** An overview of evolution in plant 5S DNA. Plant Syst Evol 183 : 169-181

**Senequier S (1992).** Etude de la variabilité génétique de l'ADN ribosomique chez plusieurs clones de vitroplants de *Quercus robur* L. et *Quercus petraea* Liebl. DEA en Biologie Forestière, 29 pp. Université Henri Poincaré, Nancy I

**Szymanski M, Specht T, Barciszewska MZ, Barciszewski J, Erdmann VA (1998).** 5S rRNA data bank. Nucleic Acids Res 26 : 156-159

**Trontin (1993).** L'ADN ribosomique 5S chez les mélèzes (*Larix decidua* Mill., *Larix x eurolepis* Henry, *Larix x leptoeuropaea* Dengler) : éléments d'organisation, stabilité au sein de cultures embryogènes. DEA en Biologie Forestière, 20 pp. Université Henri Poincaré, Nancy I

**Udovicic F, McFadden G, Ladiges PY (1995).** Phylogeny of *Eucalyptus* and *Angophora* based on 5S rDNA spacer sequence data. Mol Phyl Evol 4 : 247-256

# CHAPITRE III : Etude de l'ADNr 5S chez les chênes pédonculé et sessile

## I- Les chênes

### 1- Importance, origine et distribution

Les chênes (*Quercus*) constituent le groupe d'arbres (rarement des arbustes) le plus important de la famille des *Fagaceae* (plusieurs centaines d'espèces) qui comprend également les hêtres (*Fagus*, une dizaine d'espèces), les *Nothofagus* (une cinquantaine d'espèces) et les châtaigniers (*Castanea*, une quinzaine d'espèces). Ils constituent sans doute un des genres d'arbres forestiers les plus riches en espèces, mais aussi un des plus controversés. Selon les auteurs, 200 à 600 espèces sont en effet répertoriées et, dans sa revue récente, Nixon (1993) dénombre 394 à 448 espèces en fonction des critères de classification adoptés. L'histoire des chênes commence probablement il y a 90 à 130 Ma à la fin de l'ère secondaire (Crétacé inférieur) mais les plus vieux fossiles connus datent seulement du milieu de l'ère tertiaire (Oligocène, -38 Ma). L'expansion et la diversification du genre *Quercus* seraient intervenues dans plusieurs régions du monde à la fin de l'ère tertiaire au cours du Miocène (-25 Ma) et du Pliocène (-5 Ma), y compris dans le Pacifique Sud où les chênes ne poussent plus aujourd'hui (Nixon 1993). Modelées en grande partie par la séparation des continents (effective au cours de l'Oligocène) et plus récemment par le flux et le reflux des grandes glaciations qui caractérisent l'ère quaternaire (Pléistocène, à partir de -2 Ma, Huntley et Birk 1983), les aires de distribution des espèces actuelles sont en effet uniquement réparties dans les zones tempérées, subtropicales et même tropicales de l'hémisphère boréal. C'est sur le continent américain que se trouve la plus forte concentration de chênes (Nixon 1993). Depuis les hauts-plateaux mexicains, ils s'étendent à travers la partie est de l'Amérique du Nord (environ 60 espèces), l'ouest des Etats-Unis (environ 25 espèces) et l'Amérique Centrale (150-200 espèces) pour atteindre les Andes colombiennes (1 espèce) et même les Antilles (1 espèce). Sur l'ancien continent, la plupart des chênes (environ 150 espèces) rayonnent depuis le Moyen-Orient en passant par l'Inde vers l'Asie tempérée et plus au sud vers la péninsule malaysienne (Nixon 1993). En revanche, un plus petit nombre d'espèces occupent l'Europe tempérée et la région méditerranéenne, du Proche-Orient à l'Afrique du Nord-Ouest : une dizaine (8 à plus de 12) selon Nixon (1993) et 19 selon Schwarz (1993). Au moins 9 d'entre elles se trouvent en France (*Q. robur* L., *Q. petraea* (Matt.) Liebl., *Q. pubescens* Willd., *Q. pyrenaica* Willd., *Q. frainetto* Ten., *Q. cerris* L., *Q. suber* L., *Q. ilex* L., *Q. coccifera* L.).

**Tableau 17.** Synopsis du genre *Quercus*.

| Sous-genre | Section | Répartition | Sous-section[a] | Espèce(s) type(s) |
|---|---|---|---|---|
| *Cyclobalanopsis* | | Asie de l'Est et Malaisie | A | *Q. lineata, Q. oxyodon* |
| | | | B | *Q. myrsinifolia* |
| | | | C | *Q. glauca* |
| | | | D | *Q. acuta* |
| *Quercus* (*Euquercus*) | *Lobatae* (*Erythrobalanus*) | Est de l'Amérique du Nord | A | *Q. arkansana, Q. marilandica* |
| | | | B | *Q. nigra, Q. laevis* |
| | | | C | *Q. falcata* |
| | | | D | *Q. ilicifolia* |
| | | | E | *Q. palustris, Q. georgiana* |
| | | | F | *Q. velutina* |
| | | | G | *Q. rubra, Q. coccinea, Q. ellipsoidalis* |
| | | | H | *Q. phellos, Q. imbricaria, Q. laurifolia* |
| | | Etats-Unis de l'Ouest | A | *Q. agrifolia* |
| | | | B | *Q. hypoleucoides* |
| | | Amérique Centrale | A | *Q. crassifolia* |
| | | | B | *Q. crassipes* |
| | *Protobalanus* | Ouest de l'Amérique du Nord | | *Q. chrysolepsis* |
| | | | | *Q. vacciniifolia* |
| | | | | *Q. palmeri* |
| | *Quercus* (*Lepidobalanus*) | Ancien Monde | A | *Q. lanata, Q.lodicosa* |
| | | | B | *Q. engleriana* |
| | | | C | *Q. ilex* |
| | | | D | *Q. aliena, Q. glandulifera, Q. mongolica* |
| | | | E | *Q. boissieri, Q. canariensis, Q. faginea* |
| | | | F | *Q. hartwissiana* |
| | | | G | *Q. petraea, Q. robur, Q. pubescens* |
| | | Est de l'Amérique du Nord | A | *Q. bicolor, Q. michauxii, Q. montana* |
| | | | B | *Q. stellata* |
| | | | C | *Q. lyrata* |
| | | | D | *Q. alba* |
| | | | E | *Q. macrocarpa* |
| | | Ouest de l'Amérique du Nord | A | *Q. sadleriana* |
| | | | B | *Q. douglasii* |
| | | | C | *Q. garryana, Q. lobata* |
| | | | D | *Q. gambelii* |
| | | | E | *Q. engelmanii* |
| | | Amérique Centrale | A | *Q. glabrescence, Q. polymorpha* |
| | | | B | *Q. reticulata* |
| | *Mesobalanus* | Europe et Asie de l'Ouest | A | *Q. dentata* |
| | | | B | *Q. pontica* |
| | | | C | *Q. frainetto, Q. macranthera, Q. pyrenaica* |
| | *Cerris* | Europe et Asie de l'Ouest | A | *Q. gilliana, Q. phillyreoides* |
| | | | B | *Q. baroni* |
| | | | C | *Q. alnifolia* |
| | | | D | *Q. coccifera* |
| | | | E | *Q. suber* |
| | | | F | *Q. libani, Q. castaneifolia, Q. ithaburensis* |
| | | | G | *Q. acutissima, Q. variabilis* |
| | | | H | *Q. cerris* |

D'après Camus (1936-1954) et Nixon (1993).

[a] La lettre indique une sous-section proposée par Camus (1936-1954).

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

## 2- Classification générale du genre *Quercus*

La définition, la phylogénie, la systématique et l'évolution des espèces de chênes sont encore très mal connues malgré leur importance sylvicole, écologique, économique et même sociale (Manos *et al.* 1999, Samuel *et al.* 1998). Le concept biologique d'espèce[10] (voir Arnold 1997) est en effet difficile à manipuler dans le genre *Quercus* car l'hybridation interspécifique y est très fréquente (Whittemore et Schaal 1991, Rushton 1993, Kleinschmit 1993, Bussoti et Grossoni 1998). En outre, l'existence d'un important polymorphisme intraspécifique qui se traduit par de nombreuses formes écologiques complique la situation et soulève d'importants problèmes de nomenclature. Une même espèce porte ainsi souvent plusieurs noms et quelquefois le même nom peut désigner des taxons différents. En comparant les traités les plus anciens aux travaux plus récents (voir Bussoti et Grossoni 1998), il apparaît clairement que la tendance générale est à une simplification de la taxinomie, à la réduction du nombre d'espèces et au déclassement de différentes entités au rang de sous-espèces. Le nombre d'espèces demeure toutefois assez imprécis car il subsiste de nombreuses contradictions entre la seule monographie du genre disponible (Camus 1936-1954), les contributions de Schwarz (1964-1993), Krüssmann (1986) et la revue de Nixon (1993). Ce dernier recommande de suivre largement la classification générale proposée par Camus (1936-1954) en raison de sa cohérence avec les analyses phylogénétiques les plus récentes basées sur le polymorphisme de restriction de l'ADN chloroplastique et sur les séquences des ITS de l'ADNr (voir Manos *et al.* 1999). Une synopsis récente est présentée dans le **Tableau 17**. Les groupements principaux (sous-genres et sections) sont surtout basés sur les caractères des fleurs et des fruits comme la longueur et la forme du style, la présence ou l'absence d'indumentum sur la paroi interne du gland (endocarpe) et la position des ovules avortés de la graine (les ovaires contiennent 6 ovules dont un seul se développe en graine). Certains auteurs envisagent d'élever le sous-genre *Cyclobalanopsis* cantonné aux zones subtropicales ou à climat tempéré chaud de l'Asie de l'Est et de la Malaisie au rang de genre (Nixon 1993). Le deuxième sous-genre (*Quercus* ou *Euquercus*) est distribué dans le monde entier. Il est divisé en 5 sections : *Lobatae* (les chênes rouges américains), *Protobalanus* (le petit groupe des chênes intermédiaires de l'ouest des Etats-Unis), *Quercus*, *Cerris* et *Mesobalanus* (les chênes blancs). Plusieurs d'entre elles (*Lobatae*, *Quercus* et *Cerris*) correspondent à des sous-genres dans la classification de Schwarz (1993). Cette dernière comporte même un quatrième sous-genre (*Sclerophyllodrys*) qui regroupe les deux sous-sections représentées par *Q. ilex* (section *Quercus*, C) et *Q. coccifera* (section *Cerris*, D) dans le tableau synoptique. L'individualisation de ce groupe a été très récemment confortée par des données anatomiques (Uzunova *et al.* 1997) et une seconde analyse de séquence des ITS de l'ADNr (Samuel *et al.* 1998) qui paraît toutefois partiellement en contradiction avec les résultats de Manos *et al.* (1999).

[10] Selon ce concept, les espèces sont des groupes de populations naturelles réellement ou potentiellement interfertiles qui sont isolés les uns des autres par des barrières reproductives.

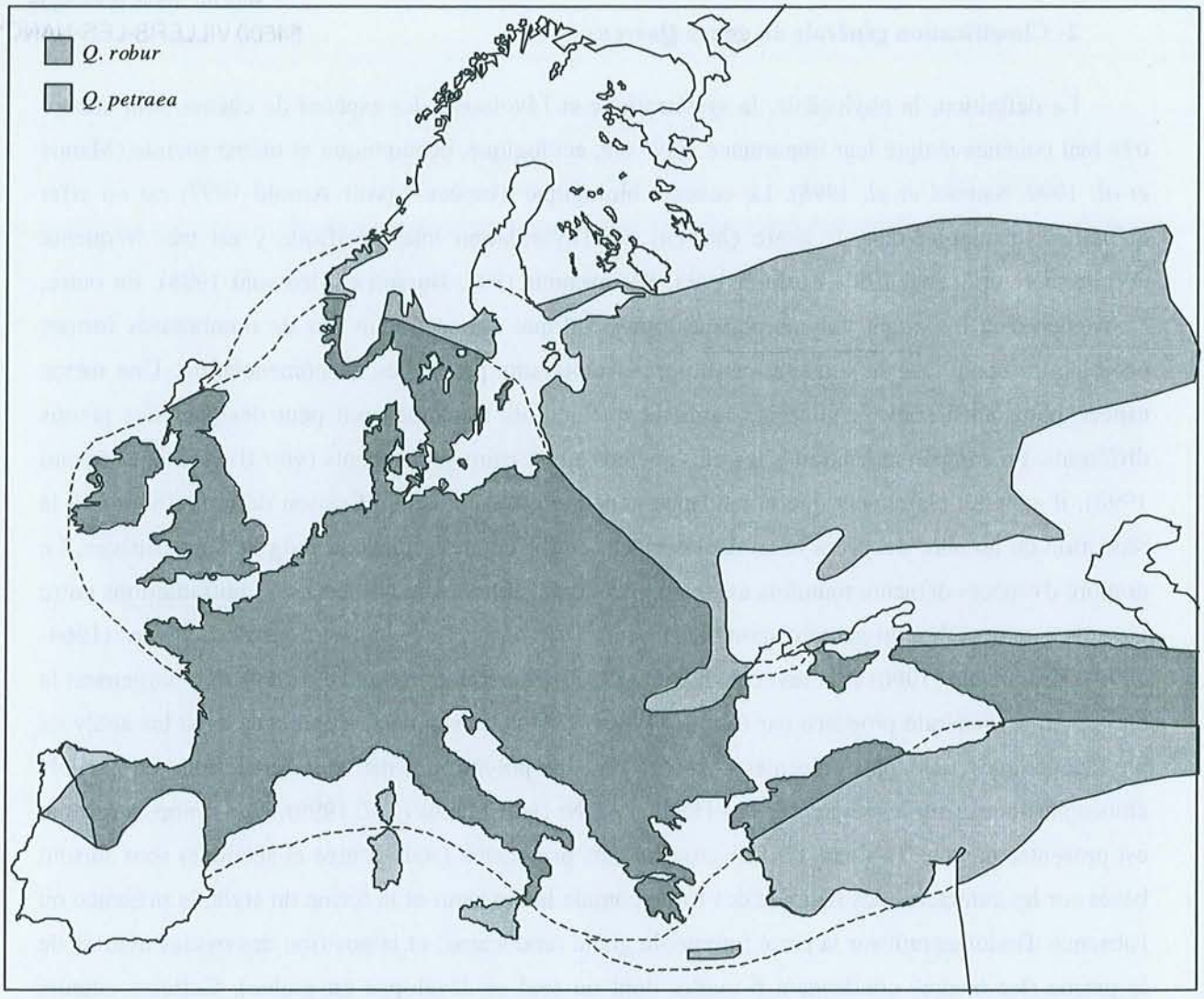



**Figure 22.** Aires de distribution de *Quercus robur* et *Quercus petraea.* D'après Kleinschmit (1993).

### 3- Les chênes pédonculé et sessile

Les chênes pédonculé (*Quercus robur* L = *Q. pedunculata* Ehrh) et sessile (*Quercus petraea* (Matt.) Liebl. = *Q. sessilis* Ehrh. = *Q. sessiliflora* Salish.) appartiennent à la section *Quercus* du sous-genre *Quercus* (**Tableau 17**). Ce sont deux espèces dominantes et sympatriques des forêts de feuillus de l'Europe tempérée (27 % de l'aire totale) dont l'importance économique est considérable. En France, elles représentent à elles seules plus de 30 % de la surface forestière et environ 15 % de la production courante de bois d'oeuvre (mètres cubes produits par an). L'aire de distribution du chêne sessile s'étend depuis le nord de l'Espagne jusqu'à la frontière russo-polonaise, de l'Ecosse vers les Balkans et la mer Caspienne et du sud de l'Italie au sud de la Suède et de la Norvège (**Figure 22**). Celle du chêne pédonculé s'étend au-delà de ces limites, plus à l'ouest vers le Portugal, plus au nord dans la péninsule Scandinave, plus à l'est jusqu'à l'Oural et plus au sud vers la Turquie. Les deux espèces constituent souvent des peuplements mélangés mais occupent généralement des niches écologiques distinctes (Becker et Lévy 1982, Grandjean et Sigaud 1987) : le chêne pédonculé a une

préférence marquée pour les milieux humides (bas de versant et sols alluviaux) tandis que le chêne sessile occupe plutôt les terrains à végétation xéro-acidophile à neutres mésophile (il tolère bien la sécheresse et l'acidité).

Des différences morphologiques (Dupouey 1983, Grandjean et Sigaud 1987, Dupouey et Badeau 1993, Aas 1993, Bacilieri *et al.* 1995, Kleinschmit *et al.* 1995) et anatomiques (Feuillat *et al.* 1997) permettent de distinguer les deux espèces malgré la forte variabilité intraspécifique, en particulier du chêne sessile. Les valeurs moyennes d'une trentaine de paramètres les plus discriminants portant sur différents aspects de la morphologie des rameaux de l'année, des feuilles et des fruits (Dupouey et Badeau 1993) et sur des caractéristiques anatomiques du bois (Feuillat *et al.* 1997) d'arbres adultes sont indiquées dans le **Tableau 18**. Certains caractères foliaires très discriminants comme la pilosité, la longueur du pétiole, l'angle de l'auricule à la base du limbe ou bien encore le nombre de lobes peuvent être utilisés efficacement pour la classification taxinomique de jeunes plants (Bacilieri *et al.* 1996).

Dans les stations mixtes cependant, des individus morphologiquement intermédiaires ou combinant les caractères-types des deux espèces sont systématiquement présents. Leur fréquence est très variable dans les différentes parties de l'aire naturelle de ces chênes. Relativement faible dans plusieurs populations françaises (5 %, Dupouey 1983, 3,7 %, Granjean et Sigaud 1987, 3,5 % Dupouey et Badeau 1993, 2,8 % Bacilieri *et al.* 1995), d'Allemagne et de Pologne (3 %, Aas 1993), elle atteindrait près de 50 % (dont 26 % d'arbres strictement intermédiaires) dans certaines populations du nord-ouest de l'Europe, particulièrement en Grande-Bretagne (revue par Rushton 1993). Ces individus inclassables qui produisent un pollen peu viable (Rushton 1978) sont souvent installés sur des terrains à végétation neutre mésophile, c'est-à-dire dans des conditions de sol moyennes par rapport aux exigences stationnelles respectives des deux espèces (Grandjean et Sigaud 1987). De nombreuses expériences de croisements contrôlés ont par ailleurs permis d'établir que les chênes pédonculé et sessile sont largement compatibles et peuvent donc produire des hybrides (Rushton 1977, Steinhoff 1993, 1998, Kleinschmit *et al.* 1995). Le taux de réussite de ces croisements est cependant très variable selon les combinaisons et plus important dans le sens *Q. robur* [femelle] x *Q. petraea* [mâle] que dans le sens inverse (41,3-50,0 % contre 11,1-24,5 %). L'existence de flux de gènes asymétriques dans les populations naturelles a aussi été démontrée par Bacilieri *et al.* (1996) dans une station mixte (417 arbres) du nord-ouest de la France. Dans cette contribution génétique (allozymes), les auteurs ont examiné 7 loci pour 3 cohortes de chênes sessiles et pédonculés constituées chacune d'arbres adultes, de leurs graines (qui représentent le pool pollinique ou d'ovules de l'année d'étude) ou de jeunes plants (3-6 ans) poussant à proximité (dans un rayon de 1 m). Pour plusieurs loci, les fréquences alléliques du pool pollinique "pédonculé" se sont avérées significativement différentes des fréquences estimées pour la cohorte de chênes pédonculés adultes et plus proches de celles de la cohorte de chênes sessiles adultes. Ce résultat indique clairement que le

chêne sessile contribue à la descendance du chêne pédonculé (17 % à 48 % selon la population de référence utilisée).

**Tableau 18.** Quelques paramètres morphologiques et anatomiques distinctifs de *Q. robur* et *Q. petraea*.

| Paramètre | *Q. robur* | *Q. petraea* |
|---|---|---|
| Rameau de l'année[a] | | |
| Longueur des poils tecteurs (mm) | 0,0 | 0,6 |
| Densité des poils tecteurs (%) | 0 | 9 |
| Feuille[a] | | |
| Longueur des poils tecteurs à la face inférieure du limbe (mm) | 0,3 | 1,7 |
| Densité des poils tecteurs à la face inférieure du limbe (%) | 4 | 47 |
| Longueur des poils tecteurs sur la nervure principale (mm) | 0,9 | 2,1 |
| Densité des poils tecteurs sur la nervure principale (%) | 8 | 45 |
| Nombre de nervures intercalaires | 4,9 | 0,5 |
| Longueur relative des nervures intercalaires (%) | 49 | 4 |
| Longueur du pétiole (mm) | 7 | 16 |
| Longueur des poils tecteurs sur le pétiole (mm) | 0,1 | 1,9 |
| Densité des poils tecteurs sur le pétiole (%) | 1 | 24 |
| Angle de l'auricule à la base du limbe (°) | 21 | 74 |
| Nombre de lobes | 10,2 | 13,2 |
| Position relative de la largeur maximum (%) | 57 | 54 |
| Pédoncule fructifère[a] | | |
| Longueur (mm) | 61 | 15 |
| Diamètre (mm) | 1,2 | 2,5 |
| Longueur des poils tecteurs (mm) | 0,4 | 1,7 |
| Densité des poils tecteurs (%) | 4 | 36 |
| Caractéristiques anatomiques du bois[b] | | |
| Surface occupée par le parenchyme dans un cerne (%) | 57 | 52 |
| Surface occupée par les fibres dans un cerne (%) | 29 | 36 |
| Bois précoce | | |
| Largeur (mm) | 0,8 | 0,72 |
| Nombre d'assises de vaisseaux | 2,4 | 2,0 |
| Nombre de vaisseaux par cerne (sur une largeur tangentielle de 1 mm) | 5,46 | 4,51 |
| Nombre de vaisseaux par unité de surface (mm²) d'un cerne | 2,6 | 2,2 |
| Surface occupée par les vaisseaux dans un cerne (%) | 14,4 | 11,5 |
| Surface occupée par les fibres (%) | 7,7 | 10,1 |
| Bois tardif | | |
| Largeur (mm) | 1,36 | 1,49 |
| Proportion (%) | 61 | 67 |
| Surface occupée par les fibres (%) | 42 | 49 |
| Aspect des groupes de vaisseaux | Large | Etroit |
| Surface occupée par les fibres dans un cerne (%) | 29 | 36 |
| Surface moyenne de la zone des fibres (mm²) | 0,47 | 0,67 |
| Surface occupée par le parenchyme (%) | 58 | 51 |

D'après Dupouey et Badeau (1993) et Feuillat *et al.* (1997).

[a] Valeurs moyennes significativement différentes au seuil de 5 % établies à partir de 80 populations de 8 forêts de la plaine de Lorraine (10 arbres dominants sélectionnés au hasard par population).

[b] Valeurs moyennes significativement différentes au seuil de 5 % basées sur l'analyse de 27 chênes pédonculés et 31 chênes sessiles provenant de 14 stations mixtes de la Forêt de Cîteaux.

La tendance est la même pour les jeunes régénérations de 3 à 6 ans indiquant que les graines hybrides ne sont pas éliminées par la sélection naturelle dans les premières phases de leur établissement. L'étude montre aussi l'existence de flux de gènes directionnels au niveau intraspécifique : pour le locus AAP-A les fréquences alléliques du pool pollinique "sessile" apparaissent en effet supérieures à celles, déjà élevées, qui caractérisent la cohorte des chênes sessiles adultes. Toutes ces observations confortent la théorie de l'hybridation introgressive selon laquelle, après fécondation préférentielle du chêne pédonculé par le pollen du chêne sessile, ce sont essentiellement des back-cross avec le chêne sessile qui se produiraient. Les auteurs suggèrent enfin que ces flux de gènes directionnels intra- et interspécifiques constituent une caractéristique génétique qui renforce considérablement la dynamique de succession de ces deux espèces. Le chêne pédonculé, colonisateur des milieux ouverts, est en effet progressivement remplacé par le chêne sessile qui s'adapte mieux aux conditions de croissance en peuplement dense (Rameau 1990).

**Tableau 19.** Caractéristiques du génome de *Q. robur* et *Q. petraea.*

| Origine des arbres étudiés | Taille du génome (pg/2C ± SE)[a] | Proportion de GC (%) | Nombre, taille des chromosomes et/ou formule chromosomique[b] | Référence |
|---|---|---|---|---|
| *Quercus robur* | | | | |
| Grosshansdorf (Allemagne) | 1,59 ± 0,04 (Fe) | nd | 10m+2sm$^{cs}$+10sm+2st+1B (2,49-6,26 µm) | Ohri et Ahuja (1990) |
| Lorraine (nord-est de la France) | 1,84 ± 0,01 (FC) | 42,0 | nd | Favre et Brown (1996) |
| Foisnard (Ile de France) | 1,88 ± 0,01 (FC) | 38,9 | 24 | Zoldos *et al.* (1998) |
| Orsay (Ile de France) | 1,88 ± 0,02 (FC) | 38,8 | 24 | |
| Cesma (nord de la croatie) | 1,88 ± 0,01 (FC) | 39,8 | 24 | |
| Pozega (nord-est de la croatie) | 1,88 ± 0,02 (FC) | 40,3 | 24 | |
| *Quercus petraea* | | | | |
| Grosshansdorf (Allemagne) | 1,58 ± 0,03 (Fe) | nd | 10m+2sm$^{cs}$+10sm+2st+1B (2,49-6,26 µm) | Ohri et Ahuja (1990) |
| Lorraine (nord-est de la France) | 1,87 ± 0,01 (FC) | 41,7 | nd | Favre et Brown (1996) |
| Dourdan (Ile de France) | 1,88-2,00 ± 0,03 (FC) | 40,0 | 24 + 1B, 2B ou 3B | Zoldos *et al.* (1998) |
| Djakovo (nord-est de la Croatie) | 1,88 ± 0,02 (FC) | 39,5 | 24 | |

nd : non déterminé.

[a] SE : erreur standard sur la moyenne ; Fe : microdensitométrie par la technique de Feulgen ; FC : Cytométrie de flux.

[b] m : métacentrique ; sm : submétacentrique ; cs : constriction secondaire ; st : subtélomérique ; B : chromosome surnuméraire.

C'est probablement l'existence de ces importants flux interspécifiques de gènes qui expliquent la faible différenciation génétique des chênes pédonculé et sessile. La taille du génome, le pourcentage de GC et les formules chromosomiques (nombre, taille et morphologie des chromosomes) sont identiques (Ohri et Ahuja 1990, Favre et Brown 1996, Besendorfer *et al.* 1997,

Zoldos *et al.* 1998, 1999, **Tableau 19**) malgré quelques variations intraspécifiques (Zoldos *et al.* 1998) provoquées par la présence occasionnelle de chromosomes surnuméraires considérés soit comme des chromosomes B (Ohri et Ahuja 1990) soit comme des loci NOR détachés (Besendorfer *et al.* 1997). Aucun marqueur moléculaire diagnostique, cytoplasmique ou nucléaire, de l'une ou l'autre espèce à l'échelle de l'aire de distribution n'a pu être détecté jusqu'à présent. La différenciation est totalement inexistante dans le cas des génomes cytoplasmiques à transmission maternelle exclusive (Dumolin-Lapègue *et al.* 1998). La structuration géographique des variations de l'ADN chloroplastique ou mitochondrial est en effet similaire pour les deux espèces qui partagent donc localement (dans les stations mixtes en particulier) les mêmes cytotypes (Kremer *et al.* 1991, Petit *et al.* 1993a, 1997, Ferris *et al.* 1993, Dumolin-Lapègue *et al.* 1997, 1998), y compris avec d'autres espèces européennes de la même section (*Q. pubescens* et *Q. faginea*) ou de la section *Mesobalanus* (*Q. frainetto*, *Q. macranthera* et *Q. pyrenaica*). Aucune différenciation n'a pu être non plus mise en évidence à l'aide de marqueurs nucléaires. La différenciation allozymique entre les deux espèces est très faible (3,0-5,0 %) et semblable à la différenciation intraspécifique entre les populations (Kremer *et al.* 1991, Kremer et Petit 1993, Müller-Starck *et al.* 1993, Zanetto *et al.* 1994). De même, la comparaison des variants de longueur de l'ADNr dans une station mixte n'a révélé aucune différence significative malgré une très forte diversité moléculaire (Petit *et al.* 1993b). Au niveau des marqueurs RAPD enfin, moins de 2 % des fragments générés ont des fréquences significativement différentes entre les deux espèces (Moreau *et al.* 1994, Bodénès *et al.* 1997a). L'étude des régions génomiques qui correspondent aux plus informatifs de ces fragments (SCAR : Sequence Characterized Amplified Region) a seulement permis de montrer que la différenciation générale (à l'échelle de l'aire de répartition) est en moyenne plus importante que la différenciation locale (Bodénès *et al.* 1997b). Cette dernière semble augmenter selon un gradient latitudinal contrairement à ce qui a été observé à l'aide des marqueurs isoenzymatique (gradient longitudinal, Zanetto *et al.* 1994). Sur la base de ces marqueurs SCAR, les populations de Basse Saxe et de Pologne apparaissent en effet nettement plus différenciées que les populations de la péninsule Ibérique et de Roumanie. Un cas de forte différenciation de populations autrichiennes pour un locus allozymique et 11 fragments RAPD (*a priori* spécifiques) a par ailleurs été rapporté très récemment (Samuel 1999).

## II- Caractérisation et variation des unités d'ADNr 5S chez les chênes pédonculé et sessile

Les études génétiques déjà menées sur *Q. robur* et *Q. petraea* à l'aide de marqueurs nucléaires montrent que la différenciation interspécifique est globalement assez faible mais variable selon les populations (**paragraphe précédent**). Nous avons donc abordé l'étude de la diversité moléculaire des unités d'ADNr 5S sur un échantillon de 57 chênes pédonculés (19 provenances) et 36 chênes sessiles (14 provenances) répartis du sud-ouest au nord-est de l'Europe. Ces arbres sont majoritairement issus des 11 paires géographiques de populations utilisées par Bodénès (1996) pour la recherche de marqueurs RAPD discriminants entre les deux espèces. Après amplification des espaceurs à l'aide d'amorces universelles (voir la **Partie III** de la **présentation générale**), un premier tri des fragments obtenus s'est révélé nécessaire en raison de la présence de nombreux artefacts de PCR (multimères, hétéroduplexes) liés à la structure répétée en tandem de l'ADNr 5S. La caractérisation des unités a ensuite été réalisée en utilisant une sonde homologue et des amorces spécifiques développées à partir des fragments amplifiés. Nous avons alors pu accéder au nombre d'unités et explorer leur organisation.

Les variations des unités d'ADNr 5S ont tout d'abord été examinées sur l'ensemble du dispositif géographique en utilisant différents couples d'amorces (permettant d'amplifier soit l'espaceur soit l'unité entière) et différentes méthodes d'électrophorèse (agarose et acrylamide). L'existence de fragments spécifiques (ou à fréquence spécifique) a été recherchée et une estimation de la diversité phénotypique des espaceurs a été obtenue pour chacune des deux espèces. Les classes majeures de variants électrophorétiques ont alors été systématiquement clonées à partir d'un échantillon réduit d'individus (10 par espèce) et un séquençage aléatoire a été réalisé au sein de chacune d'entre elles. Ces informations de séquence ont tout d'abord été utilisées pour accéder à la topographie des variations et à l'évolution structurale des unités. Puis, nous avons estimé le polymorphisme (nombre de sites variables dans l'échantillon) et la variation nucléotidique moyenne (mesure équivalente à la diversité nucléotidique) aux niveaux interspécifique, intraspécifique et intragénomique. Enfin, nous avons réalisé une première tentative d'ordination phylogénétique des clones d'espaceurs séquencés chez *Q. robur* et *Q. petraea* par construction d'un arbre cladistique enraciné à l'aide d'un groupe externe de séquences provenant de deux espèces de la section voisine *Cerris* (*Q. suber* et *Q. libani*).

Tous ces résultats sont présentés et discutés dans l'article qui suit, en phase de soumission.

**Characterization and variation of nuclear 5S ribosomal DNA units in two closely related European oaks, *Quercus robur* L. (pedunculate oak) and *Quercus petraea* (Matt.) Liebl. (sessile oak)**

Jean-François Trontin, Anne-Marie Collignon, Marie-France Arnould, and Jean-Michel Favre

## 1- Abstract

The molecular diversity and evolution of repeat units at the nuclear 5S ribosomal DNA locus (5S rDNA) were investigated on the whole distribution area (11 geographic provenances) of the two closely related, sympatric and interfertile pedunculate and sessile oak. Excluding a putative pseudogene, the gene region is highly conserved in both length (120 bp) and sequence within and between the two species (interindividual nucleotide diversity: 0.003). Conversely, appreciable intraspecific polymorphism was discovered in the spacer region of both pedunculate (197-233 bp in length) and sessile oak (217-233 bp in length). This polymorphism consisted mainly into sequence microheterogeneity and as many as 17 variable phenotypes were recognized (12 per species) on non-denaturing polyacrylamide gels. Based on calculations of sample relative phenotypic diversity, sessile oak appeared to be more variable than pedunculate oak (0.899 vs 0.779). However, the mean sequence differences among spacers is similar for the two species (0.023). Remarkably, this intraspecific value (i.e. interindividual + intraindividual) is not significantly higher than the intraindividual mean estimates (0.020) suggesting that some spacer selection may operate at the species level. Alternatively, such a small difference may simply reflect a slow rate of concerted evolution under the process of molecular drive in pedunculate and sessile oak. Despite relatively high level of intraspecific molecular diversity, no qualitative difference was discovered between the two species at the 5S rDNA locus and this is discussed in relation to extensive interspecific hybridization. The phylogenetic gene tree reconstruction based on spacer sequences and rooted with the congeneric outgroup species *Q. suber* and *Q. libani* is consistent with recent and interconnected evolutionary histories of pedunculate and sessile oak.

***Key words:*** evolution, molecular diversity, multigene family, ribosomal DNA, *Quercus* genome

## 2- Introduction

Nuclear 5S rDNA refers to a middle-repetitive gene family encoding the small 5S RNA component of the large subunit of eukaryotic ribosomes. In higher plants as in most higher eukaryotes (Lapitan 1992), hundreds to thousands of tandemly repeated 5S rDNA units occur at one or several chromosomal loci generally separate from the major 18S-5.8S-26S rDNA tandem arrays (NOR loci). Each repeated unit (ca. 200–900 bp) consists of a 5S rRNA gene transcribed by RNA polymerase III (Pol III), and a nontranscribed, intergenic spacer (NTS) which contains several hallmarks probably involved in transcriptional regulation (Korn 1982, Venkateswarlu *et al.* 1991).

Because repetitive sequences are considered to have "a high turnover" during evolution (Flavell *et al.* 1983), the 5S rDNA variations have been extensively studied in plants with a particular focus on inter- and intraspecific patterns of unit conservation and divergence. Most 5S rDNA unit polymorphisms were naturally discovered across species. Comparatively, intragenomic 5S rDNA diversity (i.e. intraindividual) is generally considered to be low and interpreted as resulting from intralocus concerted evolution of repeats (Arnheim 1983). However, the concerted homogenizing forces are not always sufficient to balance out the molecular processes responsible for mutation, thus allowing significant sequence heterogeneity among repeats within individual arrays as observed in several plant groups (Cronn *et al.* 1996, Kellogg and Appels 1995, Schneeberger and Cullis 1992). In addition, tandem arrays located on non-homologous chromosomes have been shown to evolve nearly independently (e.g. allopolyploid *Gossypium* species (Cronn *et al.* 1996) so that divergent unit length classes are frequently detected in a single genome (e.g. members of the *Triticeae* (Kellogg and Appels 1995, Scoles *et al.* 1988). Recent attempts to estimate the molecular diversity of such paralogous sequences within *Hordeum* species (Baum and Johnson 1998) also revealed that the intragenomic 5S rDNA heterogeneity can be further amplified by the variations between individuals or infraspecific groupings.

Although intragenomic 5S rDNA variations, especially mutations within array, appear to accumulate in equal proportion in genes and spacers (Cronn *et al.* 1996, Kellogg and Appels 1995), fixation of these differences is almost restricted to the spacer region. As a result, the 5S rRNA genes and their products are highly conserved in the plant kingdom, both with respect to length (ca. 120 bp) and nucleotide sequence (consult the 5S RNA Data Bank, Szymanski *et al.* 1999), whereas the intergenic spacer evolves more rapidly. The spacer variations proved to be useful for phylogenetic reconstruction of species relationships in important plant groups such as the wheat tribe *Triticeae* (Baum and Appels 1992, Dvorák *et al.* 1989, Kellogg and Appels 1995, Scoles *et al.* 1988), or the genera *Oryza* (McIntyre *et al.* 1992) and *Gossypium* (Cronn *et al.* 1996). Several authors also claimed that the spacer region is likely to provide useful polymorphisms at lower taxonomic ranks, allowing discrimination of closely related plant species, cultivars or lines (Ko and Henry 1996, Kolchinsky *et al.* 1991, Röder *et al.* 1992).

Most of these comprehensive studies of 5S rDNA diversity and evolution were however essentially concerned with short-lived species, usually domesticated annuals or perennial herbs. Comparatively, little is known about the 5S rDNA variations in long-lived woody plants such as forest trees. The 5S rDNA repeats have only been well-characterized (i.e. including sequence data) in a number of broad-leaved trees such as Eucalypts (Udovicic *et al.* 1995), *Acacia* (Playford *et al.* 1992), *Alnus incana* and *Betula papyrifera* (Johnson *et al.* 1992), and in the conifers *Larix decidua* (Trontin *et al.* 1999), *Pseudotsuga menziesii* (Amarasinghe and Carlson 1998), *Picea glauca* (Brown and Carlson 1997) and *Pinus radiata* (Gorman *et al.* 1992, Moran *et al.* 1992). As in non-woody plants the spacer variations appeared to be informative in elucidating species relationships (Moran *et al.* 1992, Playford *et al.* 1992, Udovicic *et al.* 1995).

An additional complexity is that these generally widespread woody plants are known to maintain high level of within species variation compared to non-woody plants (Hamrick *et al.* 1992). To date, the 5S rDNA variations between individuals have been poorly investigated, using too small sample size (2-4 trees, Udovicic *et al.* 1995) or screening spacer size variations only (Moran *et al.* 1992, Trontin *et al.* 1999). Nevertheless, appreciable polymorphisms among units from a single individual have already been detected (e.g. *Acacia* species, *Betula papyrifera*, *Larix decidua*, *Pinus radiata*). This is particularly exemplified by the situation in the larch genome where two highly divergent unit size classes were found to be organized in composite tandem array (Trontin *et al.* 1999).

Paying too little attention to intraspecific diversity may be a source of multiple pitfalls in phylogenetic and taxonomic studies. Unidentified relationships between divergent paralogous sequences may indeed cause muddled and inaccurate phylogenies if a gene tree is interpreted as a species tree (Buckler *et al.* 1997). In-depth intraspecific studies of molecular diversity (within and among individuals) are thus required to discriminate between divergent paralogues prior to such phylogenetic or taxonomic studies (Baum and Johnson 1998).

In this context, we focused on the large genus *Quercus* (the oaks; *Fagaceae*), which includes some 300 to 600 species (depending on the authors) spread over the northern hemisphere from subtropical humid and temperate forests to semi-arid areas. We initiated a comprehensive study of the molecular diversity of 5S rDNA units in two important and widespread white oaks (subgenus *Quercus*, section *Quercus*, Camus 1954) in Europe, *Quercus robur* L. (pedunculate oak) and *Quercus petraea* (Matt.) Liebl. (sessile oak). These two closely related species are broadly sympatric and nearly completely out-crossing. Asymmetric gene flow has been recorded in natural mixed stands (Bacilieri *et al.* 1996) and, on the basis of the success of interspecific controlled crosses, Kleinschmit *et al.* (1995), and more recently Steinhoff (1998) proposed to form only one group, *Quercus robur*, with differentiated subspecies *robur* and *petraea*. They have different ecological preferences (Grandjean and Sigaud 1987) and can be efficiently discriminated by morphological characteristics

and wood anatomy when multivariate approaches are used (Dupouey and Badeau 1993, Feuillat *et al.* 1997). In contrast, molecular differentiation is low. The two species have the same DNA content (Favre and Brown 1996), similar chromosome number and structure (Ohri and Ahuja 1990), they locally share the same cytoplasmic genome (Dumolin-Lapègue *et al.* 1998) and studies conducted on the nuclear genome using allozymes (Kremer *et al.* 1991, Zanetto *et al.* 1994), 18S-5.8S-26S rDNA (Petit *et al.* 1993b) or RAPD (Random Amplified Polymorphic DNA) markers (Moreau *et al.* 1994, Bodénès *et al.* 1997a) lacked to reveal any species-specific markers.

The structural organization of 5S rDNA is unknown in the genus *Quercus* and in the *Fagaceae* as well. In this contribution, we first characterize the 5S rDNA units in pedunculate and sessile oak and then describe the gene and spacer length and sequence variations between individuals from various natural populations on their whole distribution area.

## 3- Materials and Methods

### a- Plant material and DNA isolation

57 pedunculate and 36 sessile oaks were sampled throughout the major part of their European natural range, from Spain to Poland (**Figure 23**). 50 trees (23 pedunculate and 27 sessile oaks) belonged to 22 populations clustered in 11 geographic species pairs (circled and numbered) that have been used for the identification of discriminant markers between the two species (Bodénès *et al.* 1997a). Such a sampling scheme was designed to discriminate between local and general differentiation The provenances of these 93 trees are given in **Table 20**.

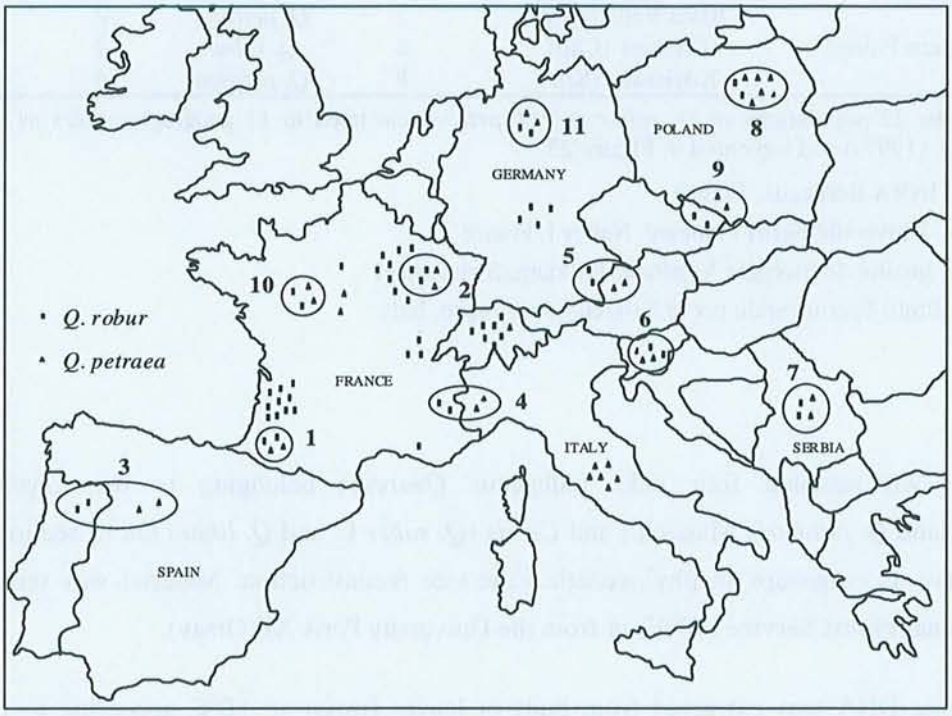



**Figure 23**. Geographical distribution of sampled oaks within Europe. Referenced populations are circled and numbered (see **Table 20**).

**Table 20.** Provenance of sampled oaks and source of materials.

| Region | Locality (abbreviation) | Population reference* | Species | Sample size | Source |
|---|---|---|---|---|---|
| Northern portugal | Viana do Minho (Vim) | 3 | *Q. robur* | 2 | *a* |
| Northern Spain | Montajo (Mon) | 3 | *Q. petraea* | 2 | *a* |
| Southwestern France | Mont de Marsan (MoM) | 1 | *Q. robur* | 2 | *a* |
| | Oloron (Olo) | 1 | *Q. petraea* | 2 | *a* |
| | Pierroton (Pie) | / | *Q. robur* | 6 | *a* |
| | Arcachon (Arc) | / | *Q. robur* | 4 | *a* |
| Centre of France | Petite Charnie (PeC) | 10 | *Q. robur* | 2 | *a* |
| | Petite Charnie (PeC) | 10 | *Q. petraea* | 2 | *a* |
| | Orsay (Ors) | / | *Q. robur* | 1 | *b* |
| | Orléans (Orl) | / | *Q. petraea* | 2 | *a* |
| Southern France | Valréas (Val) | / | *Q. robur* | 1 | *b* |
| Southern Alps | La Fajiana (Faj) | 4 | *Q. robur* | 2 | *a* |
| | Nt Dame du Cruet (DaC) | 4 | *Q. petraea* | 2 | *a* |
| Eastern France | Fontain (Fon) | / | *Q. robur* | 3 | *b* |
| Northeastern France | Lisle (Lis) | 2 | *Q. robur* | 3 | *a* |
| | Temple (Tem) | 2 | *Q. petraea* | 2 | *a* |
| | Champenoux (Cha) | / | *Q. robur* | 8 | *b* |
| Switzerland | Fribourg (Fri) | / | *Q. robur* | 7 | *c* |
| | Fribourg (Fri) | / | *Q. petraea* | 1 | *c* |
| Centre of Germany | Escherode (Esc) | 11 | *Q. robur* | 2 | *a* |
| | Escherode (Esc) | 11 | *Q. petraea* | 2 | *a* |
| | Geisenheim (Gei) | / | *Q. robur* | 4 | *c* |
| Italy | Sargiano (Sar) | / | *Q. petraea* | 6 | *d* |
| Bavaria | Stamser (Sta) | 5 | *Q. robur* | 2 | *a* |
| Austria | Ridenburg (Rid) | 5 | *Q. petraea* | 2 | *a* |
| Slovenia | Babjilozic (Bab) | 6 | *Q. robur* | 2 | *a* |
| | Kobilje (Kob) | 6 | *Q. petraea* | 3 | *a* |
| Southern Poland | Jaszkowka (Jas) | 9 | *Q. robur* | 2 | *a* |
| | Jaczowicze (Jac) | 9 | *Q. petraea* | 2 | *a* |
| Serbia | Sremska (Sre) | 7 | *Q. robur* | 2 | *a* |
| | Iriski Venac (IrV) | 7 | *Q. petraea* | 2 | *a* |
| Northeastern Poland | Chonow (Cho) | 8 | *Q. robur* | 2 | *a* |
| | Kozienice (Koz) | 8 | *Q. petraea* | 6 | *a* |

* Refers to the 22 populations of *Q. robur* and *Q. petraea* clustered in 11 geographic pairs as described in Bodénès *et al.* (1997a) and indicated in **Figure 23**.

[a] A. Kremer. INRA Bordeaux, France

[b] J.-M. Favre. Université Henri Poincaré, Nancy I, France

[c] C. Monney. Institut de Biologie Végétale, Fribourg, Switzerland

[d] F. Ducci. Istituto Sperimentale per la Selvicultura, Arezzo, Italy

In addition, we sampled four oaks (subgenus *Quercus*) belonging to the *Erythrobalanus* (*Q. rubra* L. and *Q. palustris* Muench.) and *Cerris* (*Q. suber* L. and *Q. libani* Oliv.) sections (Camus 1954) to serve as outgroups in phylogenetic gene tree reconstruction. Material was sent from the French National Forest Service (ONF) or from the University Paris XI (Orsay).

Genomic DNA was extracted from buds or leaves frozen at -80°C according to Doyle and Doyle (1987) with minor modifications (Trontin *et al.* 1999).

b- 5S rDNA amplification and electrophoresis

Spacers were selectively amplified using two universal plant primers P1 and P2 based on a comparison of nuclear 5S rRNA gene sequences (Szymanski *et al.* 1999). The whole 5S rDNA unit was amplified using two additional primers (R3 and D4) deduced from the spacer sequences previously amplified using P1 and P2. The location and sequence of these primers are shown in **Figure 24**. The restriction sites that have been introduced in the primers for cloning and sequencing purposes have no detectable effects on the banding pattern (data not shown).

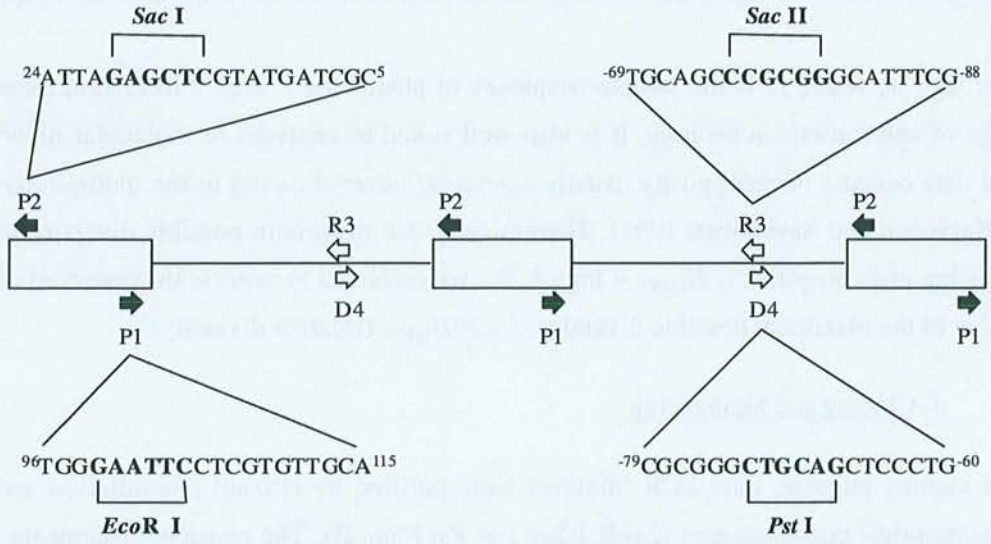



**Figure 24**. Location and sequence of the primers used to amplify the nuclear 5S rDNA in *Quercus robur* and *Quercus petraea* (P1, P2, R3, D4). The 120 bp gene region is shown as a box and the spacer as a line. The position (in bp) of each oriented primer (arrow) relative to the gene region is indicated at both 5' and 3' ends. The restriction sites *Sac* I, *Sac* II and *Eco*R I were introduced in the primers. *Pst* I has a unique, natural site in the *Quercus* spacers.

Amplifications were carried out in a 25-µl reaction volume containing 0.1-1 ng of template DNA, 20 mM Tris-HCl (pH 8.4), 2 mM $MgCl_2$, 200 µM of each dNTP, 0.2 µM of primer combinations P1/P2 or R3/D4, and 0.6 U of *Taq* DNA polymerase (Gibco BRL). Following incubation at 96°C for 3 min, 30 amplification cycles were performed in a PTC-100 thermal cycler (MJ Research Inc). Each cycle consisted of denaturation at 94°C for 40 s, annealing at 53°C for 1 min and extension at 72°C for 40 s.

10 µl of PCR products were mixed with 1 µl loading buffer [50% glycerol, 0.4% (w/v) bromophenol blue, 0.4% (w/v) xylene cyanol] and checked in standard 1.6 % agarose gels stained with ethidium bromide. For very sensitive detection of 5S rDNA polymorphisms, electrophoresis was also carried out for 6 h at constant voltage of 600 V (2500 V.h) through 0.75-mm thick, 30-cm long, non-denaturing polyacrylamide gels (8% T, 3.3% C) made up in 1x TBE (90 mM Tris-borate, 2 mM

EDTA pH 8.0) using the SequiGen apparatus (Biorad). Only a small amount of samples (2 µl) were applied onto these gels to avoid overloading and subsequent loss of resolution. At the indicated separation time (2500 V.h), DNA fragments of about 280 bp had migrated for 20–25 cm and were silver-stained (Bassam *et al.* 1991). A fine gel to gel match was realized by using the molecular weight standards DNA VI, ϕX174 or 100 bp DNA ladder (Gibco BRL).

c- Sample relative diversity

Sample 5S rDNA diversity was estimated from the amplified phenotypes revealed on acrylamide gels after silver staining. We used the Shannon's index (Lewontin 1972) to quantify the diversity,

$H = - \sum pi \log_2 pi$, where $pi$ is the sample frequency of phenotype $i$. This information measure has a wide range of applications in ecology. It is also well suited to analyses of molecular diversity using ribosomal data because heterozygosity usually cannot be inferred owing to the multiplicity of rDNA repeats (Karvonen and Savolainen 1993). Theoretically, the maximum possible diversity for a set of data consisting of $k$ categories is $H_{max} = \log_2 k$. So, we preferred to express the observed diversity as a proportion of the maximum possible diversity, $J = H/H_{max}$ (relative diversity).

d- Cloning and Sequencing

For cloning purpose, total PCR mixtures were purified by ethanol precipitation and digested with the appropriate endonucleases (*Eco*R I/*Sac* I or *Pst* I/*Sac* II). The restricted fragments were then recovered from agarose gel using the Jetsorb matrix (Genomed), ligated into pBluescript KS +/- (Stratagene), and introduced into DH5α line cells of *E. coli* (Gibco BRL) by transformation using the calcium chloride protocol (Sambrook *et al.* 1989). Prior to purification (Qiagen kits), recombinant plasmid clones were checked for insert size by PCR using the P1/P2 or R3/D4 primer combinations (reducing the annealing temperature to 50°C).

Sanger's dideoxy chain-terminating sequencing reactions were performed using dye-terminator chemistry (protocol from Applied Biosystems) and analyzed on an ABI 373 automated sequencer (Genome express SA, Grenoble, France).

e- Sequence analysis

Alignments were easily optimized by manual adjustment using SeqApp V 1.9 (Don Gilbert, Indiana University). To quantify the extent of sequence variation detected within individual, species or genome group, we used two descriptors: polymorphism and mean difference. Polymorphism is defined as the proportion of nucleotide positions that are variable. Mean difference is the average number of nucleotide differences per site between two sequences and this measure is called nucleotide diversity ($\pi$) when sequences are randomly chosen in the population. Values were calculated following Nei (1987).

Gene trees of aligned 5S rDNA sequences were generated using PAUP (Swofford 1993). Data sets were analyzed using the general heuristic algorithm with no character state weighting. Gaps were coded either as one single event (when evidently recognized as deletion or insertion) or as missing data. This treatment assures that a long gap will not be overweighted with one event per gapped site.

f- Southern hybridization and copy number estimation

Standard genomic DNA preparations were digested with restriction endonuclease and electrophoresed through 1.6 % agarose gels at 4 $V.cm^{-1}$ for 6 h. Capillary transfer of restricted DNA to positively charged nylon membrane (Euroblot) was performed for 12 h in 0.4 N NaOH (Sambrook *et al.* 1989). One 5S rDNA spacer clone (S*Qro*G102) amplified from *Q. robur* (tree G102) using the primer combination P1/P2 was excised from the vector (pBS*Qro*G102), gel purified and digoxigenin-labeled (DIG DNA labeling and detection kit, Boehringer Manheim). Hybridization with this probe (named pS*Qro*G102) was carried out at 42°C overnight in standard formamide buffer [5x SSC (0.75 M NaCl, 0.075 M sodium citrate), 50% formamide (v/v), 0.1% N-lauroylsarcosine (w/v), 0.02% sodium dodecyl sulfate (w/v), 2% blocking reagent (w/v)]. Filters were subsequently washed twice for 5 min at room temperature in 2x SSC, 0.01% SDS, and twice for 15 min at 65°C in 0.1x SSC, 0.01% SDS. Immunodetection and visualization with the colorimetric substrates NBT/X-phosphate were performed according to the manufacturer's instructions.

To estimate the 5S rDNA copy number in *Q. robur*, pure genomic DNA was extracted using the DNeasy plant mini kit (Qiagen). Precise DNA concentrations were measured by UV absorption (Genequant, Pharmacia) and fluorometry (TKO 100, Hoefer) using calf thymus DNA as internal standard (Sigma) since its base composition is similar to that reported for oak genomic DNA (42% GC, Favre and Brown 1996). Aliquots of genomic DNA (0.5 and 1.0 µg) were digested to completion with *Hae* III (this enzyme has a single recognition site within the spacer region), then electrophoresed together with dilutions of linearized vector pBS*Qro*G102 corresponding to the expected weights of 500, 1000, 2000, 4000, 6000, 8000, and 10000 copies and transferred on a nylon membrane as described above. Two filter replicates were then hybridized with the pS*Qro*G102 probe and subsequently analyzed by microdensitometry.

**4- Results**

a- Characterization of 5S rDNA repeat in *Quercus robur* and *Quercus petraea*

PCR amplification using the universal plant primers P1 and P2 produced a major fragment of approximately 280 bp and lesser amounts of 630 and 690 bp fragments (visualized on agarose gels) in virtually all trees of both species. Re-amplification of the 630 and 690 bp products yielded the 280 bp fragment, consistent with these being PCR artifacts.

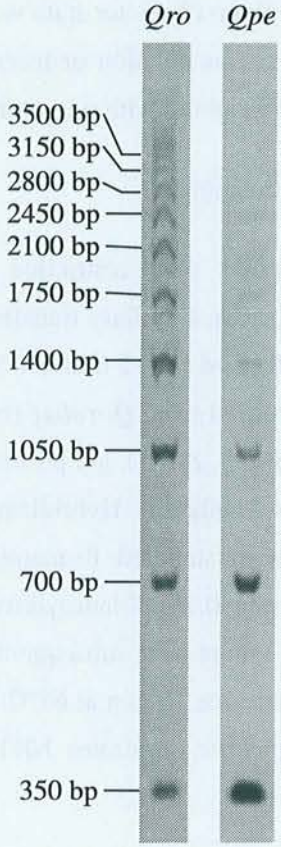



**Figure 25**. Restriction analysis of *Q. robur* (*Qro*) and *Q. petraea* (*Qpe*) genomic DNA (4 µg) digested with *Sca* I (5 U) and hybridized with the $p_{SQroG102}$ probe.

```
                                                                    Box A
Consensus          GgGTGcGATC ATACCAACAC TAATGCACCG GATCCCATCA GAACTCCGCA GTTAAGCGTG   60

GQroDF159 (Fon)    .T........ .......... .......... .......... .......... ..........
GQro3P (Pie)       .......... .......... .......... .......... .......... ..........
GQroA4 (Arc)       .......... .......... .......... .......... .......... ..........
GQpe390096-3 (Iri) .......... .......... .......... .......... .......... ..........
GQpeQS9 (Orl)      .....T.... .......... .......... .......... .......... ..........
GQpe96-3 (Tem)     .......... .......... .......... .......... .......... ..........

                                                Box C
Consensus          TTTGGGCGAG AGTAGTACTA GGATGgGTGA CCTCCTGGGA AGTCCtCGTG TTGCACCCcT  120

GQroDF159 (Fon)    .......... .......... .....C.... .......... .....C.... .........-
GQro3P (Pie)       .......... .......... .......... .......... .......... ..........
GQroA4 (Arc)       .......... .......... .......... .......... .......... ..........
GQpe390096-3 (Iri) .......... .......... .......... .......... .......... ..........
GQpeQS9 (Orl)      .......... .......... .......... .......... .......... ..........
GQpe96-3 (Tem)     .......... .......... .......... .......... .......... ..........
```

**Figure 26**. Sequence comparison of 5S genes isolated from *Q. robur* (*Qro*) and *Q. petraea* individuals (*Qpe*). Polymorphisms are indicated by standard IUPAC symbols. The provenance of each tree (DF159, 3P, A4, 390096-3, QS9 and 96-3) is signified in parentheses and can be recovered from **Table 20**. The putative binding sites of transcription factor IIIA to either the 5S gene (box A and box C) or the 5S rRNA gene product (underlined) are indicated in the majority-rule consensus sequence. EMBL accession numbers: AJ245104 (G*Qro*DF159), AJ245106 (G*Qro* 3P), AJ242948 (G*Qro*A4), AJ242949 (G*Qpe*390096-3), AJ242950 (G*Qpe*96-3), AJ245105 (G*Qpe*QS9).

A more precise investigation of these PCR products was carried out by restriction site analysis (data not shown) using endonucleases that either are known to have a recognition site within the conserved 120 bp coding region of plants (*Bam*H I, *Bst*E II, *Msp* I, *Sca* I, and *Tru*9 I) or were found to digest the 280 bp fragment (*Alu* I, *Bam*H I, *Hae* III, *Msp* I, *Pst* I, *Taq* I, and *Tru*9 I). This latter fragment was clearly identified as a monomeric PCR-product of a putative 5S rDNA spacer class approximately 230 bp long in both species. The remaining 630 bp and "690 bp" fragments respectively corresponded to dimer (i.e. homoduplexe consisting of two spacers separated by the 120 bp coding region) and "heterodimer" products (i.e. heteroduplexe between one 630 bp dimer and two 280 bp monomers).

**Figure 25** shows the Southern hybridization of the probe pS*Qro*G102, a 280 bp P1/P2 spacer clone (*Q. robur*, tree G102), to *Q. robur* and *Q. petraea* genomic DNA following digestion with *Sca* I. This restriction endonuclease has a single recognition site within the 5S rDNA coding sequence. A ladder of hybridizing fragments characteristic of DNA sequences organized in tandem arrays was observed in both species. Up to 10 resolved multimers of a 350 bp 5S rDNA repeat can be detected among the partial digest products. We concluded that a major 5S rDNA unit size class 350 bp long occurs in both pedunculate and sessile oak, with gene and spacer approximately 120 bp and 230 bp long, respectively.

The intragenomic copy number of the 5S rDNA repeat was estimated in pedunculate oak (tree DF159) by Southern blot experiments using the pS*Qro*G102 probe. Preliminary results indicated approximately 1000–2000 copies per genome 2C (data not shown) based on a diploid DNA amount of 1.84 pg per nucleus (Favre and Brown 1996).

b- 5S rRNA gene variations

We were unable to detect length polymorphisms within the coding region of either pedunculate or sessile oak in our sample. Based on the nucleotide sequence of S*Qro*G102, two primers R3 and D4 were designed to amplify the entire repeating unit (**Figure 24**). PCR amplification using these two primers produced the expected 350 bp fragment in both species (data not shown). The DNA sequences of three R3/D4 clones from each species (six individuals) were aligned and compared to known plant 5S rRNA genes (Szymanski *et al.* 1999) to determine the *Q. robur* and *Q. petraea* 5S rDNA coding regions shown in **Figure 26**. Visual inspection revealed that four out of the six clones (two from each species) are identical in both length (120 bp) and sequence (G*Qro*3P, G*Qro*A4, G*Qpe*390096-3, G*Qpe*96-3). One nucleotide difference is observed in G*Qpe*QS9 (position 6) whereas G*Qro*DF159 is 119 bp long and has in addition 3 nucleotide variations (at positions 2, 86, 106). This latter variant gene sequence could represent a pseudogene because the variations at positions 86 and 106 occur in 2 regions (named "box C" and "RNA binding domain", respectively) interacting with transcription factor IIIA (TFIIIA) that promotes the initial binding of Pol III to the DNA (Kellogg and Appels 1995). The fact that we sequenced a putative pseudogene clearly suggests that variant gene are very common within individual. The nucleotide diversity in the 5S gene region was first estimated to

be 0.017 in *Quercus robur* and 0.005 in *Quercus petraea*. However, due to the lack of fixed differences between pedunculate and sessile oak, we considered these two species as one *Quercus robur* group, as suggested by Kleinschmit *et al.* (1995) and Steinhoff (1998), and pooled the 6 clones to estimate the nucleotide diversity in 5S rRNA genes between individual plants (0.014). Excluding the putative pseudogene, the nucleotide diversity estimate is only 0.003.

The consensus coding region of the *Q. robur* group (**Figure 26**) is very similar to that reported for other angiosperms (Szymanski *et al.* 1999). Only two nucleotide differences (at positions 17 and 61) were observed with 5S rRNA genes from *Gossypium turneri*, *G. lobatum*, *G. herbaceum*, *Beta vulgaris* and *Brassica napus*. These two nucleotide positions are involved in base pairing within stem II of the generalized secondary structure model proposed for plant 5S rRNAs (Barciszewska *et al.* 1994) and appeared to be a characteristic feature of the *Q. robur* group ($^{17}$A-T$^{61}$ pairing instead of $^{17}$G-C$^{61}$ in other angiosperms).

c- Spacer variations

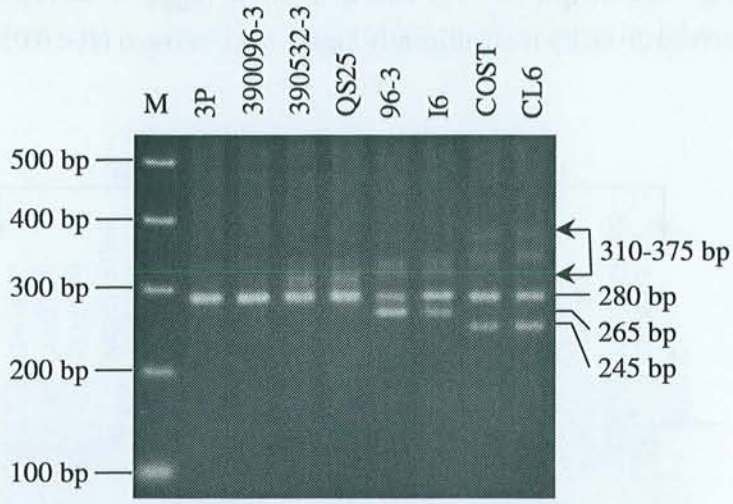
α- Sample diversity

A general survey of the P1/P2 monomer-product zone (200–400 bp) on agarose gels revealed that about 15% of sampled trees exhibited 1–3 fragments in addition to the ubiquitous 280 bp product (corresponding to the 230 bp spacer class). This polymorphism resulted in four different banding patterns (I-IV) shown in **Figure 27A**. Using restriction site analysis, the 245 bp and 265 bp fragments observed in patterns III and IV were identified as two additional spacer length variants of about 195 bp and 215 bp, respectively. Reamplification of the unstable (i.e. varying in intensity) and quite fuzzy bands observed in the 310–375 bp size zone, produced the 280 bp fragments (pattern II), the 280 bp and 265 bp fragments (pattern III), or the 280 bp and 245 bp fragments (pattern IV). This indicated that the 280 bp, 265 bp and 245 bp fragments are fully homologous, allowing formation of heteroduplexes with altered electrophoretic mobility during the final gradual cooling step of PCR amplification. Since pattern I is over-represented (**Figure 27B**), all spacer variants probably arose from the main 230 bp spacer class (see below). Interestingly, the 195 bp and 215 bp spacer variants were only detected in *Q. robur* (pattern IV) and *Q. petraea* (pattern III), but at low frequency in our samples (5.2 and 8.3 %, respectively).

Heteroduplexes observed in pattern II clearly revealed unresolved microheterogeneity within the 230 bp spacer class. To evaluate and identify this polymorphism, separation of homoduplexes was undertaken on non-denaturing polyacrylamide gels considering that in such high resolution matrices DNA fragments migrate not only as a function of their length but also as a function of their base composition (Hauser *et al.* 1998). Considerable polymorphism could be detected after silver staining. In order to exclude most dimers, heterodimers and heteroduplexes from the analysis, we focused on the monomer-product zone conveniently defined on acrylamide gel by the 200- and 300-bp fragments

of the 100 bp DNA ladder. Considering both faint and strong PCR products, the mean number of bands per individual in this zone was significantly higher in sessile oak than in pedunculate oak (2.64 vs 2.04; probability from one-tailed t-test: ~ 0.0025). Seven major spacer variant levels (electromorph classes
*a–g*) could be reliably scored providing that a fine gel to gel match is achieved (**Figure 28**). Five levels (*a–e*) were common to both species and corresponded to the microheterogeneity detected on agarose gels within the 230 bp spacer class (patterns I and II, **Figure 27**). Levels *f* and *g* corresponded to the 195 bp and 215 bp spacer variants observed in sessile oak (pattern III) and pedunculate oak (pattern IV), respectively.

**A.**




**B.**

| | I | II | III | IV |
|---|---|---|---|---|
| *Q. robur* | 89.6 | 5.2 | 0.0 | 5.2 |
| *Q. petraea* | 77.8 | 13.9 | 8.3 | 0.0 |
| Total | 85.0 | 8.6 | 3.2 | 3.2 |

**Figure 27**. Banding patterns observed on agarose gel (I, II, III and IV) after P1/P2 spacer amplification (**A**) and their frequency distribution (%) in *Q. robur* and *Q. petraea* (**B**). Only the monomer-product zone is shown. Lane M: 100 bp DNA ladder; 3P, 390532-3, COST and CL6: pedunculate oaks; 390096-3, QS25, 96-3 and I6: sessile oaks.

The 5S rDNA spacer phenotypes (major fragment patterns) and their distribution in pedunculate and sessile oaks are summarized in **Table 21**. As many as 17 variable phenotypes were recognized (12 per species) and 7 were common to both species. Many of them are poorly represented, the extreme cases being four phenotypes ("*bcg*", "*bf*", "*bce*" and "*abc*") only detected in a single individual. The most frequent phenotypes in pedunculate ("*b*") and sessile oak ("*bc*") were observed in 35% and 25% of the individuals, respectively. There were 33 trees (35.5%) with only one major variant, 57 (61.3%) with two major variants, and 3 trees (3.2%) with three major variants.

Because of the multiplicity of the 5S rDNA units within an individual (1000-2000 copies/2C), two or more spacer variants can occur within a single 5S rDNA array. Thus, diagnostic heterozygosity (different variants on homologous chromosomes) to establish genotypes from phenotypes consisting of two or more spacer variants remained uncertain although we have detected and recently mapped only one 5S rDNA locus in *Q. robur* (Barrénèche *et al.* 1998). In addition, we have preliminary results on 5S rDNA inheritance in *Q. robur* and *Q. petraea* indicating that major spacer variants cannot be routinely considered as allelic markers from this single locus (Trontin and Favre, unpublished data ; voir **Partie II** de ce **Chapitre III**). Therefore, we used the Shannon's information measure (Lewontin 1972) based on phenotype frequencies to quantify the sample relative diversity of 5S rDNA spacers in *Q. robur* ($J_{qro} = 0.779$) and *Q. petraea* ($J_{Qpe} = 0.899$). A one-tailed t-test concluded that the observed diversity is significantly higher in *Q. petraea* ($P < 0.05$).

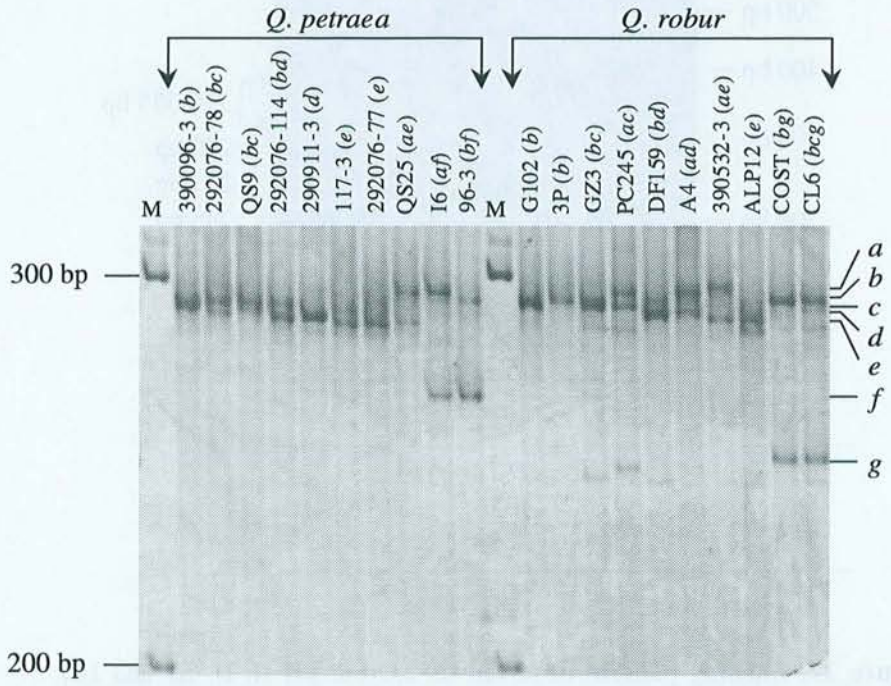



**Figure 28**. Spacer polymorphism detected on non-denaturing polyacrylamide gel in *Q. petraea* and *Q. robur* (monomer-product zone). Seven major spacer variant levels (*a–g*) can be reliably scored. Ten individuals from each species are presented and their phenotype indicated in parentheses. Provenance of each tree can be recovered from **Table 22**. M: 100 bp DNA ladder.

Fisher exact tests were finally performed on spacer variant frequencies (recovered from **Table 21**) to determine their contribution to species differentiation. At the 1 % probability level, there was only a significant difference for variant class *e* ($P < 0.001$), which was found in 11 sessile oaks (33.5 %) and only 2 pedunculate oaks (3.5 %). Most of these trees (9) were remarkably distributed in the southern parts of Europe (Portugal, Spain, Southwestern France, Italy, Slovenia and Serbia).

Considering the spacer variants *f* and *g* only detected in sessile (3 trees) and pedunculate oak (3 trees), respectively, the test is non significant for *g* ($P > 0.20$) but considered significant for *f* at $P < 0.06$.

**Table 21.** Distribution of spacer phenotypes (major fragments) in *Q. robur* and *Q. petraea*

| Phenotype* | Number of individuals | | |
|---|---|---|---|
| | *Q. robur* | *Q. petraea* | Total |
| *b* | 20 | 4 | 24 |
| *bd* | 14 | 5 | 19 |
| *bc* | 4 | 9 | 13 |
| *ab* | 4 | 2 | 6 |
| *ad* | 5 | 0 | 5 |
| *e* | 1 | 4 | 5 |
| *be* | 0 | 4 | 4 |
| *ae* | 1 | 2 | 3 |
| *a* | 2 | 0 | 2 |
| *ac* | 2 | 0 | 2 |
| *bg* | 2 | 0 | 2 |
| *d* | 1 | 1 | 2 |
| *af* | 0 | 2 | 2 |
| *bcg* | 1 | 0 | 1 |
| *bf* | 0 | 1 | 1 |
| *bce* | 0 | 1 | 1 |
| *abc* | 0 | 1 | 1 |

* Letter symbols (*a–g*) indicate the banding pattern observed on acrylamide gel (see text).

β- Nucleotide variations

We sequenced 30 P1/P2 spacer clones obtained from 10 pedunculate oaks (16 sequences) and 10 sessile oaks (14 sequences) of various provenances (**Table 22**). Most clones were selected (rather than randomly chosen) on the basis of (i) their high representativity in the genome (only major PCR products were cloned) and (ii) their different electrophoretic mobility on acrylamide gel (thought to be highly correlated with divergence). This procedure was designed to obtain estimates of nucleotide diversity and tentatively overcome the definite problem of sampling only a small fraction of the existing repeats within individual (1–3 sequences out of ca. 1000–2000 in oak). Each electromorph class detected on acrylamide gel (*a–g*) is generally represented by at least two sequences randomly chosen from two different individuals of both species. Only electromorph class *c* has a unique clone sequenced from *Q. petraea*. The size ranged from 197–198 bp (class *g*) to 230–233 bp (class *a*) with most spacers (20) being 230 bp long, thus demonstrating little intra- and interclass variation. However, 24 of the 30 sequences were different from each other, indicating a high degree of sequence polymorphism.

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

**Table 22**. Spacer clones sequenced in *Quercus* species and listed by EMBL accession number.

| Species | Tree | Provenance[a] | Clone designation | Size (bp) | Spacer class[b] | Accession number |
|---|---|---|---|---|---|---|
| *Q. robur* | A4 | Arc | S*Qro*A4*a* | 230 | *a* | AJ245074 |
| | | | S*Qro*A4*d* | 230 | *d* | AJ245080 |
| | | | S*Qro*A4 | 230 | nd | AJ245086 |
| | 390532-3 | Mon | S*Qro*390532-3*a* | 233 | *a* | AJ245075 |
| | | | S*Qro*390532-3*e* | 230 | *e* | AJ242951 |
| | COST | Gei | S*Qro*COST*b* | 230 | *b* | AJ245076 |
| | | | S*Qro*COST*g* | 198 | *g* | AJ245084 |
| | G102 | Fon | S*Qro*G102*b* | 230 | *b* | AJ242947 |
| | 3P | Pie | S*Qro*3P*b* | 230 | *b* | AJ245077 |
| | | | S*Qro*3P*d* | 230 | *d* | AJ245082 |
| | | | S*Qro*3P | 230 | nd | AJ245087 |
| | PC245 | PeC | S*Qro*PC245*c* | 230 | *c* | AJ245078 |
| | GZ3 | Fri | S*Qro*GZ3*c* | 229 | *c* | AJ245079 |
| | DF159 | Fon | S*Qro*DF159*d* | 230 | *d* | AJ245081 |
| | ALP12 | Esc | S*Qro*ALP12*e* | 230 | *e* | AJ245083 |
| | CL6 | Val | S*Qro*CL6*g* | 197 | *g* | AJ245085 |
| *Q. petraea* | I6 | Sar | S*Qpe*I6*a* | 232 | *a* | AJ245088 |
| | | | S*Qpe*I6*f* | 217 | *f* | AJ245099 |
| | QS25 | Orl | S*Qpe*QS25*a* | 233 | *a* | AJ245089 |
| | 96-3 | Tem | S*Qpe*96-3*b* | 230 | *b* | AJ245090 |
| | | | S*Qpe*96-3*f*1 | 217 | *f* | AJ245098 |
| | | | S*Qpe*96-3*f*2 | 217 | *f* | AJ245098 |
| | 390096-3 | IrV | S*Qpe*390096-3*b* | 230 | *b* | AJ245091 |
| | 292076-78 | Koz | S*Qpe*292076-78*b* | 230 | *b* | AJ245092 |
| | | | S*Qpe*292076-78*c* | 230 | *c* | AJ245093 |
| | 290911-3 | Rid | S*Qpe*290911-3*d* | 230 | *d* | AJ245094 |
| | 292076-114 | Koz | S*Qpe*292076-114*d* | 230 | *d* | AJ245095 |
| | 117-3 | Olo | S*Qpe*117-3*e* | 231 | *e* | AJ245096 |
| | 292076-77 | Koz | S*Qpe*292076-77*e* | 230 | *e* | AJ245097 |
| | QS9 | Orl | S*Qpe*QS9 | 230 | nd | AJ245100 |
| *Q. suber* | *Qsu* | Southern Europe | S*Qsu* | 306 | / | AJ245101 |
| *Q. libani* | *Qli* | Western Asia | S*Qli*S | 242 | / | AJ245102 |
| | | | S*Qli*L | 323 | / | AJ245103 |

[a] See **Table 20** for abbreviations.

[b] Refers to the seven major electromorph classes (*a*-*g*) detected on acrylamide gels in *Q. robur* and *Q. petraea* (nd: not determined).

Alignment was straightforward (**Figure 29**), requiring only a single nucleotide insertion (position 24; S*Qpe*117-3*e*) or deletion (position 90; S*Qro*GZ3*c*, S*Qro*CL6*g*, and S*Qpe*I6*a*) and several gaps in a region of relatively low conservation in the middle spacer (positions 114–148, **Figures 29-30**). Gaps introduced at positions 117–129 and 114–148 were respectively interpreted as deletions 13 bp long in sequences from electromorph class *f* (S*Qpe*96-3*f*1, S*Qpe*96-3*f*2, and S*Qpe*I6*f*) and 32 bp long in sequences from electromorph class *g* (S*Qro*COST*g* and S*Qro*CL6*g*). Small duplicated motifs (1 and 2) are indeed observed near both sides of indel-type "*f*" (TYYGR, positions 115–119 and 127–

131) and "*g*" (RATBTCG, positions 107–113 and 139–148) and such small regions of similarity could occasionally be responsible for deletion in a model of replication slippage (Scoles *et al.* 1988). Conversely, gaps introduced at positions 140–142 were most readily accounted for by the amplification in 3 sequences from electromorph class *a* (S*Qro*390532-3*a*, S*Qpe*I6*a*, and S*Qpe*QS25*a*) of a small 3 bp element (GCG) located in a short microsatellite-like region (GYGGCRGCR, positions 131–142). Nucleotide substitutions occurred across spacers with transition/transversion ratios of about 1.87 in *Q. robur* and 2.34 in *Q. petraea*. They especially accumulated in the middle spacer region prone to insertion and deletion events (ca. positions 114–148, **Figures 29-30**) and near a hot spot at position 205.

Although we have detected up to 15 substitutions between spacers of the two species (compare S*Qro*A4*a* and S*Qpe*292076-78*c*, **Figure 29**), we were unable to detect any fixed interspecific difference. Therefore, we quantified the molecular diversity of 5S rDNA spacers in *Q. robur* and *Q. petraea* as well as in the *Q. robur* group (*Q. robur* + *Q. petraea*). The proportion of nucleotide positions that are polymorphic (polymorphism) within individual was found to be similar in pedunculate and sessile oaks (**Table 23**) and intraindividual mean in the *Q. robur* group based on 7 individuals (4 pedunculate and 3 sessile oaks) was estimated to be 0.065 (gaps included) or 0.025 (gaps excluded). Polymorphism was significantly higher within species (10 individuals), as expected. 27.9% (14.7% when gaps are excluded) of nucleotide positions were polymorphic in *Q. robur* and 17.9% (11.1% when gaps are excluded) in *Q. petraea*. Within the *Q. robur* group (20 individuals) the estimate reached 33.8% (21.3% when gaps are excluded). Polymorphism is however highly dependent on sample size. We therefore calculated the less biased mean difference among spacers (**Table 23**). Remarkably, intraspecific and intraindividual values are rather identical and differences between pedunculate and sessile oak are low (gaps included) to nil (gaps excluded). Excluding gaps, the intraindividual and intraspecific mean differences among spacers in the *Q. robur* group were finally estimated to be 0.020 and 0.023, respectively.

Since two or more randomly chosen nucleotide sequences are generally available for each electromorph class (**Table 22**), intraclass and interclass nucleotide diversity ($\pi$) could finally be estimated. Excluding deletions and insertions, the $\pi$ values obtained were not significantly different between pedunculate and sessile oak. We therefore pooled the sequences to calculate the $\pi$ values in the *Q. robur* group (**Table 24**). Intraclass nucleotide diversity is considerably lower within classes *e–g* (mean $\pi$ value: 0.004) compared to classes *a–d* (mean $\pi$ value: 0.027). Considering classes *a*, *e*, *f* and *g*, interclass nucleotide diversity is higher than intraclass estimates, as expected. However, in the case of classes *b*, *c*, and *d*, interclass estimates are sometimes lower, indicating that differences in electrophoretic mobility on crosslinked gels are poorly correlated with divergence.

```
Consensus   ctCtTTTTCT TCGGATTTTT TTT-gTCGcC GGGCTtACCG TTTACGGCGG tCcAgCaCGC TcGCTTCGTa GgcTTAGTCt   80

SQroA4a     .......... .......... .......... .......... .......... .......... .......... .........C
SQro390532-3a .......... .......... .......... .......... .......... .......... .......... ..........
SQroCOSTb   .......... .......... .......... .......... .......... .......... .......... ..........
SQroG102b   .......... .......... .......... .......... .......... .......... .......... ..........
SQro3Pb     A......... .......... .......... .......... .......... .......... .......... ..........
SQroPC245c  .......... .......... .......... .....A.... .......... C......... .T........ ..........
SQroGZ3c    .......... .......... .......... .......... .......... .......... .......... ..........
SQroA4d     ...C...... .......... ....T..... .......... .......... .......... .......... ..........
SQroDF159d  .......... .......... .......T.. .......... .......... .......... .......... .C........
SQro3Pd     .......... .......... .......... .......... .......... .......... .......... ..........
SQroALP12e  .......... .......... .......... .......... .......... .......... .......... ..........
SQro390532-3e .......... .......... .......... .......... .......... .......... .......... ..........
SQroCOSTg   ...A...... .......... .......... .......... .......... .......... .......... ..........
SQroCL6g    .......... .......... .......... .......... .......... .......... .......... ..........
SQroA4      .......... .......... .......... .......... .......... .......... .........T ..........
SQro3P      .......... .......... .......... .....A.... .......... .......... .......... ..........

SQpeI6a     .......... .......... .......... .......... .......... .......... .......... ..A.......
SQpeQS25a   .......... .......... .......... .......... .......... .......... .......... ..........
SQpe96-3b   .......... .......... .......... .......... .......... .......... .......... ..........
SQpe390096-3b .......... .......... .......... .......... .......... .......... .......... ..........
SQpe292076-78b .......... .......... .......... .......... .......... ..T.T..... .......... .A........
SQpe292076-78c .C........ .......... .......... .......... .......... ..T....... .......... ..........
SQpe290911-3d .......... .......... .......... .......... .......... .......... .......... ..........
SQpe292076-114d .......... .......... .......... .......... .......... .......... .......... ..........
SQpe117-3e  .......... .......... ...T...... .......... .......... .......... .......... ..........
SQpe292076-77e .......... .......... .......... .......... .......... .......... .......... ..........
SQpe96-3f1  .......... .......... .......... .......... .......... .......... .......... ..........
SQpe96-3f2  .......... .......... .......... .......... .......... .......... .......... ..........
SQpeI6f     .......... .......... .......... .......... .......... .......... .......... ..........
SQpeQS9     .......... .......... .......... .......... .......... .....G.... .......... ..........
```

```
                                                 1         2                2                1
Consensus   GCCCCAGGCc AAAAAtACga gCCGCTGATc TCGttttggg tgctaaattg gtggcggcg- --atctcgAA ATGCcCTCGG  160

SQroA4a     .......... .........T .......... .......... .......... .......... .......... ..........
SQro390532-3a .......... .......... .......... .......... .......... .........G CG........ ..........
SQroCOSTb   .......... .......... .......... .......... .......... .......... .......... ..........
SQroG102b   .......... .......... .......... .......... .......... .......... .......... ..........
SQro3Pb     .......... .......... .......... .......... .......... .......... .......... ..........
SQroPC245c  .......... .......... .......... .......... .......... .......... .......... ..........
SQroGZ3c    .........- .......... .......... .......... ...C...... .......... .......... ..........
SQroA4d     .......... .......... .......... .......A.. .......... .......... .......... ..........
SQroDF159d  .......... .......... .......... .......... .......... .......... .......... ..........
SQro3Pd     .......... .......... .......... .......... .......C.. .......... .......... ..........
SQroALP12e  .......... .......... .......T.. .......... .......... .......... .......... ..........
SQro390532-3e .......... .......... .......T.. .......... .......... .......... .......... ..........
SQroCOSTg   .......... .......... .......... ...------- ---------- ---------- -------... ..........
SQroCL6g    .........- .......... .......... ...------- ---------- ---------- -------... ..........
SQroA4      .......... .......... .......... .......... .......... .......... .......... ..........
SQro3P      .........T .....G..CT A......... .......... .......... .......... .......... ..........

SQpeI6a     .........- .......... .......... ...A...... .......... .........G CG........ ..........
SQpeQS25a   .......... .......... .......... .......... .......... .........G CG........ ..........
SQpe96-3b   .......... .......... .......... .......... .......... .......... .......... ..........
SQpe390096-3b .......... .......... .......... .......... .......... .......... .......... ..........
SQpe292076-78b .......... ........T. .......... .......... .......... .....A.... .......... ..........
SQpe292076-78c .......... ........T. .......... .......... .......... .....A.... .......... ....G.....
SQpe290911-3d .......... .......... .......... .......... ....G..... .......... .......... ..........
SQpe292076-114d .......... .......... .......... .......... .......C.. .......... .......... ..........
SQpe117-3e  .......... .......... .......T.. .......... .......... .......... .......... ..........
SQpe292076-77e .......... .......... .......T.. .......... .......... .........A .......... ..........
SQpe96-3f1  .......... .......... .......... .......---- --------- .......... ....G..... ..........
SQpe96-3f2  .......... .......... .......... .......---- --------- .......... ....G..... ..........
SQpeI6f     .......... .......... .......... .......---- --------- .......... ....G..... ..........
SQpeQS9     .......... .......... .......... .......... .......C.. .C........ .......... ..........
```

```
Consensus      GCTGcAgCTC CCTGGTGTtA GAgTGcaTCG cTGCATtcgg ATCcttTAtA AGGATTTAAc GcGCttatGC TGAT  234

SQroA4a        .......... .......... ..A....... .......... .....C.... .......... .....GC... ....
SQro390532-3a  .......... .......... .......... .......... .....C.... .......... .......... ....
SQroCOSTb      .......... .......... .......... .......... ....C..... .......... .......... ....
SQroG102b      .......... .......... .......... .......... ....C..... .......... .......... ....
SQro3Pb        .......... .......... .......... .......... .......C. .......... .......... ....
SQroPC245c     ......A... .......... .......... .......... .......... .......... .......... ....
SQroGZ3c       .......... .......... .......... .......... ....C..... .......... .....C.... ....
SQroA4d        .......... ........G. .......... .......... ....C..... .......... .......... ....
SQroDF159d     .......... .......... ......G... .......... ....G..... .......... .T........ ....
SQro3Pd        .......... .......... .......... .......... ...A...... .......... .....C.... ....
SQroALP12e     .......... .......... .......... .......... .......... .......... .......... ....
SQro390532-3e  .......... .......... .......... .......... .......... .......... .......... ....
SQroCOSTg      .......... .......... .......... .......... ....C..... .......... .......... ....
SQroCL6g       .......... .......... .......... .......... ....C..... .......... .......... ....
SQroA4         .......... .......... .......... .........A ....C..... .......... .......... ....
SQro3P         .......... .......... .......... ........A. .......... .......... .......... ....

SQpeI6a        .......... .......... .......... .......... .......... .......... .......... ....
SQpeQS25a      .......... .......... .......... .......... .......... .......... .......... ....
SQpe96-3b      .......... .......... .......... .......... .......... .......... .......... ....
SQpe390096-3b  .......... .......... .....A.... ......A... ....C..... .......... .......... ....
SQpe292076-78b .......... .......... .......... .......T.A .......... .........T .......... ....
SQpe292076-78c .......... .......... .......... .......T.A .......... .........T ....G..... ....
SQpe290911-3d  .......... .......... .......... .......... ....C..... .......... .......... ....
SQpe292076-114d .......... .......... .......... .......... ....C..... .......... .......... ....
SQpe117-3e     .......... .......... .......... .......... .......... .......... .......... ....
SQpe292076-77e ....T..... .......... .......... .......... .......... .......... .......... ....
SQpe96-3f1     .......... .......... .......... T......... .......... .......... .......... ....
SQpe96-3f2     .......... .......... .......... T......... .......... .......... .......... ....
SQpeI6f        .......... .......... .......... T........A .......... .......... .......... ....
SQpeQS9        .......... .......... .......... .......... .......... .......... .......... ....
```

**Figure 29**. Aligned nucleotide sequences of 30 cloned 5S rDNA spacers from *Q. robur* (*Qro*) and *Q. petraea* (*Qpe*). Polymorphisms are indicated by standard IUPAC symbols. Clone characteristics are indicated in **Table 22**. Small duplicated motifs (1 and 2) probably implicated in replication slippage (see text) are boxed in the majority-rule consensus sequence. Shared polymorphisms among all sequences sampled within a given spacer class (*a–g*) are shaded.

χ- Phylogenetic comparison of major 5S rDNA spacer classes (*a–g*)

Despite substantial spacer variation, only a few informative sites were detected. Visual inspection of alignment (**Figure 29**) revealed that shared polymorphism (i.e. shared by all sequences sampled within a given electromorph class) only occur within class *g* (32 bp deletion, positions 114–148), *f* (13 bp deletion, positions 117–129; "G" at position 145; "T" at position 191) and *e* ("T" at position 110). Remarkably, the unique shared polymorphism within class *e* is responsible for the difference in electrophoretic mobility detected on acrylamide gel (see **Figure 28**) because this is the only change observed in the two sequences S*Qro*390532-3*e* and S*Qro*ALP12*e* compared to the consensus sequence. In contrast, there are no shared polymorphisms within classes *a–d*, but 3 out of 4 sequences from class *a* were found to have an insertion (GCG) at positions 140–142.

Phylogenetic analysis of the 30 spacers from the *Q. robur* group was conducted on the data matrix shown in **Figure 29**. Since morphological and molecular data sets support the red oaks (section *Erythrobalanus*, Camus 1954) as ancient and monophyletic group (Nixon 1993), two trees belonging to the species *Q. rubra* L. and *Q. palustris* Muench. were initially included as outgroup in order to root the gene trees. Because of inappropriate sequence divergence of the spacer clones isolated from

these two species we have not included them here. Several spacers were instead isolated from two not so diverged species (*Q. suber* L. and *Q. libani* Oliv.) belonging to the section *Cerris* (Camus 1954). This section is considered by Nixon (1993) to be a sister group to the white oak *sensu stricto* (section *Quercus* of Camus 1954) and combined parsimony analysis of ITS sequences (Internal Transcribed Spacer of 18S-5.8S-26S rDNA) and chloroplast DNA restriction sites strongly support for the monophyly of this group (Manos *et al.* 1999).

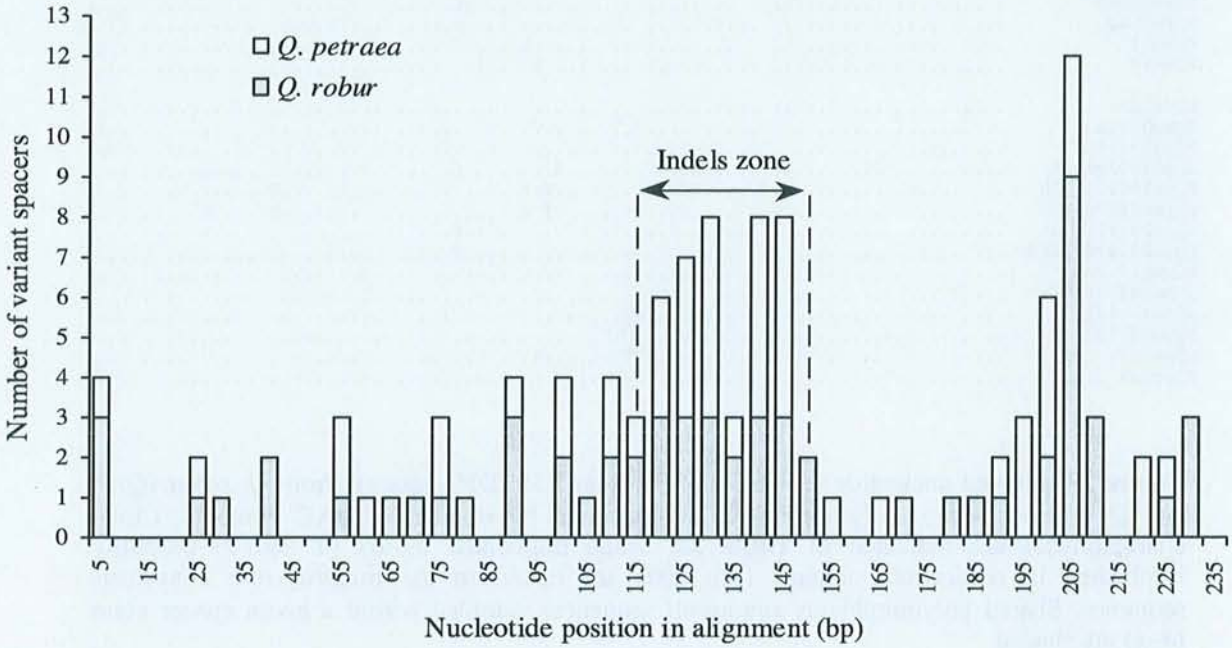



**Figure 30**. Synthetic diagram showing the distribution of variant sites across thirty 5S rDNA spacers in *Q. robur* and *Q. petraea*. Vertical axis is number of spacers with difference at a given site in reference to the consensus sequence.

P1/P2 spacer amplifications produced one major band (ca. 350 bp) in *Q. suber* and two major bands (ca. 290 bp and 370 bp) in *Q. libani* (data not shown). One clone of each of these major bands were sequenced (**Table 22**). The *Q. suber* spacer is 306 bp long and the *Q. libani* spacers are 242 bp and 323 bp long, respectively. Alignment with the majority-rule consensus sequence of the *Q. robur* group is shown in **Figure 31**. The two *Q. libani* spacers (S*QliS* and S*QliL*) mainly differ by one large tandemly duplicated region of the 5' end (83 bp) that initiates in the 5S gene. This duplication is old enough to have accumulated 11 substitutions. Excluding gaps, the observed polymorphism (0.054) is quite high compared to that of individuals from the *Q. robur* group (see **Table 23**). Interestingly, the same 83 bp duplication was found in the *Q. suber* spacer and the mean difference (gaps excluded) with the *Q. libani* spacers is low (0.030), suggesting that these two species are closely related. Disregarding the duplication, alignment with the consensus sequence of the *Q. robur* group was straightforward and required only a few insertions of gaps (mean sequence identity of about 81%). The *Cerris* sequences therefore provided a good outgroup for the phylogenetic reconstruction of

spacer classes detected in the *Q. robur* group. Excluding indels, there are 24 synapomorphies between the *Cerris* and *Q. robur* groups.

**Table 23**. Polymorphism and mean differences among spacers from *Q. robur* and *Q. petraea**.

| Taxon | N | Polymorphism | | Mean difference | |
|---|---|---|---|---|---|
| | | Gaps included | Gaps excluded | Gaps included | Gaps excluded |
| ***Q. robur*** | **16** | **0.279** | **0.147** | **0.057** | **0.023** |
| 390532-3 | 2 | 0.021 | 0.009 | 0.021 | 0.009 |
| A4 | 3 | 0.057 | 0.057 | 0.037 | 0.038 |
| COST | 2 | 0.143 | 0.005 | 0.141 | 0.005 |
| 3P | 3 | 0.052 | 0.052 | 0.034 | 0.035 |
| Intraindividual mean | | 0.068 | 0.031 | 0.058 | 0.022 |
| ***Q. petraea*** | **14** | **0.179** | **0.111** | **0.047** | **0.023** |
| I6 | 2 | 0.094 | 0.023 | 0.094 | 0.023 |
| 292076-78 | 2 | 0.022 | 0.022 | 0.021 | 0.022 |
| 96-3 | 3 | 0.065 | 0.009 | 0.043 | 0.006 |
| Intraindividual mean | | 0.060 | 0.018 | 0.053 | 0.017 |
| ***Q. robur* group** | **30** | **0.338** | **0.213** | **0.052** | **0.023** |
| Intraindividual mean | | 0.065 | 0.025 | 0.056 | 0.020 |

* Bold entries represent intraspecific (*Q. robur*; *Q. petraea*), or intragroup values (*Q. robur* group = *Q. robur* + *Q. petraea*). Normal entries represent intraindividual values and means from *Q. robur* (trees 390532-3, A4, COST, 3P), *Q. petraea* (trees I6, 292076-78, 96-3) or the*Q. robur* group. N = the number of clones sequenced per taxon.

**Table 24**. Intraclass nucleotide diversity (diagonal) and interclass mean difference (below diagonal) among 5S rDNA spacers (gaps excluded) in the *Q. robur* group.

| N* | Spacer class | *a* | *b* | *c* | *d* | *e* | *f* | *g* |
|---|---|---|---|---|---|---|---|---|
| 4 | *a* | 0.018 | | | | | | |
| 6 | *b* | 0.019 | 0.020 | | | | | |
| 3 | *c* | 0.033 | 0.030 | 0.047 | | | | |
| 5 | *d* | 0.025 | 0.022 | 0.035 | 0.025 | | | |
| 4 | *e* | 0.016 | 0.017 | 0.030 | 0.021 | 0.005 | | |
| 3 | *f* | 0.021 | 0.022 | 0.034 | 0.023 | 0.018 | 0.003 | |
| 2 | *g* | 0.018 | 0.014 | 0.028 | 0.013 | 0.014 | 0.014 | 0.005 |

* N = the number of clones sequenced per electromorph class (see **Table 22**)

The consensus gene tree resulting from maximum parsimony analysis is shown in **Figure 32**. The 13 bp and 32 bp deletions (variant classes *f* and *g*, respectively) as well as the GCG insertion (variant class *a*) were coded as one single event. Other gaps were coded as missing data. As expected,

variant spacers from classes *e–g* and three spacers from class *a* clustered together due to the shared polymorphisms described above. Interestingly, spacers from class *f* that were only found in three sessile oaks (8.4%) appeared as the outermost branch of the *Q. robur* group. This position is clearly supported by two shared polymorphisms with the three *Cerris* sequences ("G" and "T" at positions 145 and 191). This suggests that these two structural features of class *f* can be considered as the most ancient in term of evolutionary time (but not the 13 bp deletion). In contrast, the shared polymorphism detected within classes *a* (GCG insertion), *e* ("T" at position 110), and *g* (32 bp deletion) may have established more recently.

```
                                T-rich region
SQro/SQpe  cccctCTCTT TTTCTTCGGA TTTTTTTT-- ---GTCGCCG GGCTTACCGT TTACGGCGGT CCAGCACGCT CGCTTCGTAG GCT  83
SQliS      ....... ..... .....C. ........TT T--C...TT. ......T... .....T.... .......... .....T.... .T.
SQliL 1    ....... ..... .....C. ........TT T--..TT. ..G..GT... .......... .......... ....CT.C.T ..A
SQsu 1     ....... ..... .....C. .........-- ---..TT. ..G..GT... .......... .......... ....CT.C.T ..A

SQliL 2    .....T..A. ....CCA.C. ........TT TTTC...T.. ......T... .....T.... .......... .....T.... ... 166
SQsu 2     .....T..A. ....CCT.C. .........-- ---C...TT. ......T... .C...C.... .......... .....T.... ...

SQro/SQpe  TAGTCTGCCC CAGGCCAAAA ATACGAGCCG CTGATCTCGT TTTGGGTGCT AATTTGGTGG CGGCGATCTC GAAATG---- -----C 169
SQliS      ...C.C.... .......... .C........ .A........ .......... ..AGC-..C. ......CG.. .....CGGCG GAAGC.
SQliL      ...CGC.... .......... .C........ .G........ .......... ..AGC-..C. ......CG.. .....CCGCC CAAGC. 252
SQsu       ...C.C.... .......... .C........ .A........ .......... ..A------- --....CG.. .....CCGCG GAAGC.

                                                                                AT-rich      GC-rich
                                                                                region       region
SQro/SQpe  CCTCGGGCTG CAGCTCCCTG GTGTTAGAGT GCATCGCTGC ATTCGGATCC TTT-ATAAGG ATTTAACGCG CTTATGCTGA T 250
SQliS      .......T.. .GC...T.C. .C....AG.. ...C..T... G......... ...T....A ...A.....  .......... .
SQliL      .......T.. ..C...T.C. .C....AG.. ......T... G......... ...T....A ...A.....  .......... . 333
SQsu       .......T.. ..C...T.C. .C....AG.A ......T... G......... ...T....A ...A.....  .......... .
```

**Figure 31**. Nucleotide sequences of cloned 5S rDNA spacers from *Q. suber* (S*Qsu*) and *Q. libani* (S*Qli*S and S*Qli*L) aligned with the consensus sequence of the *Q. robur* group (S*Qro*/S*Qpe*). Polymorphisms are indicated by standard IUPAC symbols. The two elements (1 and 2) of a large tandemly duplicated region (83 bp) that starts 5 bp upstream the 3' end of the 5S gene (bold lower case letters) in *Q. suber* and *Q. libani* (clone S*Qli*L) are aligned to facilitate comparison. The putative regulatory elements referred to in the discussion are boxed.

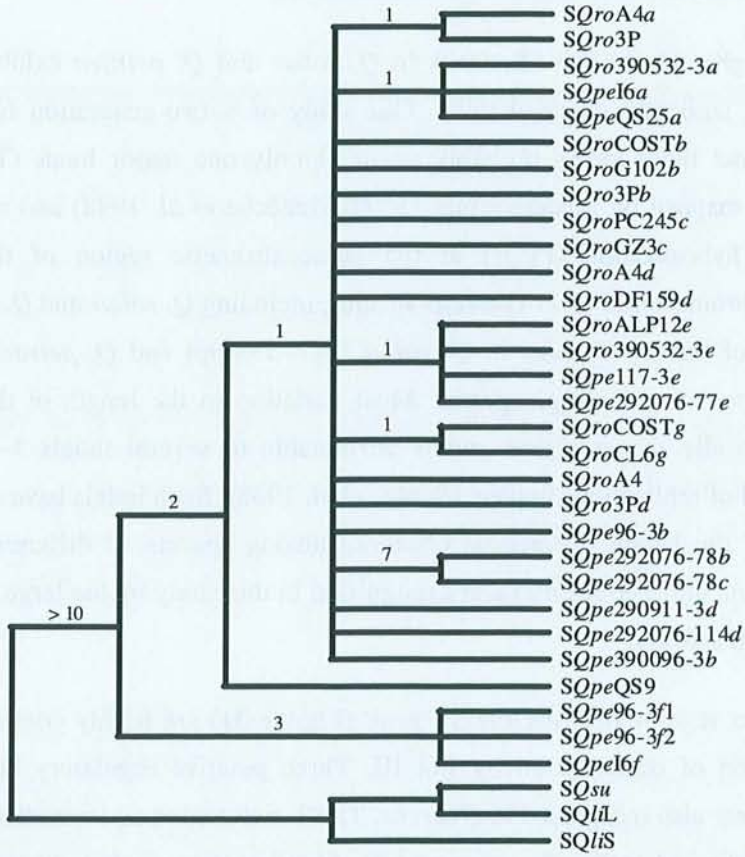



**Figure 32**. Consensus gene tree resulting from maximum parsimony analysis of 30 cloned 5S rDNA spacers from *Q. robur* (S*Qro*) and *Q. petraea* (S*Qpe)*. Topologies shown are the 50% majority-rule consensus of 1000 trees. Numbers above branch segments indicate the number of additional steps that are required for each resolved clade to collapse (decay values). Length of each constituent tree = 106 steps; consistency index = 0.925; retention index = 0.914. The outgroup taxon consisting of sequences from *Q. suber* (S*Qsu)* and *Q. libani* (S*Qli*) was constrained to be monophyletic. Spacer clones characteristics are given in **Table 22**.

## 5- Discussion

### a- *Quercus* 5S rDNA structure and organization

As in other higher plants, the 5S rDNA in *Q. robur* and *Q. petraea* exhibits a conventional pattern of clustered, tandemly repeated units. Our study of a two-generation full-sib pedigree in *Q. robur* indicated that these arrays probably occur at only one major locus (Trontin and Favre, unpublished results) mapped on linkage group "5" (Barrénèche *et al.* 1998) and recently located by fluorescent *in situ* hybridization (FISH) at the paracentromeric region of the second largest metacentric pair of chromosomes of 11 *Quercus* species, including *Q. robur* and *Q. petraea* (Zoldos *et al.* 1999). The size of the repeat units in *Q. robur* (317–353 bp) and *Q. petraea* (337–353 bp) is similar to that reported in other angiosperms. Most variation in the length of the 5S rDNA units occurs within the middle spacer region and is attributable to several indels 3–32 bp long easily explained by a model of replication slippage (Scoles *et al.* 1988). Such indels have also been shown to account for most of the length differences observed among spacers of different *Quercus* species (Trontin and Favre, unpublished results) and exemplified in this study by the large duplications in the *Q. suber* and *Q. libani* spacers.

The two spacer regions flanking the 5S gene (**Figure 31**) are highly conserved and probably involved in regulation of transcription by Pol III. Three putative regulatory elements commonly observed in plants were also recognized in *Quercus*. The T-rich region (i) immediately downstream of the 5S gene and duplicated in the *Q. suber* and *Q. libani* spacers is thought to facilitate efficient transcription termination (Korn 1982). The AT-rich (ii) and GC-rich motifs (iii) upstream of the 5S gene [$^{-31}$TTT(T)ATAA$^{-24}$ and $^{-15}$CGCGC$^{-11}$] have been implicated in transcriptional regulation of 5S rDNA (Venkateswarlu *et al.* 1991). It is noteworthy that the hot spot of substitutions observed at position 205 in *Q. robur* and *Q. petraea* (**Figures 29-30**) clearly defines the start of the AT-rich motif and probably reflects functional constraints at these sites.

### b- 5S rDNA molecular diversity in pedunculate and sessile oak

Unlike 5S gene, appreciable spacer polymorphism was revealed in *Q. robur* and *Q. petraea*. Although our samples sizes were small, up to 12 phenotypes solely based on major fragments were detected in each species and the corresponding relative phenotypic diversities are apparently very high (0.779 in *Q. robur* and 0.899 in *Q. petraea*). Obviously, it is difficult to compare our estimates to others published measurements of rDNA phenotypic diversity in plant owing to the different methods used to investigate the diversity and the varying sample sizes. For example, a survey on agarose gel of 92 accessions representing the major barley growing areas of the world estimated the level of rDNA relative diversity to be only 0.572 using RFLP methods (Zhang *et al.* 1992). Similarly, Karvonen and Savolainen (1993) obtained 0.585 investigating four populations (97 trees) of *Pinus sylvestris* in Finland. Comparatively, we used a much more sensitive method that surely improved phenotype

discrimination (Hauser *et al.* 1998). Most 5S rDNA spacer polymorphisms (*a–e*) were indeed only detected on conventionally crosslinked (3.3% C) non-denaturing polyacrylamide gels (8% T) and sequence analysis revealed that single nucleotide differences of homoduplexes can be resolved (e.g. the difference in mobility of electromorph *e*). However, this simple method did not reveal a number of silent substitutions as demonstrated by the relatively high nucleotide diversity within each electromorph class (especially classes *a–d*, **Table 24**).

The observed 5S rDNA spacer diversity consisted mainly into accumulation of substitutions (microheterogeneity) and very little 5S rDNA spacer length variation was discovered comparatively to other species such as in the wheat tribe *Triticeae* (Baum and Johnson 1998, Kellogg and Appels 1995, Scoles *et al.* 1988). In barley for example, up to 8 unit length groups varying in size between 233 and 607 bp were recognized (Baum and Johnson 1998). The relative length uniformity of the spacers may be related to the probable occurrence of a single 5S rDNA locus in *Q. robur* and *Q. petraea* (Barrénèche *et al.* 1998). Considerable spacer length variation was indeed generally observed within species when 5S arrays were detected at more than one chromosomal location as in the *Triticeae* (Dvorák *et al.* 1989). In these cases interlocus concerted evolution is not considered to be an important factor of evolution so that chromosome-specific arrays may evolve independently (Cronn *et al.* 1996). This is particularly outlined by the data reported on flax (Schneeberger and Cullis 1992) and Monterey pine (Gorman *et al.* 1992, Moran *et al.* 1992) which revealed a highly dispersed organization concomitant with a high level of 5S rDNA length polymorphism.

Sequence analysis particularly highlighted the high degree of intraspecific spacer microheterogeneity at the 5S rDNA locus (80% of the spacer clones are different from each other), especially within individual. The significant mean number of bands per individual revealed on conventional non-denaturing polyacrylamide gel in *Q. robur* (2.035) and *Q. petraea* (2.639) also clearly demonstrated that variant sequences are very common within array. The intraindividual mean difference among spacers (gaps excluded) was found to be 0.022 in *Q. robur* and 0.017 in *Q. petraea* (overall mean: 0.020). These values are in the lower range of nucleotide diversity estimates recently obtained over spacers in 28 diploid wheat relatives (0.000-0.092, overall mean: 0.030, Kellogg and Appels 1995) as well as in 20 diploid and allopolyploid cottons (0.020-0.154, overall mean: 0.058, Cronn *et al.* 1996). Similarly, Samuel *et al.* (1998) recently detected higher ITS sequence polymorphism (10% of nucleotide and indel variation) among rDNA repeats within one individual of *Q. pubescens*. However, it should be remembered that differences in electrophoretic mobility on crosslinked gels were found to be poorly correlated with sequence divergence and thus the mean difference values obtained in *Q. robur* and *Q. petraea* are likely to be underestimates of nucleotide diversity. Intraindividual sequence polymorphism may be interpreted as a result of mutation events across the entire array indicating that intralocus concerted evolution is not strong enough to homogenize new variants in *Quercus*. Alternatively, it can be considered that intra- or interspecific

gene flow (see below) is sufficient to keep introducing variants into an array (Kellogg and Appels 1995).

The intraspecific mean nucleotide difference among spacers (based on 10 individuals per species) is not significantly greater than the intraindividual (and probably underestimated) values in *Q. robur* (0.023 vs 0.022) and *Q. petraea* (0.023 vs 0.017). Such a situation may simply results from the process of fixation of variants in a population (molecular drive, Dover 1982) as a consequence of relatively weak concerted evolution in oak. A slow rate of interchromosomal homogeneisation relative to the fast rate of chromosome turnover within a population (generation time) could indeed ensure that, at any given time, the differences between individuals would be kept to a minimum with respect to the changing sequence composition of the family (Dover 1982). Alternatively, high variation within an individual indicated that there is little selection (i.e. elimination of polymorphism) within array and this can be easily explained by the buffer effect of high copy number against the deleterious effects of mutation in any copy (Kellogg and Appels 1995), but similar levels of intraspecific and intraindividual molecular diversity suggest that some spacer selection may operate at the species level. This limitation of intraspecific polymorphism can be directly related to the maintenance of effective transcriptional regulation of the 5S gene involving some spacer regions (as discussed above). Another valid explanation implies that strong selection for functional gene product can also partly affect the associated spacer region. Kellogg and Appels (1995) and Cronn *et al.* (1996) have recently demonstrated in the *Triticeae* and *Gossypium* species that despite similar level of intraindividual nucleotide diversity in both gene and spacer, there are considerably fewer fixed difference between species in the gene than in the spacer. They concluded that selective mechanisms intensively acted on the gene region to eliminate mutations. Similarly, the very low nucleotide diversity observed in our *Quercus* study for the gene (0.003) relative to that for the spacer (0.023) further supports intense gene selection at the species level. Kellogg and Appels (1995) excluded that such selection also acted on the spacer because they observed higher interspecific than within species variation. However, they generally studied only one individual per species and thus did not distinguish the intraindividual and intraspecific levels. Moreover, the amount of spacer variation between some of the species they investigated is the same as that within individual, suggesting as in *Q. robur* and *Q. petraea* that evolutionary forces acting on the gene could also partially affect the linked spacer.

c- <u>The low differentiation of pedunculate and sessile oak at the 5S rDNA locus</u>

Despite the clearly established morphological distinctiveness between *Q. robur* and *Q. petraea* (Dupouey and Badeau 1993, Feuillat *et al.* 1997) and an appropriate sampling scheme for the identification of discriminant markers (most of the trees were collected in species pairs across the distribution range), our study of 5S rDNA variation confirms that the two species are very similar genetically. As most other molecular studies based on isozymes (Kremer *et al.* 1991, Zanetto *et al.*

1994), chloroplast and mitochondrial DNA (Dumolin-Lapègue *et al.* 1998), 18S-5.8S-26S rDNA (Petit *et al.* 1993b) and RAPD markers (Moreau *et al.* 1994, Bodénès *et al.* 1997a), we did not find any qualitative 5S rDNA difference between pedunculate and sessile oak (specific variant, fixed nucleotide substitution) but only quantitative differences. As already depicted by allozyme variations (Zanetto *et al.* 1994), similarities between RAPD profiles (Moreau *et al.* 1994) and nucleotide diversity inferred from RAPD (Bodénès *et al.* 1997a), the Shannon's information measure of 5S rDNA phenotypic diversity indeed indicated that sessile oak is more variable than pedunculate oak (0.899 vs 0.779). Moreover, our results also provided new nuclear markers exhibiting significant frequency differences between the two species (classes *e* and *f*), but studies at the population level are clearly needed to infer their importance.

The lack of clear qualitative nuclear ribosomal DNA differences between the two species was highly unexpected. Because of concerted evolution of repeats superimposed with selection and genetic drift (Arnheim 1983, Dover 1982) tandemly repeated DNA sequences are indeed very often considered to evolve faster than other genomic sequences [e.g. repetitive families in *Gossypium* species (Zhao *et al.* 1998)] and thus constitute excellent species markers. Considering the relatively high level of intraspecific molecular diversity observed in 5S rDNA (preceding paragraph) and rDNA (Petit *et al.* 1993b), it is therefore quite surprising that the concerted homogenizing forces (primarily unequal crossing-over and gene conversion) failed to convert even a small part of the observed polymorphism into fixed differences between species.

To date, fixed rDNA differences between oaks have only been observed at higher taxonomic level. Investigating populations by restriction site analysis, Whittemore and Schaal (1991) found that a 7.8 kb length variant is specific to a group of 3 species native to the eastern United States (*Q. alba*, *Q. macrocarpa*, and *Q. michauxii*). Similarly, in their infrageneric phylogenetic study of *Quercus* based on ITS sequences from nuclear rDNA, Samuel *et al.* (1998) reported evidence of extensive divergence between relatively distant taxa. However, the 7 base changes they detected between two ITS clones from *Q. robur* and *Q. petraea* are unlikely to represent fixed differences in regard to our own results (we have detected up to 15 substitutions between 5S rDNA spacers of the two species and none of them was found to represent a fixed difference). This particularly highlights the need of thorough intraspecific investigations prior to studies of interspecific relationships, especially in oaks (Whittemore and Schaal 1991).

How could be interpreted such a low differentiation at the 5S rDNA and others investigated nuclear loci ? Extensive interspecific hybridization and introgression between pedunculate and sessile oak (Bacilieri *et al.* 1996, Steinhoff 1998) may be advocated as a major cause. Such interfertility has been shown to commonly occur between sympatric oak species (Rushton 1993, Whittemore and Schaal 1991) and interpreted as either reflecting recent divergence of the species (i.e. they have not completely diverged and still share a large part of the genome) or resulting from prolonged and

repeated contacts in the glacial refuges during the quaternary history (Huntley and Birks 1983). In addition, it can be stressed that the number of generations since the last postglacial recolonization by oaks took place (7000–15000 years ago) is probably too low (50–100 generations of sexual reproduction) to have generated any detectable fixed nuclear DNA difference, even in the case of sympatric species with more effective reproductive barriers. In these schemes, fragments with significant frequency differences between pedunculate and sessile oak (such as 5S rDNA spacer variants *e* and *f*) may either originate from a common ancestor and be progressively selected or result from mutations that occurred after divergence and were subsequently exchanged. The basal position of variant class *f* in the gene tree reconstruction (2 synapomorphies with the congeneric outgroup species *Q. suber* and *Q. libani*) is congruent with the first proposal. In contrast, the phylogenetic characteristics of variant class *e* indicated more recent interspecific gene flow. Thus, pedunculate and sessile oaks may have relatively recent and interrelated evolutionary history.

Bodénès *et al.* (1997a) have recently provided extensive evidence using RAPD data that the genomic regions differentiating these two closely related species are relatively rare and exhibited higher within species nucleotide diversity (estimated to be on average 0.031 from a set of 36 discriminant fragments) than other non informative regions (on average 0.023 from a set of 376 non discriminant fragments). Remarkably, the mean nucleotide difference (gaps excluded) among 5S rDNA spacers sequenced in *Q. robur* and *Q. petraea* (0.023 on average) is very similar to the estimates of nucleotide diversity obtained by Bodénès *et al.* (1997a) from non discrimant RAPD fragments, suggesting that the 5S rDNA locus is not in the genomic regions currently responsible for nucleotide divergence between the two species. Nevertheless, it is noteworthy that among the SCARs (sequence characterized amplified regions) obtained from the discriminant RAPD markers, one (B12-500) exhibited high sequence identity with ribosomal RNA genes of maize and yellow lupin (Bodénès *et al.* 1997a).

***Acknowledgements:*** We are indebted to A. Kremer (INRA Bordeaux) for providing most of the material used in this study and for helpful discussions.

**6- Conclusion : rappel des principaux résultats obtenus**

Cette étude a permis de montrer que l'ADNr 5S nucléaire des chênes pédonculé et sessile est, comme chez beaucoup d'espèces, composé de milliers d'unités répétées en tandem qui mesurent ici de 317 à 353 pb. Peu de variations ont été révélées sur gel d'agarose (4 profils électrophorétiques dont l'un, composé uniquement d'unités de 350 pb, est observé chez plus de 85 % des arbres examinés) mais une importante microhétérogénéité de longueur et de séquence des unités, en particulier des espaceurs, devient apparente après migration sur gel d'acrylamide non dénaturant dans des conditions résolutives (25 cm de migration, gel à 8 %).

Comme la plupart des travaux antérieurs conduits à l'aide d'autres types de marqueurs cytoplasmiques ou nucléaires (Kremer *et al.* 1991, Ferris *et al.* 1993, Petit *et al.* 1993a,b, Müller-Starck *et al.* 1993, Zanetto *et al.* 1994, Moreau *et al.* 1994, Bodénès *et al.* 1997a,b, Dumolin-Lapègue *et al.* 1998), et malgré l'importante diversité moléculaire des espaceurs de cette famille multigénique (diversité phénotypique de Shannon, divergence nucléotidique moyenne), aucune différence qualitative n'a pu être mise en évidence entre ces deux espèces proches, sympatriques et interfertiles. Toutefois, la diversité phénotypique calculée à partir des profils électrophorétiques sur gel d'acrylamide (7 niveaux de migration correspondant aux catégories d'espaceurs ou classes électrophorétiques *a-g*), apparaît nettement plus élevée chez le chêne sessile (0,899) que chez le chêne pédonculé (0,779) confirmant ainsi les résultats obtenus à l'aide de marqueurs isoenzymatiques (Zanetto *et al.* 1994) ou RAPD (Moreau *et al.* 1994, Bodénès *et al.* 1997a). De plus, une catégorie d'espaceurs (*e*) qui présente dans l'échantillon étudié une différence de fréquence significative entre les deux espèces (33,5 % des chênes sessiles et 3,5 % des chênes pédonculés) a été identifiée. Les arbres porteurs de ce marqueur sont en majorité (9/11) distribués au sud de l'Europe, près des refuges glaciaires présumés (Ferris *et al.* 1993, 1998, Dumolin-Lapègue *et al.* 1997). Deux autres catégories d'espaceurs à faible fréquence ont enfin été observées soit uniquement chez le chêne sessile (*f* : 8,3 %), soit uniquement chez le chêne pédonculé (*g* : 5,2 %).

L'organisation structurale et fonctionnelle des unités d'ADNr 5S des chênes pédonculé et sessile est similaire à celle des autres plantes angiospermes. La taille du gène est très peu variable (119-120 pb) et 4 des 6 clones séquencés dans cette étude à partir de 6 individus de différentes provenances (3 par espèce) sont identiques. La séquence consensus ne diffère qu'à 2 positions nucléotidiques (paire de base au centre de la région bicaténaire II (impliquée dans la structure secondaire des ARNr 5S, **Figure 17**) des gènes d'autres angiospermes telles que les cotonniers. L'estimation de la diversité nucléotidique de cette région transcrite est toutefois relativement élevée ($\pi = 0{,}014$) car l'un des gènes séquencés s'est avéré assez divergent. Ce clone est probablement inopérant (pseudogène) car sa taille est seulement de 119 pb et il comporte en outre 3 mutations ponctuelles dont deux sont situées dans les zones importantes d'interaction du gène (boîte C) ou de l'ARNr 5S avec le facteur de transcription TFIIIA (voir **Partie I-3** du **Chapitre II** et **Figure 17**). La diversité nucléotidique diminue fortement lorsque ce probable pseudogène est exclu du calcul ($\pi = 0{,}003$).

Les régions de l'espaceur qui bordent le gène (environ 25 pb) sont relativement bien conservées chez tous les chênes examinés dans ce travail (*Q. robur*, *Q. petraea*, *Q. suber*, *Q. libani*) et probablement impliquées dans la régulation de la transcription par la Pol III (voir la **Partie I-1** du **Chapitre II**). Nous avons ainsi identifié trois motifs régulateurs classiquement reconnus chez les plantes : un motif poly T immédiatement en aval du gène (responsable de la terminaison de la transcription) et les motifs riches en AT (-24 pb en amont du gène) et en GC (-11 pb en amont du gène) qui jouent probablement un rôle important au cours de l'initiation de la transcription.

La région centrale de l'espaceur est en revanche beaucoup plus variable et largement responsable des différences de longueur (197-233 pb) et de séquence (variation nucléotidique moyenne : 0,023) observées entre les espaceurs des 7 catégories (*a-g*) identifiées sur gel d'acrylamide. Les petits motifs répétés présents dans cette zone sont en effet probablement impliqués dans des phénomènes de glissement de brins (voir la **Partie II-2-c-α** du **Chapitre I**) qui se traduisent par des délétions et/ou des insertions de plus ou moins grande amplitude (3-32 pb dans notre échantillon). Les mutations ponctuelles présentent une distribution similaire avec cependant un "point chaud" près de la position 205 qui précède immédiatement le motif régulateur riche en AT et fournit ainsi une bonne définition topographique de l'extrémité 3' conservée de l'espaceur (25 pb).

Les mécanismes de l'évolution concertée n'ont apparemment pas suffisamment joué pour permettre l'homogénéisation complète de l'ADNr 5S, ni chez le chêne pédonculé, ni chez le chêne sessile. La variation intragénomique moyenne (0,020) estimée à partir de 2-3 séquences par individu (4 chênes pédonculés et 3 chênes sessiles) est très similaire à la variation intraspécifique moyenne (0,023) déduite de l'analyse de 30 séquences issues d'une vingtaine d'individus (10 chênes pédonculés et 10 chênes sessiles). Cette faible différence peut être interprétée comme résultant du processus de "molecular drive" décrit par Dover (1982). Dans cette hypothèse, le brassage chromosomique intervenant entre les individus à chaque génération serait considérablement plus rapide que l'homogénéisation intragénomique (évolution concertée) des unités d'ADNr 5S (voir **Partie II-3-d** du **Chapitre I**) qui se réaliserait à une vitesse très faible, approximativement égale à celle de la spéciation (Kellogg et Appels 1995, Cronn *et al.* 1996, voir la **Partie III-4-b** du **Chapitre II**). Dans une telle situation, les niveaux de diversité intragénomique et intraspécifique attendus sont équivalents. Une autre explication envisageable repose sur l'existence d'une sélection partielle s'exerçant sur les espaceurs, soit directement pour maintenir la fonction des éléments régulateurs (poly T, régions riches en AT et en GC), soit indirectement en raison de sa liaison avec le gène (cible de mécanismes sélectifs intenses, Kellogg et Appels 1995, Cronn *et al.* 1996).

La plupart des espaceurs (67 %) ont une taille de 230 pb. Il y a donc peu de variations de longueur au sein et entre les 7 classes électrophorétiques *a* (230-233 pb), *b* (230 pb), *c* (229-230 pb), *d* (230 pb), *e* (230-231 pb), *f* (217 pb) et *g* (197-198 pb) observées. La diversité nucléotidique est très faible au sein des classes *e-g* qui présentent des différences de fréquence entre chêne pédonculé et chêne sessile ($\pi$ = 0,004 en moyenne). Elle est beaucoup plus importante dans les classes *a-d* ($\pi$ = 0,027 en moyenne). Aucun polymorphisme commun n'a pu être détecté entre les différents membres séquencés de chacune de ces 4 classes très hétérogènes. Cela indique que la mobilité électrophorétique des fragments amplifiés sur gel d'acrylamide n'est que faiblement corrélée à la divergence nucléotidique. En revanche, des polymorphismes partagés au sein des classes *e* (T en position 110), *f* (délétion de 13 pb, G et T en positions 145 et 191) et *g* (délétion de 32 pb) ont été relevés. L'analyse phylogénétique de ces allèles a été précisée après recherche de séquences adaptées pour constituer un groupe externe à la section *Quercus* à laquelle appartiennent les chênes pédonculé

et sessile. Notre choix s'est porté sur des espaceurs isolés d'un chêne liège (*Q. suber*, un clone de 306 pb) et d'un chêne du Liban (*Q. libani*, deux clones de 242 pb et 323 pb) appartenant à la section voisine *Cerris* (**Tableau 17**) car l'alignement est aisé (81 % de similarité) et surtout complet (pas de délétions majeures entre les espaceurs de ces deux groupes). La comparaison des séquences des groupes *Cerris* et *Quercus* a permis de révéler que les bases G (position 145) et T (position 191) caractéristiques de la catégorie d'espaceurs *f* (uniquement détectée chez le chêne sessile) peuvent être considérés comme des caractères ancestraux.

## III- Héritabilité et localisation génétique des unités d'ADNr 5S nucléaire chez *Quercus robur* L.

### 1- Introduction

Comme la plupart des familles dominantes de séquences répétées en tandem du génome des eucaryotes (**Partie II-2** du **Chapitre I**), les séries d'unités (= d'homoallèles et/ou d'hétéroallèles selon qu'elles sont ou non identiques) de l'ADNr 5S nucléaire peuvent en théorie fournir de nombreux marqueurs utiles pour les études de diversité moléculaire et de différenciation génétique des populations (Avise 1994). Par comparaison aux séquences uniques, les possibilités de variation allélique de ces familles au sein du génome d'un individu sont en effet considérables (**Figure 1**). Le polymorphisme intragénomique (ou allélique) peut ainsi résider non seulement dans les séries (polymorphisme intrachromosomique ou *cis*), mais aussi entre les séries disposées sur des chromosomes homologues ou non homologues lorsque la famille occupe plusieurs loci (polymorphisme interchromosomique ou *trans*). En pratique, ce multiallélisme *cis* et *trans* qui caractérise les séquences répétées en tandem peut être pour plusieurs raisons à l'origine de sérieuses difficultés d'analyse des données et d'interprétation des résultats (Avise 1994).

Tout d'abord, l'indépendance d'évolution des différents membres d'une famille n'est pas assurée en raison de l'existence du phénomène d'évolution concertée (Zimmer *et al.* 1980, Arnheim 1983, Ohta et Dover 1983). Selon la fréquence et l'efficacité des mécanismes d'homogénéisation intragénomique (intra- et interchromosomiques) comme le crossing-over inégal ou la conversion génique, la distinction usuelle entre séquences paralogues et orthologues (requise pour les études phylogénétiques) peut ainsi se révéler très difficile voire impossible (**Figure 19**). Avant toute interprétation de la variation observée à l'aide de ce type de marqueur, il est donc nécessaire de réaliser de solides évaluations de la diversité moléculaire dans les populations examinées afin d'apprécier l'impact éventuel de l'évolution concertée (Buckler *et al.* 1997, Baum et Johnson 1998).

Ensuite, les marqueurs issus de la fraction répétée du génome peuvent présenter une hérédité non mendélienne liée aux mécanismes de l'évolution concertée. Contrairement aux séquences uniques du génome qui sont *a priori* seulement soumises aux processus de sélection naturelle (évolution adaptative) et de dérive génétique (évolution non adaptative), l'augmentation progressive de la fréquence puis la fixation d'une unité variante (ou variant allélique) plus ou moins simultanément dans une famille et dans une population ("molecular drive", Dover 1982) est en effet essentiellement provoquée par le jeu des échanges moléculaires (recombinaisons) aléatoires ou dirigés entre chromosomes homologues ou non homologues et entre chromatides soeurs (voir la **Partie II-3** du **Chapitre I**).

Enfin, l'impossibilité d'accéder rapidement et simplement au génotype d'un individu constitue un obstacle important à l'exploitation des marqueurs fournis par les familles de séquences répétées en génétique des populations. En raison du multiallélisme, les profils électrophorétiques (RFLP, PCR)

sont généralement complexes (présence de nombreuses bandes majeures et mineures) et la distinction entre polymorphismes *cis* et *trans*, requise pour l'identification des homozygotes et des hétérozygotes, devient dès lors très difficile, particulièrement lorsque plusieurs loci sont impliqués (Clegg 1989, Karvonen et Savolainen 1993).

Dans la **Partie II** précédente, l'analyse du polymorphisme des espaceurs de l'ADNr 5S a été réalisée à partir d'un dispositif expérimental représentatif de l'aire de répartition des chênes pédonculé (*Quercus robur* L.) et sessile (*Quercus petraea* (Matt.) Liebl.). Nous avons à cette occasion détecté une forte diversité phénotypique qui repose sur l'identification de 7 classes principales (*a-g*) de variants électrophorétiques. Le fait que la plupart des arbres testés (90/93) ne présentent qu'un ou deux de ces fragments majeurs dans la zone des monomères suggère que le déterminisme génétique de leur héritabilité est assez simple, impliquant un seul locus et au moins 7 allèles différents. Toutefois, l'identification des hétérozygotes et donc l'estimation de la diversité génétique sur cette base demeure délicate car les analyses de séquences ont révélé une diversité nucléotidique assez élevée au sein de certaines de ces classes (*a-d* en particulier). D'autre part, les phénotypes comportent souvent des bandes supplémentaires de moindre intensité et plus de deux fragments majeurs ont été amplifiés chez 3 individus.

Dans cette troisième partie, nous avons donc cherché à déterminer le mode d'hérédité des variants électrophorétiques obtenus après amplification des espaceurs de l'ADNr 5S chez le chêne pédonculé. Pour ce faire, nous avons analysé une descendance hybride F1 issue d'un croisement intraspécifique contrôlé de chênes pédonculés impliqué dans un projet de cartographie génétique dirigé par A. Kremer (INRA de Bordeaux). Les deux phénotypes parentaux se sont avérés suffisamment différents pour permettre (i) d'étudier l'héritabilité de 3 des 7 variants majeurs identifiés précédemment (*a*, *b* et *d*) et de plusieurs variants mineurs, (ii) d'estimer le nombre de loci, (iii) d'établir des génotypes et donc de distinguer le polymorphisme *cis* du polymorphisme *trans*, et enfin (iv) d'accéder à la localisation génétique de l'ADNr 5S (Barrénèche *et al.* 1998, voir **Annexe**).

## 2- Matériel et méthodes

### a- Matériel végétal

L'héritabilité des marqueurs de l'ADNr 5S a été étudiée à partir d'une descendance F1 de chênes pédonculés (94 individus) issue d'un croisement contrôlé d'arbres adultes (plus d'une centaine d'années). Le parent mâle (A4) est localisé près d'Arcachon et le parent femelle (3P) au sein de la station de recherche forestière de l'INRA de Bordeaux (Pierroton) distante d'une quarantaine de km. Le croisement a été réalisé au printemps 1992 selon la méthode décrite par Zanetto *et al.* (1996). Les extraits d'ADN génomique, obtenus à partir de feuilles selon la procédure utilisée par Moreau *et al.* (1994), nous ont été fournis par A. Kremer.

b- Amplification des espaceurs de l'ADNr 5S

Le couple d'amorces universelles utilisé pour amplifier les espaceurs de l'ADNr 5S nucléaire a été directement sélectionné à partir de la séquence consensus des gènes 5S de plantes supérieures disponibles (Szymanski *et al.* 1999). Il s'agit des amorces directe (F) et indirecte (B) suivantes, situées aux extrémités du gène 5S (voir **Figure 21**) :

F : 5' TGG GAA GTC CTC GTG TTG CA 3' (positions 96 à 115 de la séquence consensus)
B : 5' ATT AGT GCT GGT ATG ATC GC 3' (positions 24 à 5 de la séquence consensus)

Nous avons aussi utilisé les amorces P1 et P2 décrites dans la **Partie II-3-b** de ce **Chapitre III** qui permettent de cloner les espaceurs amplifiés. Elles dérivent respectivement des précédentes par introduction d'un site de restriction par l'endonucléase *Eco*R I (substitution G→T en position 102) ou *Sac* I (substitution T→A en position 19 et G→C en position 15).

Chaque amplification a été réalisée dans un volume réactionnel total de 25 µl. Le mélange contient 0,1-1 ng d'ADN génomique, 0,2 µM de chacune des amorces F et B ou P1 et P2, 200 µM de chacun des 4 désoxyribonucléotides (dATP, dCTP, dGTP et dTTP), 20 mM de tampon Tris-HCl (pH 8,4), 50 mM de KCl, 2 mM de $MgCl_2$ et 0,6 unités d'ADN polymérase *Taq* (Gibco BRL). Après une étape de dénaturation à 96°C pendant 3 min, les 96 échantillons (2 arbres parentaux et 94 descendants) subissent simultanément (en microtubes de 200 µl sans huile minérale) 30 cycles d'amplification à l'aide d'un thermocycleur PTC-100 muni d'un couvercle chauffant (MJ Research, Inc). Un cycle consiste en une étape de dénaturation à 94°C pendant 40 s, une étape d'hybridation des amorces à 55°C (F/B) ou 53°C (P1/P2) pendant 1 min et une étape d'élongation des fragments par l'ADN polymérase à 72°C pendant 40 s.

c- Séparation et détection des produits de l'amplification

Après un premier contrôle sur gel d'agarose à 1,5 %, l'électrophorèse des produits de PCR a été réalisée en conditions non dénaturantes sur gel de polyacrylamide (8 % T et 3,3 % C) de 0,75 mm d'épaisseur confectionné dans du tampon TBE 1X (90 mM de Tris base, 90 mM d'acide borique, 2 mM d'EDTA pH 8,0, eau ultrapure). La polymérisation du gel à température ambiante (pendant au moins 2 h) est initiée sans dégazage par ajout de 0,05 % (p/v) de persulfate d'ammonium (APS) et 0,1 % de TEMED. Après mélange avec 10 % (v/v) de tampon de charge [50 % de glycérol, 0,4 % (p/v) de bleu de bromophénol, 0,4 % (p/v) de xylène cyanol], 2 à 5 µl de produits de PCR sont rapidement déposés dans les 68 puits d'un gel de 38 cm de largeur et 30 cm de longueur (cuve SequiGen, Biorad) à l'aide d'une pipette multicanaux (2 gels pour 96 échantillons). Plusieurs marqueurs de taille moléculaire (DNA VI, ϕX174 et "100 bp DNA ladder", Gibco BRL) sont en outre ajoutés sur chaque gel. L'électrophorèse est ensuite réalisée pendant 6 h à 600 V constants (soit au total 2500 V.h).

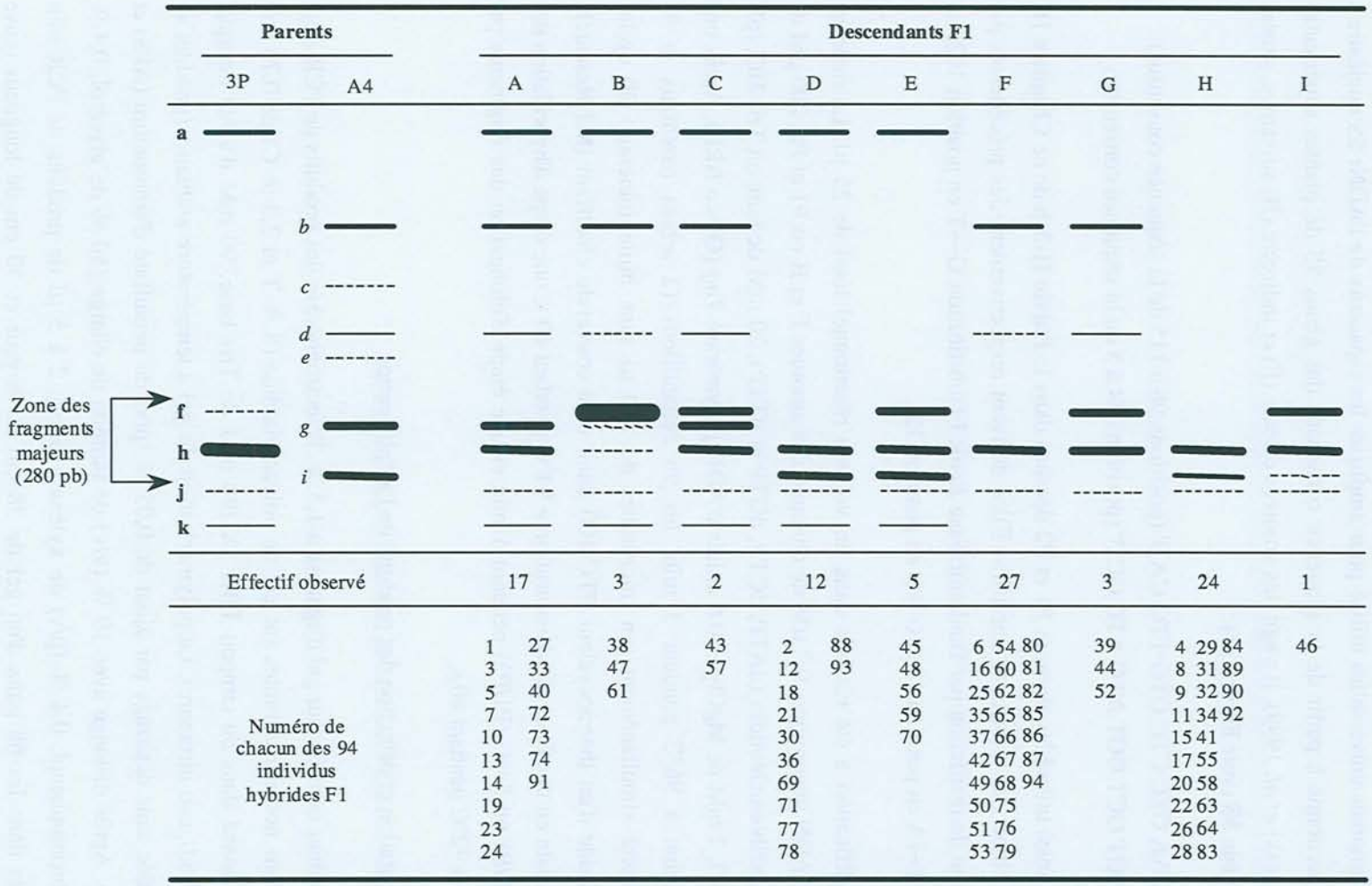



**Figure 33**. Phénotypes (A-I) observés sur gel d'acrylamide (zone des monomères : 250-350 pb) après amplification des espaceurs de l'ADNr 5S (amorces universelles F et B) dans une population F1 de chêne pédonculé (94 descendants) issue d'un croisement contrôlé entre les individus 3P (parent femelle) et A4 (parent mâle). L'intensité des différents fragments amplifiés chez les parents et les descendants (a-k) est figurée par l'épaisseur variable des traits.

Dans ces conditions, les fragments amplifiés de 250-300 pb (zones des monomères) parcourent une distance d'au moins 20 à 25 cm et sont relativement bien séparés. La détection des fragments est effectuée soit sous UV après incubation du gel pendant 30 min dans une solution de bromure d'éthidium à 0,5 µg/ml, soit par coloration au nitrate d'argent en suivant la procédure de Bassam *et al.* (1991). Les profils électrophorétiques des deux arbres parentaux (3P et A4) et des 94 descendants ont été déterminés sur la base de 2 amplifications différentes et 4 passages sur gels d'acrylamide colorés au bromure d'éthidium (2 passages) ou au nitrate d'argent (2 passages).

**3- Résultats**

a- Phénotypes des arbres 3P et A4

Après amplification des espaceurs à l'aide des amorces universelles F et B et électrophorèse sur gel d'agarose (1,5 à 2 %) aucun polymorphisme exploitable n'a pu être mis en évidence entre les deux arbres parentaux 3P et A4, quel que soit le niveau considéré (zones des monomères, dimères et trimères), même après digestion des produits de PCR par différentes enzymes de restriction (*Alu* I, *BamH* I, *Msp* I, *Pst* I, *Taq* I et *Tru* 9I). En revanche, sur gel d'acrylamide en conditions non dénaturantes, les deux individus ne présentent aucun marqueur commun dans la zone des monomères comprise approximativement entre 250 pb et 350 pb (**Figure 33**). Le phénotype du parent femelle (3P) est composé d'une bande majeure (**h**) et de 4 bandes de moindre intensité (**a** = **k** > **j** > **f**). Le phénotype du parent mâle (A4) comporte quant à lui 2 bandes majeures d'intensité similaire (*g* et *i*) ainsi que 4 bandes mineures (*b* > *d* > *c* = *e*). Les amplifications réalisées à l'aide des amorces P1 et P2 conduisent à des profils identiques et permettent de vérifier que les fragments majeurs *g*, **h** et *i* correspondent respectivement aux trois niveaux de variants électrophorétiques *a*, *b* et *d* identifiés précédemment (voir la **Figure 28**). La comparaison des séquences d'un clone P1/P2 de chacun de ces 3 variants majeurs (S*Qro*A4*a*, S*Qro*A4*d*, S*Qro*3P*b*) ne révèle aucune différence de longueur (230 pb). Seules quelques substitutions peuvent être détectées, en nombre équivalent (7 à 11), et avec des taux de transitions/transversions similaires (**Tableau 25**). Chaque clone ne présente par ailleurs que 2 à 6 différences nucléotidiques par rapport à la séquence consensus des espaceurs établie dans la **Partie II** de ce **Chapitre III** (consulter l'alignement de la **Figure 29**). Ils sont donc fortements apparentés.

b- Phénotypes des hybrides F1

Neuf phénotypes différents (A-I) sont observés dans la descendance (**Figure 33**) et tous sont constitués d'une combinaison des bandes parentales. Aucune bande nouvelle n'a été identifiée. Quatre phénotypes (D, A, H et F) sont prépondérants (12 à 27 individus) ; les 5 autres (I, C, B, G et E) ne sont représentés que par 1 à 5 individus. L'analyse de la ségrégation des marqueurs parentaux est indiquée dans le **Tableau 26**. Les bandes *c* et *e* de l'arbre A4 présentent une intensité trop faible après coloration au nitrate d'argent pour permettre de déterminer leur fréquence dans la descendance. Les marqueurs **h**, **a**, **k**, **j**, (3P) et *i*, *b*, *d* (A4) ségrègent de façon mendélienne tandis qu'une sévère

distorsion est observée pour la bande mineure **f** (3P) et la bande majeure *g* (A4). La ségrégation des bandes **a** et **k** d'une part (3P), et *b* et *d* d'autre part (A4), apparaît liée.

**Tableau 25.** Comparaison des séquences nucléotidiques d'un clone de chacun des 3 fragments majeurs détectés chez les chênes pédonculés 3P (variant **h**) et A4 (variants *g* et *i*)[a].

| Variant | Clone (taille en pb)[b] | S*Qro*3Pb (230) | S*Qro*A4a (230) | S*Qro*A4d (230) |
|---|---|---|---|---|
| **h** | S*Qro*3Pb (230) | *2 (1,00)* | | |
| *g* | S*Qro*A4a (230) | 8 (1,00) | *6 (1,00)* | |
| *i* | S*Qro*A4d (230) | 7 (1,33) | 11 (1,20) | *5 (1,50)* |

[a] Diagonale (en Italique) : nombre de substitutions et taux de transitions/transversions (entre parenthèses) de chaque clone par rapport à la séquence consensus des espaceurs établie précédemment (voir **Figure 29**). Sous la diagonale : nombre de substitutions/tranversions (entre parenthèses) pour les comparaisons deux à deux des 3 clones.

[b] Nomenclature des clones selon le **Tableau 20**.

**Tableau 26.** Analyse de la ségrégation des fragments parentaux (classés par intensité décroissante) dans la descendance hybride F1 du croisement 3P X A4 (94 individus).

| Fragment parental | $F_{obs}(\%)^a$ | $F_{th}(\%)^b$ | $\chi^2(1ddl)^c$ | Interprétation génétique |
|---|---|---|---|---|
| Femelle 3P | | | | |
| **h** | 100 | 100 | 0 | Un locus homozygote pour **h** |
| **a** | 41,5 | 50 | 2,72 (p > 0,05) | Un locus hétérozygote pour **a** |
| **k** | 41,5 | 50 | 2,72 (p > 0,05) | Un locus hétérozygote pour **k** |
| **j** | 100 | 100 | | Un locus homozygote pour **j** |
| **f** | 14,9 | 50 | 46,34 (p > 0,005) | Distorsion de ségrégation |
| Mâle A4 | | | | |
| *g* | 23,4 | 50 | 26,59 (p > 0,005) | Distorsion de ségrégation |
| *i* | 44,7 | 50 | 1,06 (p > 0,10) | Un locus hétérozygote pour *i* |
| *b* | 55,3 | 50 | 1,06 (p > 0,10) | Un locus hétérozygote pour *b* |
| *d* | 55,3 | 50 | 1,06 (p > 0,10) | Un locus hétérozygote pour *d* |
| *c* | 0 | / | / | / |
| *e* | 0 | / | / | / |

[a] Fréquence observée

[b] Fréquence théorique

[c] Test du $\chi^2$ à un degré de liberté

c- Analyse génétique des marqueurs à hérédité mendélienne **h, a, k, j** (3P) et *i, b, d* (A4)

α- Génotype du parent femelle 3P

L'examen de la **Figure 33** montre que les groupes de marqueurs **h, a, k, j** (phénotypes A-E ; 39 individus) d'une part, et **h, j** (phénotypes F-I ; 55 individus) d'autre part, ségrègent dans la descendance hybride F1 selon un rapport 1:1 ($\chi^2_{(1\ ddl)} = 2,72$ ; $p > 0,05$). Ce résultat est compatible avec un modèle génétique impliquant l'existence d'un locus composé des séries alléliques composites

ou mosaïques (**hakj**) et (**hj**). L'arbre 3P présente donc un génotype globalement hétérozygote (**hakj**)/(**hj**) pour ce locus. Il faut cependant noter qu'il est homozygote pour les variants alléliques **h** (majeur) et **j** (mineur).

**Tableau 27.** Génotypes proposés pour les 94 hybrides F1 du croisement 3P X A4 sur la base des fragments à hérédité mendélienne **h**, **a**, **k**, **j** (3P) et *i*, *b*, *d* (A4).

| Génotype | Phénotypes | $N_{obs}$ [a] | $N_{th}$ [b] | N° des hybrides F1 |
|---|---|---|---|---|
| (**hakj**)/(*bd*) | A, B, C | 22 | 23,5 | 1, 3 ,5 ,7 ,10 ,13 ,14 , 19, 23, 24, 27, 33, 38, 40, 43, 47, 57, 61, 72, 73, 74, 91 |
| (**hakj**)/(*i*) | D, E | 17 | 23,5 | 2, 12, 18, 21, 30, 36, 45, 48, 56, 59, 69, 70, 71, 77, 78, 88, 93 |
| (**hj**)/(*bd*) | F, G | 30 | 23,5 | 6, 16, 25, 35, 37, 39, 42, 44, 49, 50, 51, 52, 53, 54, 60, 62, 65, 66, 67, 68, 75, 76, 79, 80, 81, 82, 85, 86, 87, 94 |
| (**hj**)/(*i*) | H, I | 25 | 23,5 | 4, 8, 9, 11, 15, 17, 20, 22, 26, 28, 29, 31, 32, 34, 41, 46, 55, 58,63, 64, 83, 84, 89, 90, 92 |

[a] Effectif observé
[b] Effectif théorique

β- Génotype du parent mâle A4

La même analyse fait ressortir que le groupe de marqueurs liés *b*, *d* (phénotypes A-C, F, G ; 52 individus) et le marqueur *i* (phénotypes D, E, H, I ; 42 individus) ségrègent aussi dans la descendance hybride F1 (**Figure 33**) selon un rapport 1:1 ($\chi^2$ (1 ddl) = 1,06 ; $p > 0,10$). Ce résultat est donc compatible avec un modèle génétique impliquant un unique locus composé d'une série allélique apparemment homogène (*i*) et d'une série allélique composite (*bd*). Le génotype de l'arbre A4 [(*bd*)/(*i*)] est donc globalement hétérozygote pour ce locus. Il est intéressant de remarquer qu'il est également hétérozygote pour le variant allélique majeur *i*.

χ- Génotypes proposés pour les descendants

Sur la seule base des fragments à hérédité mendélienne, il est possible de regrouper les 9 phénotypes hybrides F1 (**Figure 33**) en 4 classes génotypiques (**Tableau 27**) correspondant aux différentes combinaisons des 4 séries alléliques (= 4 allèles différents au sens génétique du terme) fournies par les génotypes parentaux 3P [(**hakj**) et (**hj**)] et A4 [(*bd*) et (*i*)]. La ségrégation de ces marqueurs codominants dans la descendance suit le rapport attendu 1:1:1:1 (22:17:30:25, $\chi^2$ (3 ddl) = 3,79, $p > 0,1$) et suggère un déterminisme génétique simple gouverné par un seul locus hétérozygote

chez les deux parents étudiés. Nous proposons la nomenclature 5SA pour faire référence à ce locus d'ADNr 5S chez le chêne pédonculé.

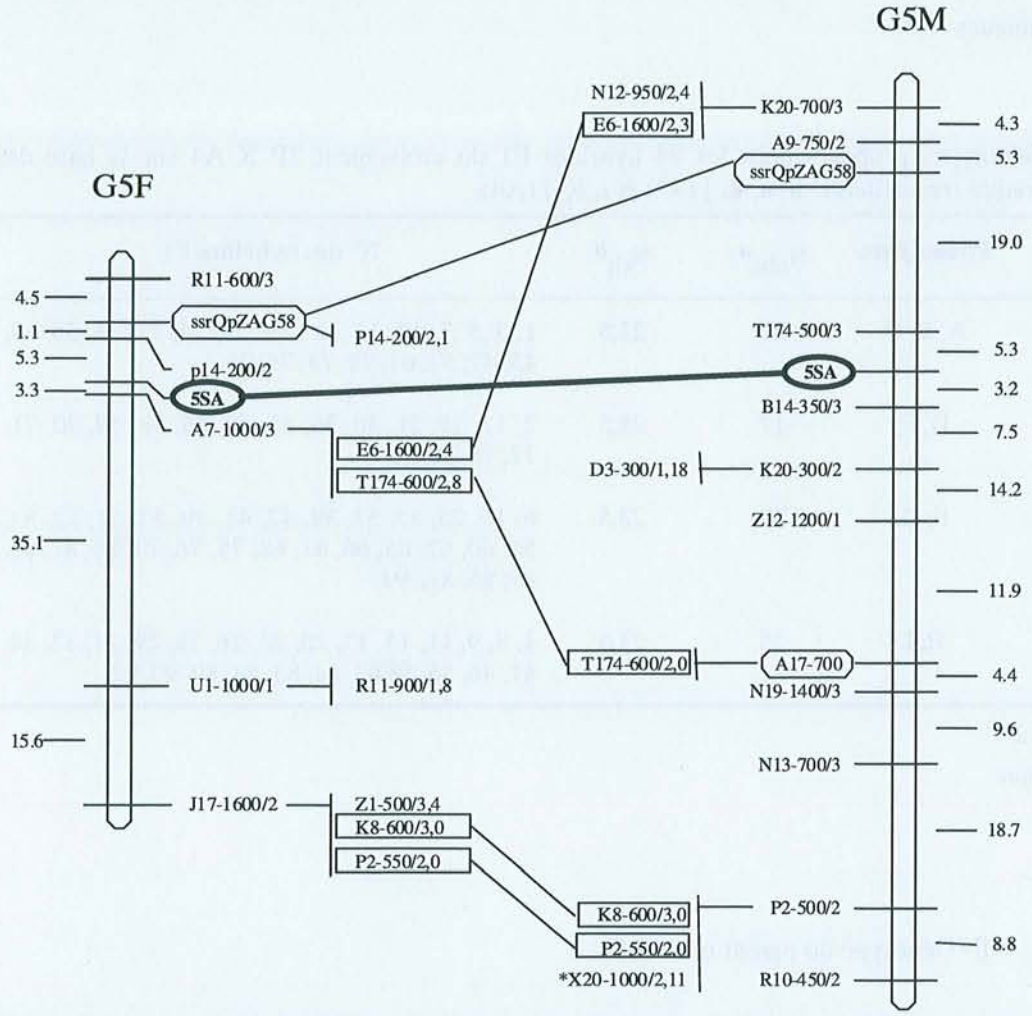



**Figure 34**. Localisation génétique du locus 5SA sur les groupes de liaison G5M (parent femelle 3P) et G5F (parent mâle A4) de chêne pédonculé. Les distances génétiques entre marqueurs sont indiquées pour chaque groupe (en cM). La nomenclature utilisée pour chaque marqueur est donnée en **Annexe** (Barrénèche *et al.* 1998). A l'exception de A17-700 (locus SCAR) et ssrQpZAG58 (locus microsatellite), les loci cartographiés sont des marqueurs RAPD. Les marqueurs communs aux deux groupes de liaison G5M et G5F sont reliés par des traits (marqueur d'ancrage).
* marqueur présentant une distorsion de ségrégation.

### δ- Cartographie génétique du locus 5SA

Grâce à sa configuration hétérozygote, il a été possible de localiser le locus 5SA sur les deux cartes génétiques construites pour chacun des deux génotypes parentaux 3P et A4 (Barrénèche *et al.* 1998, **Annexe**). L'analyse de liaison des marqueurs codominants de l'ADNr 5S avec d'autres marqueurs (RAPD, SCAR, microsatellites, minisatellites et isozymes) a été réalisée par l'équipe d'Antoine Kremer (INRA de Bordeaux). Le locus 5SA a été affecté sans ambiguïté aux groupes de

liaison G5F du parent femelle 3P (environ 64,9 cM) et G5M du parent mâle A4 (environ 112,2 cM) qui comportent respectivement 13 et 19 autres marqueurs, essentiellement RAPD (**Figure 34**).

d- Cas des marqueurs à hérédité non mendélienne **f** (3P) et *g* (A4)

La faible intensité du marqueur **f** rend sa détection difficile et irrégulière chez le parent 3P. Il est cependant clairement identifié chez les 14 descendants (**Figure 33**) des classes phénotypiques I (1 individu), C (2 individus), B et G (3 individus chacun) et enfin E (5 individus). Il est même le marqueur majeur chez les 3 arbres de phénotype B. Ce marqueur mineur **f** s'associe apparemment au hasard avec les allèles **hakj** et **hj** du locus 5SA lors de la gamétogénèse de l'arbre 3P et se retrouve donc dans les 4 classes génotypiques issues du croisement 3P X A4 précédemment décrites (**Tableau 27**). Toutefois, sa faible fréquence (14,9 %) ne permet pas d'accéder au déterminisme génétique de son héritabilité.

Le marqueur *g* est clairement identifié chez le parent mâle A4 (marqueur majeur) ainsi que dans la descendance avec une fréquence de 23,4 %. Comme pour le marqueur **f** précédent, cette fréquence peu élevée ne permet pas de proposer un modèle de ségrégation mendélien. On peut seulement constater que seuls les 22 individus de la classe génotypique (**hakj**)/(*bd*) (**Tableau 27**) sont porteurs de ce marqueur (phénotypes A, B et C, **Figure 33**). Il est donc préférentiellement (mais pas toujours) associé avec l'allèle (*bd*) du locus 5SA lors de la gamétogénèse du parent mâle A4. L'analyse des descendants révèle *a priori* la disjonction des groupes gamétiques (*bd*, *g*), (*bd*) et (*i*) dans un rapport 1:1:2 (22:30:42, $\chi^2$ (2 ddl) = 2,42 ; $p > 0,1$) au cours de la méiose.

**4- Discussion**

La plupart des marqueurs (7/9) de l'ADNr 5S (majeurs ou mineurs) détectés par amplification des espaceurs à l'aide d'amorces universelles chez les deux chênes pédonculés impliqués dans le croisement contrôlé étudié se sont avérés présenter un mode d'hérédité mendélien. C'est la situation qui a été décrite dans les quelques études comparables menées chez diverses plantes, aussi bien dans le cas de l'ADNr 5S (*Lycopersicon esculentum*, Lapitan *et al.* 1991, *Hordeum vulgare*, Kanazin *et al.* 1993, Laurie *et al.* 1993, *Beta vulgaris*, Schondelmaier *et al.* 1997) que dans le cas de l'ADNr (*Hordeum vulgare*, Saghai-Maroof *et al.* 1984, *Pisum sativum*, Ellis *et al.* 1984, Polans *et al.* 1986, *Triticum aestivum*, Snape *et al.* 1985, *Pinus sylvestris*, Karvonen et Savolainen 1993). L'analyse des phénotypes observés dans la descendance hybride F1 du croisement 3P X A4 a permis de montrer que l'héritabilité de ces marqueurs mendéliens est gouvernée par un locus principal (5SA) qui a pu être localisé sur le groupe de liaison G5 de la carte génétique récemment publiée (Barrénèche *et al.* 1998, **Annexe**). Les deux arbres parentaux sont hétérozygotes au locus 5SA qui permet donc, comme 4 autres loci RAPD (P2-550, K8-600, T174-600, E6-1600) et le locus microsatellite ssr*Qp*ZAG58 (**Figure 34**) d'ébaucher une carte consensus des groupes de liaison G5 mâle et femelle.

A notre connaissance, la cartographie génétique de l'ADNr 5S n'a été réalisée que chez la tomate (Lapitan *et al.* 1991), l'orge (Kanazin *et al.* 1993, Laurie *et al.* 1993) et la betterave sucrière (Schondelmaier *et al.* 1997). Dans chacun de ces trois cas, les auteurs ont pu accéder assez facilement à la localisation physique de l'ADNr 5S sur les chromosomes par la technique d'hybridation *in situ* de sondes fluorescentes (FISH : Fluorescent *In Situ* Hybridization). Comme les sites chromosomiques occupés par d'autres familles de séquences répétées en tandem dominantes du génome nucléaire (ADNr, divers satellites), les loci d'ADNr 5S peuvent donc être utilisés pour établir la correspondance entre groupes de liaison génétique et chromosomes. L'incorporation de tels jalons cytologiques dans les cartes de liaison est indispensable pour mieux comprendre les relations existant entre distance physique et distance génétique, et développer des stratégies de sélection assistée par marqueur (après identification de QTL) et/ou de clonage de gènes. Les cartes de liaison ne fournissent en effet que peu d'informations sur les distances physiques entre loci car la distribution des événements de recombinaison le long des chromosomes n'est pas aléatoire. La plupart interviennent dans les régions distales, ce qui se traduit, notamment, par des compressions de marqueurs autour du centromère (Tanksley *et al.* 1992, Laurie *et al.* 1993, Schondelmaier *et al.* 1997).

Comme d'autres espèces du genre *Quercus* (D'Emerico *et al.* 1995), le génome du chêne pédonculé comporte un nombre de base diploïde de 24 chromosomes (voir **Tableau 19**) qui sont difficiles à distinguer en raison de leur petite taille, même après révélation des bandes C par la technique de coloration au Giemsa (Ohri et ahuja 1991, Besendorfer *et al.* 1996). Seules 3 paires chromosomiques peuvent être identifiées formellement : les chromosomes N°1 métacentriques et de grande taille (6,24-6,26 µm) ; les chromosomes N°2, submétacentriques et porteurs d'une constriction secondaire ; et enfin, les chromosomes N°11 de type subtélocentrique (Ohri et Ahuja 1991). Il est donc difficile d'établir une correspondance entre les 12 groupes de liaison (Barrénèche *et al.* 1998, **Annexe**) et les 12 paires de chromosomes du chêne pédonculé, même si l'on peut penser que la paire N°1 correspond sans doute au groupe de liaison G2, le plus important aussi bien chez le parent femelle 3P (G2F : 124,8 Centimorgans) que chez le parent mâle A4 (G2M : 191,8 Centimorgans). La création de lignées aneuploïdes chez cette espèce diploïde est difficilement envisageable (Schondelmaier *et al.* 1997). Les stratégies utilisées pour associer groupes de liaison et chromosomes, et corréler cartes physique et génétique du chêne pédonculé sont donc largement dépendantes des approches cytogénétiques de haute résolution comme l'hybridation *in situ* employées récemment par Besendorfer *et al.* (1997) et Zoldos *et al.* (1999) pour les chênes. L'utilisation d'une sonde hétérologue de blé par ces auteurs a en effet permis de démontrer la présence des gènes codant pour les ARNr 18S, 5,8S et 26S au niveau de la constriction secondaire et des zones adjacentes subterminales de l'unique paire de chromosomes subtélocentriques N°11 (locus majeur) ainsi qu'au niveau de la constriction secondaire paracentromérique de la paire de chromosomes submétacentrique N°2 (locus mineur). Il serait donc intéressant d'identifier les groupes de liaison correspondant à ces deux loci NOR en étudiant l'héritabilité de marqueurs de l'ADNr, issus par exemple de l'amplification par PCR de régions variables comme les ITS ou certaines portions de l'IGS. Zoldos *et al.* (1999) ont par

ailleurs clairement détecté deux sites d'hybridation d'une sonde hétérologue d'ADNr 5S chez 11 espèces de chênes, incluant les chênes pédonculé et sessile, qui correspondent probablement au locus 5SA. Les signaux fluorescents ont été détectés au niveau de la région paracentromérique de la seconde paire de chromosomes métacentriques (par la taille) qui peut donc être associée avec le groupe de liaison génétique N°5. Des investigations complémentaires visant à préciser le nombre et la position des loci de l'ADNr et de l'ADNr 5S chez le chêne pédonculé ont été engagées dans notre équipe à l'aide de sondes homologues telle que pS*Qro*G102 décrite dans la **Partie II** de ce **Chapitre III**.

La composition des séries d'unités disposées sur chaque chromosome homologue au locus 5SA est généralement hétérogène. C'est le cas des séries alléliques composites (**hakj**) et (**hj**) de l'arbre 3P et (*bd*) de l'arbre A4. L'homogénéisation intrachromosomique des unités d'ADNr 5S (intramoléculaire ou entre chromatides soeurs) par les mécanismes de l'évolution concertée (Arnheim 1983) n'est donc pas suffisante ou n'a pas encore suffisamment joué pour éliminer totalement la variation. Seule la série allélique (*i*) de l'arbre A4 est apparue comme étant homogène. Toutefois, la forte diversité nucléotidique ($\pi = 0{,}025$) estimée à partir des séquences de 5 clones de variants électrophorétiques de la classe *d* définie dans la **Partie II** de ce **Chapitre III**, à laquelle le fragment majeur *i* appartient (clone S*Qro*A4*d*), suggère que les conditions d'électrophorèse utilisées, pourtant résolutives (parfois jusqu'à une mutation ponctuelle détectée), n'ont peut-être pas permis de révéler l'hétérogénéité existant dans cette série allélique (*i*). Le fait que l'on puisse observer des marqueurs spécifiques de chaque série (**a**, **k**, *b*, *d* ou *i*), indique en outre que la fréquence des échanges intrachromosomiques (entre chromatides soeurs) est supérieure à la fréquence des échanges interchromosomiques (entre chromosomes homologues) au locus 5SA de *Quercus robur*. Des conclusions similaires ont d'ailleurs été rapportées chez divers organismes dans le cas des séries de l'ADNr (Seperak *et al.* 1988, Karvonen et Savolainen 1993, Schlötterer et Tautz 1994, Copenhaver et Pikaard 1996) et d'autres familles de séquences répétées en tandem (Reed *et al.* 1994, Liao *et al.* 1997, Liao 1999). Ceci explique probablement la configuration hétérozygote généralement observée pour les loci occupés par des séquences répétées en tandem (déséquilibre de liaison intraindividuel, Dover 1989). Le modèle d'évolution concertée intralocus des séries multigéniques proposé par Liao *et al.* (1997) impliquant de fréquentes homogénéisations intrachromosomiques et de rares interactions entre chromosomes homologues (consulter la **Figure 10** et la **Partie II-3-c-$\chi$** du **Chapitre I**) paraît donc bien être applicable aux séries de l'ADNr 5S du chêne pédonculé. Dans cette hypothèse, on peut considérer que les marqueurs spécifiques de chaque série (**a**, **k**, *b*, *d* ou *i*) résultent sans doute de mutations relativement récentes qui sont, soit en cours d'élimination par les mécanismes d'homogénéisation intrachromosomique (c'est le cas le plus fréquent en l'absence de conversion génique biaisée, Karvonen et Savolainen 1993), soit en phase de transfert horizontal (ou *cis*) et de diffusion dans les populations ("molecular drive", Dover 1982) où elles sont soumises à la sélection naturelle (Flavell *et al.* 1986). Les marqueurs communs aux deux séries du locus 5SA (**h** et **j** de l'arbre 3P) seraient quant à

eux plus anciens et auraient subi une diffusion *trans* au cours des rares échanges intervenant entre chromosomes homologues.

A côté de ces marqueurs à hérédité mendélienne (**h**, **a**, **k**, **j** et *b*, *d*, *i*), deux fragments majeur (*g*) et mineur (**f**) présentent de sévères distorsions de ségrégation dans la descendance (fréquence faible). Plusieurs causes peuvent être avancées pour expliquer ces distorsions. Elles pourraient d'abord être dues à des difficultés expérimentales résultant de la structure particulière des séries d'unités d'ADNr 5S. Par exemple, l'accessibilité des séries de l'ADNr 5S aux amorces universelles F et B lors des cycles d'amplification par PCR pourrait être localement rendue difficile. Buckler *et al.* (1997) ont ainsi montré que la structure secondaire adoptée par certains ITS variants de l'ADNr de maïs est suffisamment stable pour perturber, voire interdire, leur amplification. Une situation similaire pourrait être envisagée pour les unités d'ADNr 5S. Les gènes 5S peuvent en effet théoriquement adopter une structure secondaire proche de celle des ARNr 5S, ce qui empêcherait la fixation des amorces F et B. Si les différents variants d'une série sont réellement organisés en plusieurs sous-séries homogènes adjacentes, comme semblent l'indiquer les résultats de Copenhaver et Pikaard (1996) concernant l'ADNr d'*Arabidopsis thaliana* (au lieu d'être mélangés), il est concevable que certains d'entre eux (et certains seulement) soient irrégulièrement amplifiés s'ils se trouvent dans une zone de la série où les gènes sont peu accessibles. Cela pourrait être le cas du marqueur **f**, de faible intensité chez le parent femelle 3P, d'intensité moyenne chez les arbres de phénotypes C, E, G, I, et d'intensité très forte chez les 3 arbres de phénotype B (**Figure 33**). Si **f** provient réellement du locus 5SA cartographié, il est probablement en configuration hétérozygote car il est présent dans les 4 classes génotypiques (**Tableau 27**). Pour vérifier cette hypothèse d'une influence de la structure secondaire des gènes 5S sur l'amplification des espaceurs, certains auteurs suggèrent d'introduire dans le milieu réactionnel des agents dénaturants comme le diméthylsulfoxyde (DMSO, Varadaraj et Skinner, 1994) et/ou des nucléotides qui facilitent (après incorporation par l'ADN polymérase) la dénaturation des produits de PCR, comme la 7-deaza-2'-déoxyguanosine triphosphate ($c^7$dGTP, Innis 1990). Une telle explication n'est cependant pas envisageable dans le cas du fragment majeur *g* qui est régulièrement et efficacement amplifié chez les seuls individus de la classe génotypique (**hakj**)/(*bd*). L'absence d'amplification de ce marqueur chez les individus de la classe génotypique (**hj**)/(*bd*) est troublante si on considère qu'il appartient au moins à l'une des deux séries du locus 5SA (configuration hétérozygote).

Les distorsions de ségrégation des marqueurs **f** et *g* pourraient aussi être expliquées par des phénomènes biologiques fondés soit sur des caractéristiques structurales particulières du génome nucléaire des chênes pédonculés 3P et A4, soit sur les mécanismes de l'évolution concertée. Plusieurs scénarios peuvent être envisagés dans le premier cas. Il paraît peu probable que certaines combinaisons alléliques au locus 5SA puissent induire des effets létaux dans la descendance en raison du grand nombre de copies de ces gènes à répétition de dosage (1000-2000 unités/2C, voir la **Partie II** de ce **Chapitre III**). En revanche, le locus 5SA pourrait se trouver physiquement proche de loci

impliqués dans les phénomènes de dépression de consanguinité (liée à l'homozygotie) auxquels le chêne pédonculé paraît assez sensible. De jeunes semis issus d'autofécondations contrôlées présentent en effet de nombreux caractères défavorables comme une capacité de germination faible, une déficience chlorophyllienne prononcée et une croissance réduite (Kleinschmit et Kleinschmit 1996, Barrénèche *et al.* 1998, **Annexe**). La détection de nombreux marqueurs qui ségrègent anormalement dans la descendance du croisement 3P X A4 (18 % des marqueurs, Barrénèche *et al.* 1998, **Annexe**) est compatible avec une telle hypothèse. Cependant, la majorité de ces marqueurs "distordus" (65 %) sont rassemblés dans les 3 groupes de liaison G2, G4 et G9 alors qu'un seul se trouve dans le groupe de liaison G5. Dans l'état actuel de nos connaissances, il semble donc difficile d'admettre que le locus 5SA soit situé dans une zone du génome inductrice de distorsions de ségrégation par dépression de consanguinité. Si tel était le cas, l'héritabilité de l'ensemble des marqueurs du locus 5SA (et pas seulement les marqueurs **f** et *g*) devrait en être affectée. Une conclusion plus définitive ne pourra réellement être apportée que lorsque la carte génétique, encore incomplète, aura été saturée à l'aide d'autres marqueurs.

L'existence éventuelle d'un second locus d'ADNr 5S (virtuellement nommé 5SB) sur un ou plusieurs chromosomes surnuméraires (chromosomes B) dans le génome des chênes pédonculé 3P et A4 peut constituer une autre piste intéressante à envisager pour tenter d'expliquer les distorsions de ségrégation des marqueurs **f** et *g*. Ces chromosomes surnuméraires (c'est-à-dire non indispensables), nettement plus petits et qui ne se recombinent pas avec les autres chromosomes du complément diploïde (chromosomes A), ont en effet un mode d'hérédité non mendélien très irrégulier (revue détaillée dans Jones 1995). Bien que généralement dépourvus de gènes majeurs, ils portent souvent des familles de séquences répétées en tandem (Beukeboom 1994) et plusieurs exemples de chromosomes B véhiculant des unités d'ADNr 18S-5,8S-26S sont maintenant connus chez les plantes (une vingtaine de cas, Jones 1995). La présence d'ADNr 5S sur de telles structures est donc envisageable.

Chez le chêne pédonculé, Ohri et Ahuja (1990) ont effectivement détecté un chromosome B unique de faible taille (moins de 2 µm) et très hétérochromatique (probablement riche en séquences répétées) en plus des 12 paires de chromosomes A dans le génome de plusieurs individus. De même, Besendorfer *et al.* (1996) ont observé un ou deux chromosomes B de taille différente et riches en GC dans certaines cellules de plusieurs arbres d'une population du nord de la Croatie (voir **Tableau 19**). Sur la base de différentes techniques de coloration, ces derniers auteurs indiquent quee ces chromosomes B correspondent probablement à des loci NOR (organisateurs nucléolaires probablement riches en ADNr 18S-5,8S-26S) détachés des chromosomes A (voir aussi Besendorfer *et al.* 1997). Cependant, aucune preuve moléculaire directe de la présence d'ADNr sur ces corpuscules n'a encore été apportée. Zoldos *et al.* (1998) n'ont en revanche pas pu mettre en évidence de chromosomes surnuméraires au sein de 4 populations françaises et croates de chênes pédonculés. Jusqu'à présent, les observations que nous avons réalisées sur des apex racinaires de différents

génotypes maintenus en conditions de culture hydroponique sont aussi restées vaines. Les cartes génétiques des individus 3P et A4 (Barrénèche *et al.* 1998, **Annexe**) ne révèlent pas non plus de groupes de liaison excédentaires (il n'y en a que 12 dans les deux cas). Cela n'exclut toutefois pas la présence de chromosomes B chez ces deux arbres. Leur taille réduite par rapport aux chromosomes A (Ohri et Ahuja 1990, Besendorfer *et al.* 1996) constitue probablement un obstacle important à la détection de marqueurs spécifiques. Dans le cas des petits groupes de liaison G12M et G12F par exemple, seuls deux marqueurs (sur un total de 307) ont pu être cartographiés. Comme la présence de chromosomes B chez le chêne pédonculé est apparemment variable selon les populations, les individus ou parfois les cellules d'un même individu (c'est une situation courante chez les plantes, Jones 1995), il faudrait pouvoir établir un bilan caryologique précis des arbres 3P et A4.

Si l'on retient néanmoins cette hypothèse de l'existence de chromosomes B chez le chêne pédonculé, il est intéressant de remarquer que dans le cas du parent mâle A4, la production apparente des 3 pools gamétiques (*bd, g*), (*bd*) et (*i*) dans un rapport 1:1:2 au cours de la méiose est compatible avec un modèle génétique impliquant le locus principal 5SA [porteur des séries alléliques (*bd*) et (*i*)] et un deuxième locus 5SB situé sur un unique chromosome B (**Figure 35A**). Deux conditions doivent cependant être remplies. La première est que le chromosome B ségrège préférentiellement avec le chromosome A porteur de la série allélique (*bd*) pendant la méiose I ; la seconde, qu'il n'y ait pas disjonction des chromatides soeurs du chromosome B à la méiose II ou, de manière équivalente, que le chromosome B se réplique de façon asynchrone par rapport aux chromosomes A. Ces deux conditions semblent plausibles à la lecture des articles de synthèse de Beukeboom (1994) et de Jones (1995). On peut ajouter que Ohri et Ahuja (1990) ont effectivement observé des chromosomes B se divisant tardivement au cours de la méiose II (à la fin de l'anaphase) chez le chêne pédonculé. La faible différence de séquence nucléotidique (seulement 11 substitutions) enregistrée entre les variants majeur *i* (locus 5SA) et *g* (locus 5SB éventuel) constitue toutefois un argument en défaveur de ce scénario. En effet, les unités d'ADNr 5S situées à des loci différents sont normalement très divergentes en raison de l'absence ou de la très faible fréquence des mécanismes d'homogénéisation entre chromosomes non homologues (voir la **Partie III-4-a** du **Chapitre II**).

La ségrégation 1:1:2 des groupes de marqueurs (*bd, g*), (*bd*) et (*i*) au cours de la méiose de l'arbre A4 peut aussi être expliquée par l'intervention des mécanismes de l'évolution concertée, sans qu'il soit nécessaire de faire appel à la présence d'un chromosome B (**Figure 35B**). Admettons par exemple que le marqueur *g* soit à l'origine un constituant de la série allélique (*bd*) du locus 5SA [série allélique que nous noterons (*bdg*)] de l'arbre A4. Cela est plausible, puisque le marqueur *g* ségrège uniquement avec les marqueur *b* et *d* (**Figure 33**).

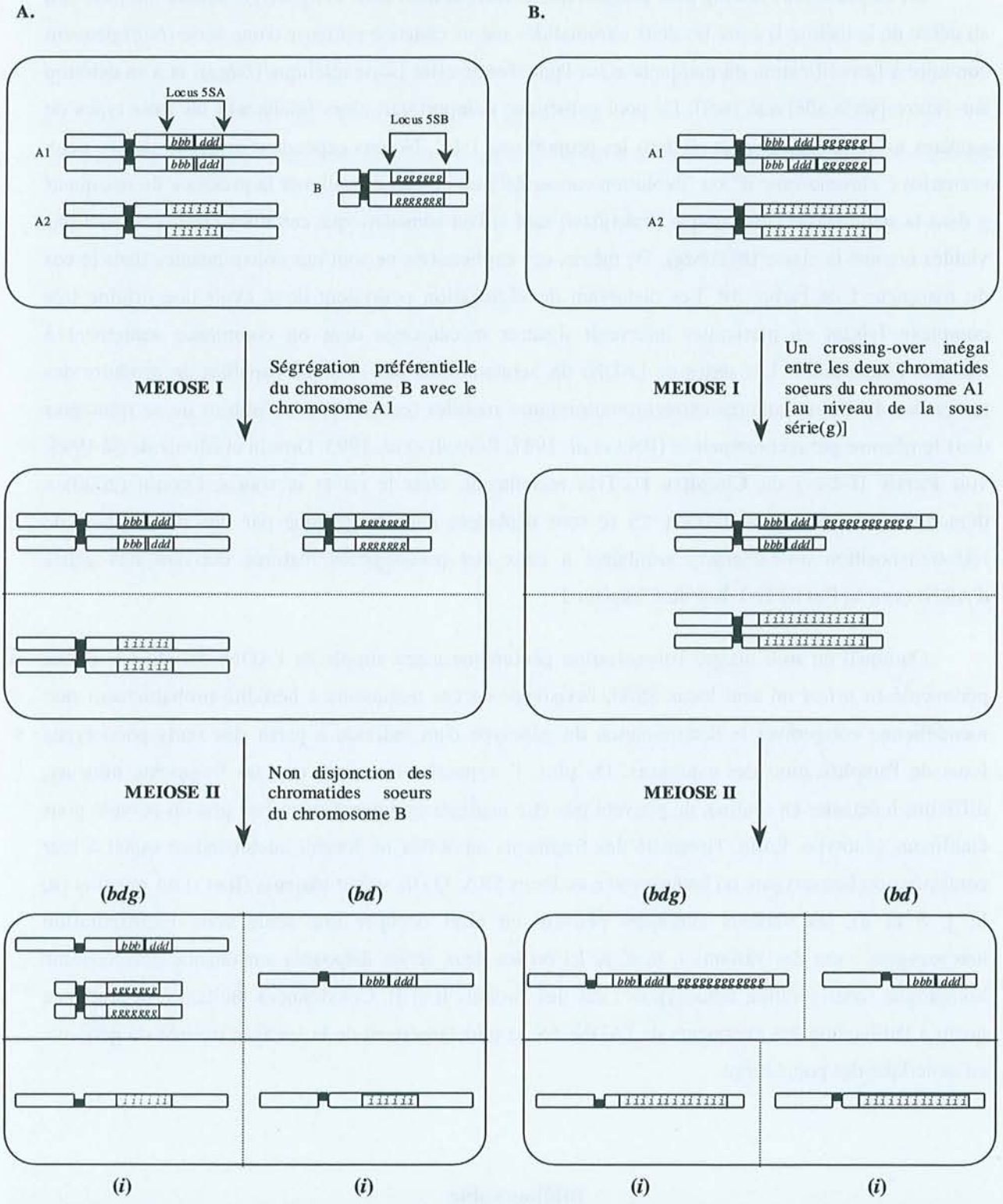



**Figure 35**. Deux modèles génétiques hypothétiques permettant d'expliquer la ségrégation des 3 groupes de marqueurs de l'ADNr 5S (*bdg*), (*bd*) et (*i*) dans un rapport 1:1:2 lors de la gamétogénèse du chêne pédonculé A4.

**A.** Modèle basé sur la présence d'un chromosome B (porteur du marqueur *g*) qui ségrège préférentiellement avec le chromosome A1 (porteur des marqueurs *b* et *d*) au cours de la méiose I et dont les chromatides soeurs ne se disjoignent pas au cours de la méiose II. Un modèle équivalent fait intervenir un chromosome B qui ne se réplique pas ou de manière asynchrone par rapport aux chromosomes A.

**B.** Modèle impliquant un crossing-over inégal entre les deux chromatides soeurs du chromosome A1 (porteur ici des 3 marqueurs *b*, *d* et *g*) au cours de la méiose I qui produit une amplification de la sous-série (*g*) sur l'une d'entre elles et une délétion sur l'autre.

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

Un ou plusieurs crossing over inégaux qui se réaliseraient dans la lignée germinale (au plus tard au début de la méiose I) entre les deux chromatides soeurs chacune porteuse d'une série (*bdg*) peuvent conduire à l'amplification du marqueur *g* sur l'une d'entre elles [série allélique (*bdgg*)] et à sa délétion sur l'autre [série allélique (*bd*)]. Le pool gamétique comporterait alors finalement les trois types de gamètes notés (*bdgg*), (*bd*) et (*i*) dans les proportions 1:1:2. Notons cependant qu'aucun de ces deux scénarios ("chromosome B" ou "évolution concertée") ne permet d'expliquer la présence du marqueur *g* dans la seule classe génotypique (**hakj**)/(*bd*) sauf si l'on considère que certaines classes ne sont pas viables comme la classe (**hj**)/(*bdg*). De même, ces explications ne sont pas convaincantes dans le cas du marqueur **f** de l'arbre 3P. Les distorsion de ségrégation pourraient donc avoir une origine très complexe faisant en particulier intervenir d'autres mécanismes dont on commence seulement à mesurer l'importance. Les séries de l'ADNr 5S seraient ainsi par exemple capables de produire des molécules d'ADN circulaires extrachromosomiques mobiles (cccDNA) susceptibles de se réintégrer dans le génome par recombinaison (Pont *et al.* 1987, Renault *et al.* 1993, Drouin et Moniz de Sá 1995, voir **Partie II-2-c-χ** du **Chapitre I**). Très récemment, chez le rat et la souris, Drouin (2000) a démontré que des copies d'ADNr 5S se sont déplacées dans le génome par des mécanismes de rétrotransposition probablement similaires à ceux des pseudogènes maturés dérivant des gènes d'ARNt (voir la **Partie II-1-b-β** du **Chapitre I**).

Quoiqu'il en soit, malgré l'organisation génomique assez simple de l'ADNr 5S chez le chêne pédonculé (*a priori* un seul locus 5SA), l'existence de ces marqueurs à hérédité probablement non mendélienne compromet la détermination du génotype d'un individu à partir des seuls phénotypes issus de l'amplification des espaceurs. De plus, il apparaît clairement que les fragments mineurs, difficiles à détecter en routine, ne peuvent pas être négligés et doivent donc être pris en compte pour établir un génotype. Enfin, l'intensité des fragments amplifiés ne fournit aucun indice quant à leur configuration homozygote ou hétérozygote au locus 5SA. Qu'ils soient majeurs (**h** et *i*) ou mineurs (**a**, **k**, **j**, *b* et *d*), les variants alléliques peuvent en effet occuper une seule série (configuration hétérozygote : cas des variants *i*, *b*, *d*, **a**, **k**) ou les deux séries disposées sur chaque chromosome homologue (configuration homozygote : cas des variants **h** et **j**). Ces données incitent à la prudence quant à l'utilisation des marqueurs de l'ADNr 5S, et plus largement de la fraction répétée du génome, en génétique des populations.

## Bibliographie

**Aas G (1993).** Taxonomical impact of morphological variation in *Quercus robur* and *Quercus petraea*: a contribution to the hybrid controversy. Ann Sci For 50, suppl. 1 : 107s-113s

**Amarasinghe V, Carlson JE (1998).** Physical mapping and characterization of 5S rRNA genes in douglas-fir. J Hered 89 : 495-500

**Arnheim N (1983).** Concerted evolution of multigene families. Dans *Evolution of genes and proteins*, pp 38-61. Edité par Nei M, Koehn RK. Sinauer associates, Sunderland MA

**Arnold ML (1997).** *Natural hybridization and evolution.* Oxford University Press, New York, Oxford

**Avise JC (1994).** *Molecular markers, natural history and evolution.* Chapman & Hall, New York

**Bacilieri R, Ducousso A, Kremer A (1995).** Genetic, morphological, ecological and phenological differentiation between *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. in a mixed stand of northwest of France. Sylvae Genet 44 : 1-10

**Bacilieri R, Ducousso A, Petit RJ, Kremer A (1996).** Mating system and directional gene flow in a mixed oak stand. Evolution 50 : 900-908

**Barciszewska MZ, Erdmann VA, Barciszewski J (1994).** A new model for the tertiary structure of 5S ribonucleic acid in plants. Plant Mol Biol Rep 12 : 116-131

**Barrénèche T, Bodénès C, Lexer C, Trontin JF, Fluch S, Streiff R, Plomion C, Roussel G, Steinkellner H, Burg K, Favre JM, Glössl J, Kremer A (1998).** A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers. Theor Appl Genet 97 : 1090-1103

**Bassam BJ, Caetano-Anollés G, Gresshoff PM (1991).** Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal Biochem 196 : 80-83

**Baum BR, Appels R (1992).**Evolutionary change at the 5S DNA loci of species in the *Triticeae*. Plant Syst Evol 183 : 195-208

**Baum BR, Johnson DA (1998).**The 5S rRNA gene in sea barley (*Hordeum marinum* Hudson *sensu lato*): sequence variation among repeat units and relationship to the X haplome in barley (*Hordeum*). Genome 41 : 652-661

**Becker M, Lévy G (1982).** Le dépérissement du chêne en forêt de Tronçais. Les causes écologiques. Ann Sci For 39 : 439-444

**Besendorfer V, Zoldos V, Siljak-Yakovlev S, Javornik B, Papes D (1997).** Interspecific and interindividual variability of NORs and its restriction pattern in *Quercus* species. Dans *Diversity and adaptation in oak species*, pp 90-97. Edité par Steiner KC. Proceeding of the second meeting of working party 2.08.05, genetics of *Quercus*, of the IUFRO

**Besendorfer V, Zoldos V, Peskan T, Krsnik-Rasol M, Littvay T, Papes D (1996).** Identification of potential cytogenetical and biochemical markers in bioindication of common oak forests. Phyton 36 : 139-146

**Beukeboom LW (1994).** Bewildering Bs: an impression of the 1st B-chromosome conference. Heredity 73 : 328-336

**Bodénès C (1996).** Différenciation moléculaire entre chêne sessile (*Quercus petraea* (Matt.) Liebl.) et chêne pédonculé (*Quercus robur* L.) Thèse de l'Université de Bordeaux I

**Bodénès C, Laigret F, Joandet S, Kremer A (1997a).** Detection of genomic regions differentiating two closely related oak species *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. Heredity 78 : 433-444

**Bodénès C, Labbé T, Pradère S, Kremer A (1997b).** General vs. local differentiation between two closely related white oak species. Mol Ecol 6 : 713-724

**Brown GR, Carlson JE (1997).** Molecular cytogenetics of the genes encoding 18S-5.8S-26S rRNA and 5S rRNA in two species of spruce (*Picea*). Theor Appl Genet 95 : 1-9

**Buckler ES, Ippolito A, Holtsford TP (1997).** The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. Genetics 145 : 821-832

**Bussotti F, Grossoni P (1998).** Des problèmes dans la classification des chênes. Taxonomie en Europe et région méditérranéenne. Forêt Méditérranéenne 19 : 267-279

**Camus A (1936-1954).** Les chênes - Monographie du genre *Quercus*. Le Chevalier, Paris

**Clegg MY (1989).** Molecular diversity in plant populations. Dans *Plant population genetics, breeding and genetic resources*, pp 98-115. Edité par Brown AHD, Clegg MT, Kahler AL, Weir BS. Sinauer Associates, Sunderland, Masschusetts

**Copenhaver GP, Pikaard CS (1996).** Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in *Arabidopsis thaliana* suggesting local spreading of variants as the mode for gene homogenization during concerted evolution. Plant J 9 : 273-282

**Cronn RC, Zhao X, Paterson AH, Wendel JF (1996).** Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. J Mol Evol 42 : 685-705

**D'Emerico S, Bianco P, Medagli P, Schirone B (1995).** Karyotype analysis in *Quercus* spp. (*Fagaceae*). Silvae genet 44 : 66-70

**Dover G (1982).** Molecular drive: a cohesive mode of species evolution. Nature 299 : 111-117

**Dover GA (1989).** Linkage disequilibrium and molecular drive in the rDNA gene family. Genetics 122 : 249-252

**Doyle JJ, Doyle JL (1987).** A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19 : 11-15

**Drouin G (2000).** Expressed retrotransposed 5S rRNA genes in the mouse and rat genome. Genome 43 : 213-215

**Drouin G, Moniz de Sá M (1995).** The concerted evolution of 5S ribosomal genes linked to the repeat units of other multigene families. Mol Biol Evol : 481-493

**Dumolin-Lapègue S, Demesure B, Fineschi S, Le Corre V, Petit RJ (1997).** Phylogeographic structure of white oaks throughout the European continent. Genetics 146 : 1475-1487

**Dumolin-Lapègue S, Pemonge MH, Petit RJ (1998).** Association between chloroplast and mitochondrial lineages in oaks. Mol Biol Evol 15 : 1321-1331

**Dupouey JL (1983).** Analyse multivariable de quelques caractères morphologiques de populations de chênes (*Quercus robur* L. et *Quercus petraea* (Matt.) Liebl.) du Hurepoix. Ann Sci For 40 : 265-282

**Dupouey JL, Badeau V (1993).** Morphological variability of oaks (*Quercus robur* L., *Quercus petraea* (Matt.) Liebl., *Quercus pubescens* Willd) in northeastern France: preliminary results. Ann Sci Forest 50, Suppl 1 : 35s-40s

**Dvorák J, Zhang HB, Kota RS, Lassner M (1989).** Organization and evolution of the 5S ribosomal RNA gene family in wheat and related species. Genome 32 : 1003-1016

**Ellis THN, Davies DR, Castleton JA, Bedford ID (1984).** The organization and genetics of rDNA length variants in peas. Chromosoma 91 : 74-81

**Favre JM, Brown S (1996).** A flow cytometric evaluation of the nuclear DNA content and GC percent in genomes of European oak species. Ann Sci Forest 53 : 915-917

**Ferris C, Oliver RP, Davy AJ, Hewitt GM (1993).** Native oak chloroplasts reveal an ancient divide across Europe. Mol Ecol 2 : 337-344.

**Ferris C, King RA, Väinölät R, Hewitt GM (1998).** Chloroplast DNA recognizes three refugial sources of European oaks and suggests independent eastern and western immigrations to Finland. Heredity 80 : 584-593

**Feuillat F, Dupouey JL, Sciama D, Keller R (1997).** A new attempt at discrimination between *Quercus petraea* and *Quercus robur* based on wood anatomy. Can J Forest Res 27 : 343-351

**Flavell RB, Jones J, Lonsdale DM, O'Dell M (1983).** Higher plant genome structure and the dynamics of genome evolution. Dans *Advances in gene technology. Molecular genetics of plants and animals*, pp 47–59. Edité par Downey K, Voellmy RN, Ahmed F, Schultz J. Academic Press, New York

**Flavell RB, O'Dell M, Sharp P, Nevo E, Beiles A (1986).** Variation in intergenic spacer of ribosomal DNA of wild wheat, *Triticum dicoccoides* in Israel. Mol Biol Evol 3 : 547-558

**Gorman SW, Teasdale RD, Cullis CA (1992).** Structure and organization of the 5S rRNA genes (5S DNA) in *Pinus radiata* (*Pinaceae*). Plant Syst Evol 183 : 223-234

**Grandjean G, Sigaud P (1987).** Contribution à la taxonomie et à l'écologie des chênes du Berry. Ann Sci Forest 44 : 35-66

**Hamrick JL, Godt MJW, Sherman-Broyles SL (1992).** Factors influencing levels of genetic diversity in woody plant species. New Forests 6 : 95-124

**Hauser MT, Adhami F, Dorner M, Fuchs E, Glössl J (1998).** Generation of co-dominant PCR-based markers by duplex analysis on high resolution gels. Plant J 16 : 117-125

**Huntley B, Birks HJ (1983).** An atlas of past and present pollen maps for Europe: 0–13000 years ago. Cambridge University Press

**Innis MA (1990).** PCR with 7-deaza-2'-deoxyguanosine triphosphate. Dans *PCR protocols: a guide to methods and applications*, pp 54-59. Edité par Innis MA, Gelfand DH, Sninsky JJ, White TJ. Academic Press, Inc., San Diego

**Johnson DA, Chan CCY, Gottlob-McHugh SG, Mackenzie K, Marengère L, Prud'homme MC (1992).** Structure of the 5S rRNA genes in birch (*Betula papyrifera*) and alder (*Alnus incana*). Genome 35 : 337-341

**Jones RN (1995).** B chromosomes in plants. New phytol 131 : 411-434

**Kanazin V, Ananiev E, Blake T (1993).** The genetics of 5S rRNA encoding multigene families in barley. Genome 36 : 1023-1028

**Karvonen P, Savolainen O (1993).** Variation and inheritance of ribosomal DNA in *Pinus sylvestris* L. (Scots pine). Heredity 71 : 614-622

**Kellogg EA, Appels R (1995).** Intraspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. Genetics 140 : 325-343

**Kleinschmit J (1993).** Intraspecific variation of growth and adaptative traits in European oak species. Ann Sci Forest 50, Suppl 1 : 166s-185s

**Kleinschmit JRG, Kleinchmit J (1996).** Artificial hybridization between *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. Dans *Inter- and intra-specific variation in European oaks: evolutionary implications and practical consequences*, pp 69-87. Edité par Kremer A, Mühs H. European Commission, EUR 16717EN, Luxembourg

**Kleinschmit JRG, Bacilieri R, Kremer A, Roloff A (1995).** Comparison of morphological and genetic traits of pedunculate oak (*Q. robur* L.) and sessile oak (*Q. petraea* (Matt.) Liebl.). Silvae genet 44 : 5-6

**Ko HL, Henry RJ (1996).** Specific 5S ribosomal RNA primers for plant species identification in admixtures. Plant Mol Biol Rep 14 : 33-43

**Kolchinsky A, Kolesnikova M, Ananiev E (1991).** "Portraying" of plant genomes using polymerase chain reaction amplification of ribosomal 5S genes. Genome 34 : 1028-1031

**Korn LJ (1982).** Transcription of *Xenopus* 5S ribosomal RNA genes. Nature 295 : 101-105

**Kremer A, Petit RJ, Zanetto A, Fougère V, Ducousso A, Wagner DB, Chauvin C (1991).** Nuclear and organelle gene diversity in *Q. robur* and *Q. petraea*. Dans *Genetic variation in European forest trees*, pp 141-166. Edité par Müller-Starck G, Ziehe M. Sauerländer's Verlag, Frankfurt-am-Main

**Kremer A, Petit RJ (1993).** Gene diversity in natural populations of oak species. Ann Sci Forest 50, Suppl 1 : 186s-202s

**Krüssmann G (1986).** *Manual of cultivated broad-leaved trees and shrubs*. Vol 2. BT Batsford LTD, London

**Lapitan NLV (1992).** Organization and evolution of higher plant nuclear genomes. Genome 35 : 171-181

**Lapitan NLV, Ganal MW, Tanksley SD (1991).** Organization of the 5S ribosomal RNA genes in the genome of tomato. Genome 34 : 509-514

**Laurie DA, Pratchett N, Devos KM, Leitch IJ, Gale MD (1993).** The distribution of RFLP markers on chromosome 2(2H) of barley in relation to the physical and genetic location of 5S rDNA. Theor Appl Genet 87 : 177-183

**Lewontin RC (1972).** The apportionment of human diversity. Evol Biol 6 : 381-398

**Liao D (1999).** Concerted evolution : molecular mechanism and biological implications. Am J Hum Genet 64 : 24-30

**Liao D, Pavelitz T, Kidd JR, Kidd KK, Weiner AM (1997).** Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the *RNU2* locus) involves rapid intrachromosomal homogenization and rare interchromosomal gene conversion. EMBO J 16 : 588-598

**Manos PS, Doyle JJ, Nixon KC (1999).** Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (*Fagaceae*). Molecular Phylogenetics and Evolution 12 : 333-349

**McIntyre CL, Winberg B, Houchins K, Appels R, Baum BR (1992).** Relationships between *Oryza* species (*Poaceae*) based on 5S DNA sequences. Plant Syst Evol 183 : 249-264

**Moran GF, Smith D, Bell JC, Appels R (1992).** The 5S RNA genes in *Pinus radiata* and the spacer region as a probe for relationships between *Pinus* species. Plant Syst Evol 183 : 209-221

**Moreau F, Kleinschmit J, Kremer A (1994).** Molecular differentiation between *Q. petraea* and *Q. robur* assessed by random amplified DNA fragments. Forest Genet 1 : 51-64

**Müller-Starck G, Herzog S, Hattemer HH (1993).** Intra- and interpopulational genetic variation in juvenile populations of *Quercus robur* L. and *Quercus petraea* Liebl. Ann Sci Forest 50, Suppl 1 : 233s-244s

**Nei M (1987).** *Molecular evolutionary genetics.* Columbia University Press, New York

**Nixon KC (1993).** Infrageneric classification of *Quercus* (*Fagaceae*) and typification of sectional names. Ann Sci Forest 50, Suppl 1 : 25s-34s

**Ohri D, Ahuja MR (1990).** Giemsa C-banded karyotype in *Quercus* L. (Oak). Silvae Genet 39 : 216-219

**Ohta T, Dover GA (1983).** Population genetics of multi-gene families that are dispersed into two or more chromosomes. Proc Natl Acad Sci USA 80 : 4079-4083

**Petit RJ, Kremer A, Wagner DB (1993a).** Geographic structure of chloroplast DNA polymorphism in European oaks. Theor Appl Genet 87 : 122-128.

**Petit RJ, Wagner DB, Kremer A (1993b).** Ribosomal DNA and chloroplast DNA polymorphisms in a mixed stand of *Quercus robur* and *Quercus petraea*. Ann Sci Forest 50, Suppl 1 : 41s-47s

**Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducousso A, Kremer A (1997).** Chloroplaste DNA footprints of postglacial recolonization by oaks. Proc Natl Acad Sci USA 94 : 9996-10001

**Playford J, Appels R, Baum BR (1992).** The 5S DNA units of *Acacia* species (*Mimosaceae*). Plant Syst Evol 183 : 235-247

**Polans NO, Weeden NF, Thompson WF (1986).** Distribution, inheritance and linkage relationship of ribosomal DNA spacer length variants in pea. Theor Appl Genet 72 : 289-295

**Pont G, DeGroote F, Picard G (1987).** Some extrachromosomal circular DNAs from *Drosophila* embryos are homologous to tandemly repeated genes. J Mol Biol 195 : 447-451

**Rameau (1990).** Comportement dynamique du chêne pédonculé et du chêne sessile dans les successions forestières. Rev For Fr 42 : 155-164

**Reed KM, Beukeboom LW, Eickbush DG, Werren JH (1994).** Junctions between repetitive DNAs on the PSR chromosome of *Nasonia vitripennis*: association of palindromes with recombination. Mol Evol 38 : 352-362

**Renault S, DeGroote F, Picard G (1993).** Identification of short tandemly repeated sequences in extrachromosomal circular DNAs from *Drosophila melanogaster* embryos. Genome 36 : 244-254

**Röder MS, Sorrells ME, Tanksley SD (1992).** 5S ribosomal gene clusters in wheat: pulsed field gel electrophoresis reveals a high degree of polymorphism. Mol Gen Genet 232 : 215-220

**Rushton BS (1977).** Artificial hybridization between *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. Watsonia 11 : 229-236

**Rushton BS (1978).** *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl.: a multivariate approach to the hybrid problem. 1. Data acquisition, analysis and interpretation. Watsonia 12 : 81-101

**Rushton BS (1993).** Natural hybridization within the genus *Quercus* L. Ann Sci Forest 50, Suppl 1 : 73s-90s

**Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984).** Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81 : 8014-8018

**Sambrook J, Fritsch EF, Maniatis T (1989).** Molecular cloning: a laboratory manual. Second edition. Cold Spring Harbor Laboratory Press

**Samuel (1999).** Identification of hybrids between *Quercus petraea* and *Q. robur* (*Fagaceae*): results obtained with RAPD markers confirm allozyme studies based on the *Got*-2 locus. Plant Syst Evol 217 : 137-146

**Samuel R, Bachmair A, Jobst J, Ehrendorfer F (1998).** ITS sequences from nuclear rDNA suggest unexpected phylogenetic relationships between Euro-Mediterranean, East Asiatic and North American taxa of *Quercus* (*Fagaceae*). Plant Syst Evol 211 : 129-139

**Schlötterer C, Tautz D (1994).** Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. Curr Biol 4 : 777-783

**Schneeberger RG, Cullis CA (1992).** Intraspecific 5S rRNA gene variation in flax, *Linum usitatissimum* (*Linaceae*). Plant Syst Evol 183 : 265-280

**Schondelmaier J, Schmidt T, Heslop-Harrison JS, Jung C (1997).** Genetic and chromosomal localization of the 5S rDNA locus in sugar beet (*Beta vulgaris* L.). Genome 40 : 171-175

**Schwarz O (1964).** *Quercus* L. Dans *Flora europaea*, vol. 1, première édition, pp 61-64. Edité par Tutin TG, Heywood VH, Burger NA, Valentine DH, Walters SM, Webb DA. Cambridge Universty Press, Cambridge

**Schwarz O (1993).** *Quercus* L. Dans *Flora europaea*, vol. 1, seconde édition, pp 72-76. Edité par Tutin TG, Burger NA, Chater AO, Edmonson JR, Heywood VH, Moore DM, Valentine DH, Walters SM, Webb DA. Cambridge Universty Press, Cambridge

**Scoles GJ, Gill BS, Xin ZY, Clarke BC, McIntyre CL, Chapman C, Appels R (1988).** Frequent duplication and deletion events in the 5S RNA genes and the associated spacer regions of the *Triticeae*. Plant Syst Evol 160 : 105-122

**Seperack P, Slatkin M, Arnheim N (1988).** Linkage disequilibrium in human ribosomal genes: implications for multigene family evolution. Genetics 119 : 943-949

**Snape JW, Flavell RB, O'Dell M, Hughes WG, Payne PI (1985).** Intrachromosomal mapping of the nucleolar organizer region relative to three marker loci on the chromosome 1B of wheat (*Triticum aestivum*). Theor Appl Genet 69 : 263-270

**Steinhoff S (1993).** Results of species hybridization with *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. Ann Sci For 50, suppl. 1 : 137s-143s

**Steinhoff S (1998).** Controlled crosses between pedunculate and sessile oak: results and conclusions. Allg Forst Jagdztg 169 : 163-168

**Swofford DL (1993).** PAUP: phylogenetic analysis using parsimony, version 3.1.1. Illinois Natural History Survey, Champaign, IL

**Szymanski M, Barciszewska MZ, Barciszewski J, Erdmann VA (1999).** 5S ribosomal RNA data bank. Nucleic Acids Res 27 : 158-160

**Tanksley SD, Ganal MW, Prince JP, de Vincente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messenguer R, Miller JC, Miller L, Paterson AH, Pineda O, Röder MS, Wing RA, Wu W, Young ND (1992).** High density molecular linkage maps of the tomato and potato genomes. Genetics 132 : 1141-1160

**Trontin JF, Grandemange C, Favre JM (1999).** Two highly divergent 5S rDNA unit size classes occur in composite tandem array in european larch (*Larix decidua* Mill.) and japanese larch (*Larix kaempferi* (Lamb.) Carr.). Genome 42 : 837-848

**Udovicic F, McFadden GI, Ladiges PY (1995).** Phylogeny of *Eucalyptus* and *Angophora* based on 5S rDNA spacer sequence data. Mol Phylogenet Evol 4 : 247-256

**Uzunova K, Palamarev E, Ehrendorfer F (1997).** Anatomical changes and evolutionary trends in the foliar epidermis of extant and fossil Euro-Mediterranean oaks (*Fagaceae*). Plant Syst Evol 204 : 141-159

**Varadaraj K, Skinner DM (1994).** Denaturants or cosolvents improve the specificity of PCR amplification of a G+C-rich DNA using genetically-engineered DNA-polymerases. Gene 140 : 1-5

**Venkateswarlu K, Lee SW, Nazar RN (1991).** Conserved upstream sequence elements in plant 5S ribosomal RNA-encoding genes. Gene 105 : 249-253

**Whittemore AT, Schaal BA (1991).** Interspecific gene flow in sympatric oaks. Proc Natl Acad Sci USA 88 : 2540-2544

**Zanetto A, Roussel G, Kremer A (1994).** Geographic variation of inter-specific differentiation between *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. Forest genet 1 : 111-123

**Zanetto A, Kremer A, Müller-Starck G, Hattemer HH (1996).** Inheritance of isozymes in Pedunculate oak (*Quercus robur* L.). J Hered 87 : 364-370

**Zhang Q, Saghai Maroof MA, Yang PG (1992).** Ribosomal DNA polymorphisms and the Oriental-Occidental genetic differentiation in cultivated barley. Theor Appl Genet 84 : 682–687

**Zhao X, Ji Y, Ding X, Stelly DM, Paterson AH (1998).** Macromolecular organization and genetic mapping of a rapidly evolving chromosome-specific tandem repeat family (B77) in cotton (*Gossypium*). Plant Mol Biol 38 : 1031–1042

**Zimmer EA, Martin SL, Beverley SM, Kan YW, Wilson AC (1980).** Rapid duplication and loss of genes coding for the α chains of hemoglobin. Proc Natl Acad Sci USA 77 : 2158-2162

**Zoldos V, Papes D, Brown SC, Panaud O, Siljak-Yakovlev (1998).** Genome size and base composition of seven *Quercus* species: inter-and intra-population variation. Genome 41 : 162-168

**Zoldos V, Papes D, Cerbah M, Panaud O, Besendorfer V, Siljak-Yakovlev (1999).** Molecular-cytogenetic studies of ribosomal genes and heterochromatin reveal conserved genome organization among 11 *Quercus* species. Theor Appl Genet 99 : 969-977

# CHAPITRE IV : Etude de l'ADNr 5S chez les mélèzes d'Europe et du Japon

## I- Les mélèzes

### 1- Importance, origine et distribution

Les mélèzes (*Larix*) constituent un petit groupe de gymnospermes (une dizaine d'espèces seulement, Debazac 1977, Farjon 1990) à aiguilles annuelles et caduques qui appartiennent comme les pins (*Pinus,* plus de 100 espèces), les épicéas (*Picea,* 40-50 espèces), les sapins (*Abies,* environ 40 espèces), les *Tsuga* (15 espèces), les sapins de Douglas (*Pseudotsuga*, 4-8 espèces), les cèdres (*Cedrus,* 4 espèces), l'unique espèce du genre *Pseudolarix* [*Pseudolarix amabilis* (Nels.) Rehd.] et quelques autres genres mineurs (*Cathaya*, *Hesperopeuce*, *Keteleeria*, *Nothotsuga*, Page 1990) à la grande famille des *Pinaceae* (environ 230 espèces, soit plus de 30 % des gymnospermes). L'histoire du genre *Larix* remonte probablement au Crétacé (fin de l'ère secondaire, entre -130 et -65 Ma) mais les données paléontologiques indiquent clairement que son expansion n'est intervenue qu'à partir du milieu de l'ère tertiaire (Eocène et Oligocène, entre -50 et -25 Ma). Elle aurait débuté en Amérique du Nord, origine géographique supposée des mélèzes. Les plus anciens restes fossiles connus proviennent en effet de l'ouest des Etats-Unis, dans la région centrale de l'Idaho (-45 Ma, Axelrod 1990, Schorn 1994) et de l'île canadienne Axel Heiberg, située au-delà du cercle polaire arctique (sédiments du milieu et de la fin de l'Eocène, entre -45 et -38 Ma, LePage et Basinger 1991). Dans le cas du continent eurasiatique, les plus vieux gisements fossilifères ont été découverts au nord-est de la Russie et datent seulement de l'Oligocène (entre -38 et -25 Ma).

Le genre est aujourd'hui largement représenté dans toutes les régions froides de l'hémisphère Nord (**Figure 36**), à basse altitude dans les régions septentrionales et dans les étages montagnards ou subalpins aux latitudes moyennes. La séparation des continents (en particulier la rupture de la Laurasie ou Euramérique) qui s'est achevée au début de l'Oligocène (-38 Ma) et, beaucoup plus récemment, les périodes glaciaires et interglaciaires qui se sont succédées au cours du pléistocène (entre -2 Ma et -18000 ans, maximum de la dernière glaciation) ont probablement largement contribué à la fragmentation actuelle de l'aire de répartition naturelle des mélèzes (Schorn 1994, LePage et Basinger 1995, Gernandt et Liston 1999, Semerikov *et al.* 1999). Deux grands groupes peuvent aujourd'hui être distingués sur des critères géographiques, morphologiques et génétiques : d'une part les espèces d'Europe et d'Asie et, d'autre part, les espèces d'Amérique du Nord (Debazac 1977, Farjon 1990, Semerikov et Lascoux 1999). Les espèces américaines sont au nombre de trois. *Larix laricina* (Du Roi) K. Koch a une répartition transcontinentale depuis la côte atlantique jusqu'à l'Alaska où une variété géographique (var.) a été décrite [var. *alaskensis* (Wight) Rau] tandis que les deux autres espèces ont une aire plus limitée dans le nord-ouest américain. *Larix occidentalis* Nutt. et *Larix lyalli* Parl. occupent respectivement les étages montagnard et subalpin du Sud de la Colombie Britannique (Canada) jusqu'aux chaînes du Washington, de l'Oregon, de l'Idaho et du Montana (Etats-Unis). Sur le vieux continent, certaines espèces ont une aire de répartition naturelle particulièrement vaste et des

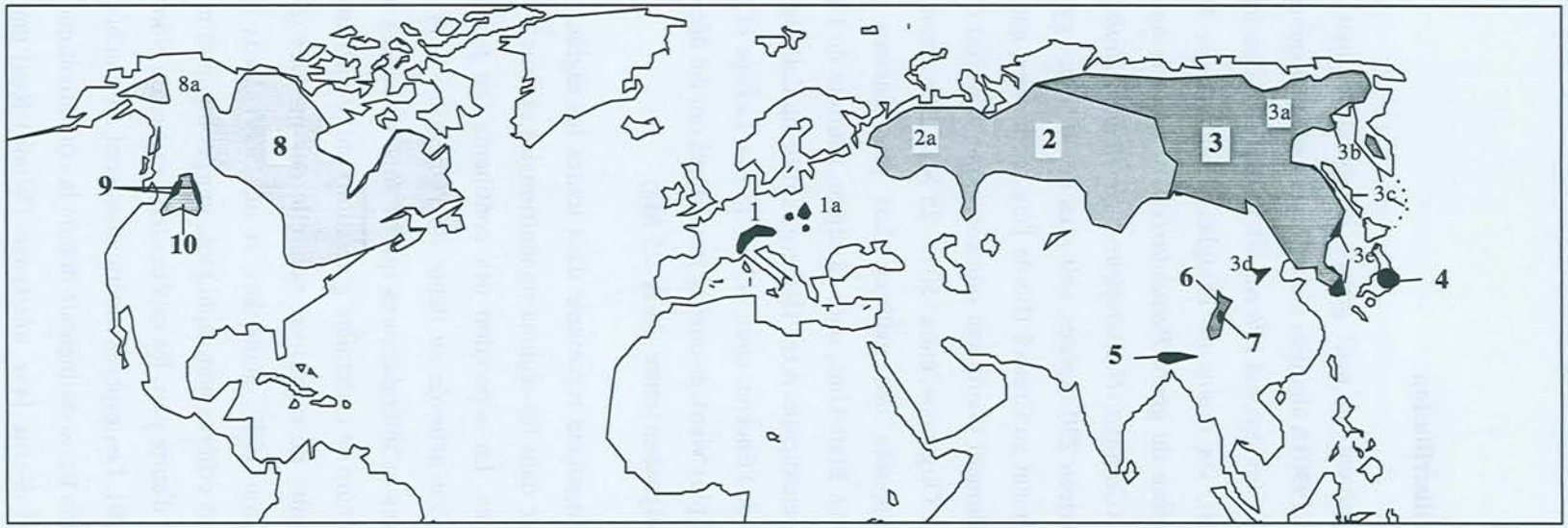



**Figure 36.** Aire de distribution des mélèzes dans le monde. Les taxa principaux (espèces reconnues) sont indiqués en gras. D'après Qian *et al.* (1995) et Semerikov *et al.* (1999).

**1**-*Larix decidua* ; 1a-var. *polonica* ; **2**-*L. sibirica* ; 2a-*L. sukaczewii* ; **3**-*L. gmelinii* ; 3a-var. *gmelinii* (*L. cajanderi*) ; 3b-var. *ochotensis* et *middendorfi* ; 3c-var. *japonica* (*L. kamtschatica*) ; 3d-var. *principis rupprechtii* ; 3e-var. *olgensis*, *komarovii* et *amurensis* ; **4**-*L. kaempferi* ; **5**-*L. griffithiana* ; **6**-*L. potaninii* ; **7**-*L. mastersiana* ; **8**-*L. laricina* ; 8a-var. *alaskensis* ; **9**-*L. lyalli* ; **10**-*L. occidentalis*.

variétés géographiques ou sous-espèces ont souvent été distinguées. *Larix sibirica* Ledeb. [= *Larix russica* (Endl.) Sabine ex Trautv] occupe ainsi tout le nord de la Russie d'Europe et la plaine de la Sibérie occidentale jusqu'au fleuve Ienisseï. En raison de leur forte différenciation morphologique et génétique , les populations de l'ouest (au-delà de l'Ob) sont souvent désignées sous le nom de *Larix sukaczewii* Dylis (Semerikov *et al.* 1999). *Larix gmelinii* (Rupr.) Kusen. a une distribution plus orientale. Il s'étend du plateau de la Sibérie centrale à la presqu'île du Kamtchatka à l'extrême nord-est (var. *gmelinii* = *L. cajanderi* Mayr), au littoral de la mer d'Okhotsk (var. *ochotensis* Kolesn. et *middendorfi* Kolesn.) et aux îles Kouriles et Sakhaline à l'est [var. *japonica* Max. = *L. kamtschatica* (Rupr.) Carr.]. Plus au sud, il rayonne jusqu'au nord de la Chine et de la Corée (var. *principis rupprechtii* Pilg.) et atteint le littoral de la mer du Japon (var. *Olgensis* Henry, *komarovii* Kolesn. et *amurensis* Kolesn.). Cinq autres espèces ont une aire de distribution beaucoup plus restreinte, voire localisée. *Larix kaempferi* (Lamb.) Carr. est endémique des zones montagneuses de l'Ile de Hondo au Japon (entre 1500 et 2000 m). *Larix griffithiana* (Lindl. et Gord.) Carr. est cantonné dans les hauteurs (2500-4000 m) de l'est de l'Himalaya (est du Népal, Sikkim et Bhoutan) tandis que *Larix potaninii* Batalin et *Larix mastersiana* Rehd. et Wils. occupent l'ouest de la Chine (en bordure du Tibet, entre 2500 et 3300 m d'altitude). *Larix decidua* Mill. enfin, occupe les montagnes de l'Europe centrale (1000-2500 m) et quelques stations de plaine en Pologne (var. *polonica* (Racib.) Domin.). C'est la seule espèce naturellement présente en France.

Les phénomènes d'hybridation interspécifique et d'introgression sont très courants dans les populations naturelles de mélèzes (Carlson et Theroux 1993, Qian *et al.* 1995, Lewandowski 1997, Semerikov *et al.* 1999). Des formes intermédiaires ont ainsi été observées dans les zones de sympatrie entre *L. occidentalis* et *L. lyalli* aux Etats-Unis et entre *L. potaninii* et *L. mastersiana* en Chine. De même, l'hybridation entre *L. sibirica* et *L. gmelinii* (qui s'étend depuis la péninsule de Taymir au nord, jusqu'à la Mongolie au sud) a permis l'émergence de l'espèce *Larix x czekanowskii* Szaf. (*L. gmelinii* X *L. sibirica*), très commune dans la région du lac Baïkal. Une origine hybride est par ailleurs suspectée pour bon nombre des principales variétés géographiques définies dans le cas des espèces confrontées à des milieux écologiquement variés. De récentes études allozymiques ont montré par exemple que la forme *polonica* de *L. decidua* est génétiquement apparentée à *L. sibirica* (Lewandowski 1997) et que plusieurs populations de mélèzes de l'est de l'Asie (var. *cajanderi*, *amurensis* et *ochotensis*) résultent de phénomènes d'hybridation introgressive entre la forme *olgensis* de *L. gmelinii* (qui est parfois élevée au rang d'espèce) et les autres variétés géographiques de cette espèce (Semerikov *et al.* 1999). Par ailleurs, des hybrides artificiels entre espèces allopatriques, comme *Larix x eurokurilensis* Rohm & Dimp. (*L. decidua* X *L. gmelinii* var. *japonica*), *Larix x eurolepis* Henry (*L. decidua* X *L. kaempferi*), *Larix x marschlinsii* Coaz (*L. kaempferi* X *L. sibirica*) ou bien encore *Larix x pendula* (Soland.) Salisb. (*L. decidua* X *L. laricina*) sont assez facilement obtenus lorsque les espèces parentales sont installées en peuplement mixte. Les barrières reproductives semblent donc très faibles dans le genre *Larix*.

**2- Classification du genre *Larix***

La classification des mélèzes est encore très largement discutée car les caractères morphologiques utilisés ne sont généralement pas suffisants pour distinguer avec précision les différents taxa. La présence d'hybrides interspécifiques ou de formes introgressées pose en outre de nombreux problèmes. Le statut (sous-espèce, espèce ou hybride) des différentes variétés géographiques décrites chez *L. laricina*, *L. decidua*, *L. sibirica* et surtout *L. gmelinii* restent par exemple difficiles à établir (Semerikov *et al.* 1999). Il en résulte que de nouvelles espèces sont encore "découvertes", comme *Larix kongboensis* localisé au sud-est du Tibet et auparavant confondu avec *Larix mastersiana* (Mill 1999). Selon les auteurs, le nombre d'espèce peut ainsi varier de 10 à plus d'une vingtaine.

Le genre *Larix* est actuellement divisé en deux sections, essentiellement sur la base de caractères concernant la morphologie des aiguilles et les inflorescences femelles matures (cônes), en particulier la taille des bractées associées aux écailles ovulifères (Farjon 1990, Schorn 1994). D'autres paramètres tels que la répartition géographique actuelle et diverses données fossiles sont également prises en compte (LePage et Basinger 1991, 1995). Les 6 espèces de la section *Multiserialis* (espèces de haute altitude et dont l'aire de répartition est assez réduite) possèdent des bractées saillantes, c'est-à-dire beaucoup plus longues (groupe I "*Aristatus*" américain: *L. lyalli* et *L. occidentalis*) ou légèrement plus longues (groupe "*Laminatus*" IIa asiatique : *L. griffithiana*, *L. mastersiana*, *L. potaninii*, *L. kaempferi*) que les écailles du cône qu'elles soutiennent. Chez les 4 espèces américaine (*L. laricina*) ou eurasiennes (*L. gmelinii*, *L. sibirica*, *L. decidua*) de la section *Pauciserialis* (ou *Larix*), les bractées sont en revanche plus courtes et donc incluses au sein du cône. Ces espèces vivant à plus faible altitude et dont l'aire de répartition est généralement plus étendue constituent le groupe "*Laminatus*" IIb. Comme les plus anciens restes fossiles des gisements de l'île Axel Heiberg possèdent des bractées courtes alors que ceux de l'Idaho présentent des bractées longues, la divergence de ces deux groupes remonterait à une époque antérieure à l'Eocène (LePage et Basinger 1991, 1995).

Les phylogénies établies récemment à partir de l'analyse RFLP de l'ADN chloroplastique (Qian *et al.* 1995, Kisanuki *et al.* 1995), des séquences nucléotidiques de la région des ITS de l'ADNr nucléaire (Gernandt et Liston 1999) et des variations allozymiques à 16 loci (Semerikov et Lascoux 1999) remettent en cause cette taxinomie basée sur des caractères morphologiques et en particulier sur la taille des bractées. Globalement concordantes, ces analyses moléculaires conduisent plutôt à diviser le genre en deux grands groupes géographiques qui correspondent respectivement aux espèces américaines et eurasiennes (discussion dans Gernandt et Liston 1999). Comme les bractées saillantes sont considérées comme un caractère ancestral car elles sont présentes à la fois dans le genre *Larix* et dans le genre très proche *Pseudotsuga* (caractère symplésiomorphique, Riesberg et Brunsfeld 1992), ces résultats semblent indiquer que la réduction de la taille des bractées se serait produite de manière indépendante en Amérique du Nord et en Eurasie (Gernandt et Liston 1999). Ces auteurs concluent qu'il n'est donc pas nécessaire de postuler l'existence de migrations relativement récentes (c'est-à-dire

jusqu'à la fin du Tertiaire, LePage et Basinger 1991, Schorn 1994), *via* un couloir situé à l'emplacement de l'actuel détroit de Béring, pour expliquer la présence d'espèces à bractées courtes sur les deux continents. Au-delà de cette partition générale en deux ensembles géographiques, la divergence nucléotidique au sein de chaque groupe est faible. Seul *L. griffithiana*, isolé dans les hauteurs de l'est de l'Himalaya, apparaît relativement distant des autres mélèzes eurasiens.

L'analyse de la variabilité allozymique (15 loci) de 32 populations de mélèzes d'Europe et du nord-est de l'Asie (*L. decidua*, *L. sibirica*, 6 variétés géographiques de *L. gmelinii*, *L. x czekanovskii*, *L. kaempferi*) révèle de même une faible différenciation interspécifique (Semerikov *et al.* 1999). Selon ces auteurs, cette faible différenciation génétique serait liée à la recolonisation post-glaciaire relativement récente (moins de 18000 ans), probablement à partir d'un petit nombre de refuges (voire un seul, situé au nord de la Sibérie où des vestiges âgés de 13000 à 21000 ans ont été découverts), et également à l'existence de flux de gènes entre les espèces (zones d'introgression). Une classification ascendante hiérarchique de ces mélèzes eurasiens a néanmoins pu être établie sur la base du calcul des distances génétiques (D, Nei 1987). *L. gmelinii* et la plupart de ses variétés ainsi que les populations de l'est du mélèze de Sibérie (*L. sibirica*) forment un groupe très peu différencié (D = 0,020 en moyenne) auquel se rattachent les populations de l'ouest du mélèze de Sibérie (D = 0,046) puis la variété *japonica* de *L. gmelinii* (D = 0,048), le mélèze d'Europe (D = 0,085) et enfin le mélèze du Japon (D = 0,133). La distance génétique la plus importante est observée entre ces deux dernières espèces (D = 0,171).

### 3- Les mélèzes d'Europe et du Japon

Comme la plupart des espèces du genre *Larix*, les mélèzes d'Europe (*Larix decidua* Mill. = *L. europaea* D.C.) et du Japon [*Larix kaempferi* (Lamb.) Carr. = *L. leptolepis* (Sieb. & Zucc)] présentent de nombreuses caractéristiques intéressantes pour les forestiers et les industriels de la filière bois. Le mélèze est ainsi parfois qualifié de "chêne des conifères" (Ferrand 1986, Collardet et Besset 1988, Pâques 1992, Arcade 1999 pour un bilan complet). C'est une essence pionnière héliophile (très tolérante à la lumière) de grande longévité (300 ans) qui résiste bien au froid et au vent (bon coefficient de forme) et s'adapte à une large gamme de sols qu'elle ne dégrade pas. Sa productivité moyenne est assez forte (plus de 10 $m^3$/ha/an). La branchaison fine permet un élagage naturel satisfaisant et l'exploitation est possible dès l'âge de 40 ans (croissance initiale rapide et rotation courte). Malgré un retrait un peu fort au séchage, le bois est généralement de très bonne qualité (constitué en grande partie de duramen à texture assez forte, mi-lourd et mi-dur), solide (propriétés mécaniques les meilleures parmi les conifères), durable (résistant aux agents chimiques et biologiques, imputrescible) et esthétique (duramen brun-rouge ou orangé, veinage décoratif).

En Europe occidentale, la plus grande part des besoins en cette essence appréciée sont couverts par le mélèze d'Europe et, dans une moindre mesure, par le mélèze du Japon. Ce sont aussi les deux seules espèces du genre utilisées pour le reboisement (le mélèze du Japon a été introduit en Europe au

19ième siècle). Toutes deux sont impliquées dans des programmes d'amélioration génétique (Pâques 1996a,b). La sélection porte essentiellement sur la résistance des branches et du tronc à la pezize (chancre provoqué par le champignon ascomycète *Lachnellula willkommii*), l'adaptation à différents types de sols défavorables (chimiquement pauvres, carbonatés), la résistance à des sécheresses estivales modérées (le mélèze du Japon ne tolère pas les étés secs), la croissance en hauteur et en volume ainsi que la rectitude du fût (souvent insuffisante à l'état naturel). En Europe centrale, les efforts d'amélioration sont essentiellement concentrés sur l'espèce autochtone *L. decidua* dont l'aire naturelle est divisée en 5 zones principales (**Figure 37**) : l'arc alpin (des Alpes maritimes françaises aux préalpes de Basse-Autriche), les monts Sudètes en République Tchèque, les Tatras en Pologne et en Slovaquie, les monts de Sainte-croix et Lysa Gora au centre de la Pologne et quelques provenances isolées dans les Alpes de Transylvanie, les Carpathes orientales et les monts Bihor en Roumanie. En France, le mélèze occupe encore une place modeste (100000 ha soit environ 2 % de la forêt résineuse) mais l'objectif principal de l'amélioration est son extension hors de son aire naturelle (Ferrand 1986, Pâques 1994) pour des reboisements en condition de plaine et de basse ou de moyenne montagne (diversification des espèces, remplacement des essences peu productives sur milieux marginaux, valorisation des terres libérées par l'agriculture). La variabilité naturelle du mélèze d'Europe est très importante. Un test comparatif (Pâques 1996a) a permis de montrer que les provenances les plus intéressantes pour le reboisement sont celles des sudètes (plasticité, niveau élevé et homogène de production, bonne résistance au chancre), du centre de la Pologne (var. *polonica*, bois de forte densité, très résistant au chancre) et de Basse-Autriche (rectitude du fût, bonne croissance). Plusieurs millions de plants sont produits chaque année en pépinière (Pâques 1996a).

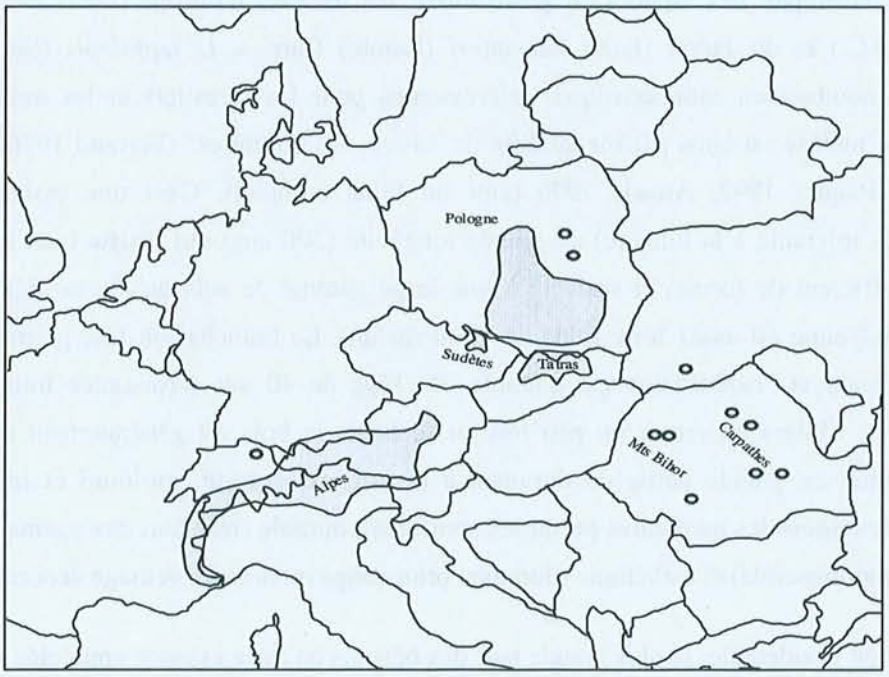



**Figure 37.** Aire de distribution naturelle du mélèze d'Europe (hachurée). D'après Pâques (1996a).

Une étude allozymique (7 loci) menée sur 7 populations représentatives de l'aire naturelle du mélèze d'Europe révèle également la forte diversité génétique de l'espèce (H = 0,223) qui réside essentiellement au sein des populations et très peu entre les populations (Maier 1992).

Le dendrogramme basé sur les distances génétiques indique toutefois que les provenances des Sudètes, de Pologne et des monts Tatras constituent un groupe relativement bien différencié parfois assimilé à une seule et même race dont l'origine hybride est suspectée (voir Lewandowski 1997). En comparaison, la variabilité naturelle du mélèze du Japon est faible pour la plupart des caractères d'intérêt économique (Pâques 1996b) et son organisation est complexe (ni clinale, ni écotypique, ni liée aux régions de provenance) ce qui ne facilite pas la sélection de matériel dans l'aire de répartition naturelle. Cette essence exotique originaire de l'île de Hondo n'est en effet représentée que par un ensemble de petites populations (seulement 7500 ha en tout) isolées sur les versants des principaux massifs volcaniques (entre 1200 et 2500 m d'altitude) eux-mêmes séparés par des vallées profondes (**Figure 38**).

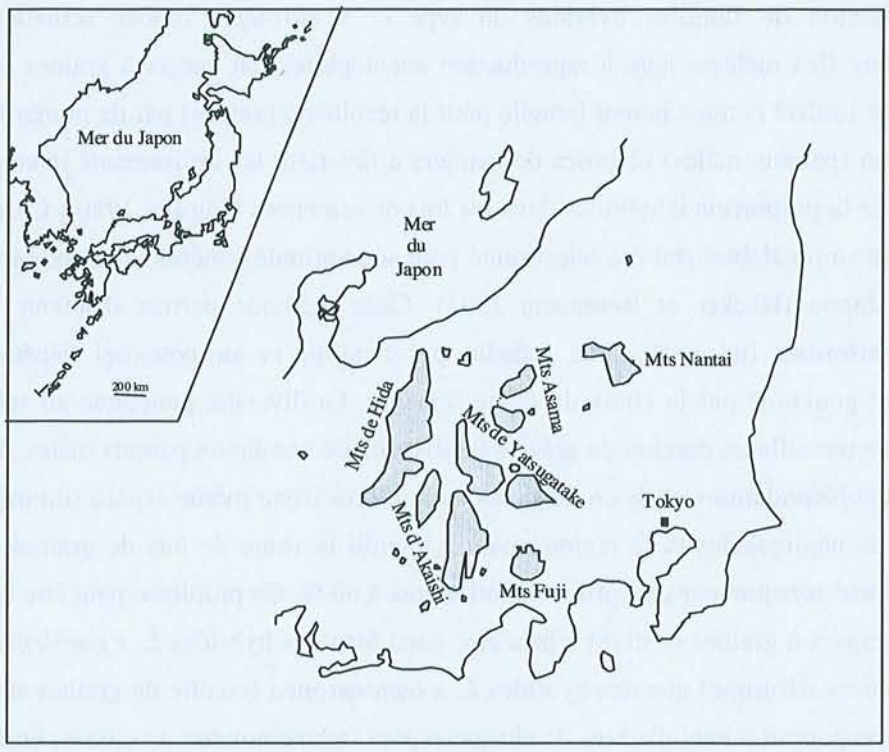



**Figure 38.** Aire de distribution naturelle du mélèze du Japon (hachurée). D'après Pâques (1996b).

Son utilisation pour le reboisement n'est envisageable que sur des sites à climat océanique en raison de son exigence en humidité atmosphérique et de sa sensibilité aux gelées de printemps (environ 12000 ha plantés en France, essentiellement dans le Limousin). En revanche, le mélèze du Japon présente un grand intérêt dans les programmes d'amélioration génétique par hybridation interspécifique avec le mélèze d'Europe (Pâques 1996b). Sa résistance presque absolue au chancre et

sa vigueur précoce sont en effet des caractères très recherchés, complémentaires des qualités du mélèze d'Europe comme la résistance à la sécheresse estivale, la bonne forme de tige et le bois d'excellente qualité technologique (Pâques 1989). Les croisements réalisés ont en effet montré que les hybrides F1 *L. x eurolepis* (*L. decidua* X *L. kaempferi*) sont particulièrement performant (en particulier pour la vigueur, la forme des fûts, l'homogénéité des plantations) par rapport aux espèces parentales (Bastien et Keller 1980, Pâques 1996a) même si cette supériorité ne s'exprime pas dans tous les milieux et qu'elle s'atténue parfois avec le temps (Pâques 1989). L'hybride réciproque *L. x leptoeuropea* Dengler (*L. kaempferi* X *L. decidua*) présente des caractéristiques similaires mais son obtention est plus difficile car les mélèzes sont protogynes (les gamètes femelles sont matures avant les gamètes mâles) et la floraison de *L. kaempferi* précède en général celle de *L. decidua* de quelques semaines (Pâques 1989, Ennos et Qian 1994). De nombreuses familles hybrides (plus de 600 selon Pâques 1989) sont actuellement testées par l'INRA sur une large gamme de sites potentiels pour le mélèze en France (une vingtaine, essentiellement dans le Massif central, à l'est et en Bretagne) afin de mettre les plus intéressantes à la disposition des reboiseurs (Le Cam 1981).

La production de familles hybrides du type *L. x eurolepis* repose actuellement sur la pollinisation libre (les mélèzes sont à reproduction anémophiles) en verger à graines d'un clone de mélèze d'Europe (utilisé comme parent femelle pour la récolte de graines) par de nombreux clones de mélèze du Japon (parents mâles) disposés de manière à favoriser les croisements interspécifiques et donc à optimiser la proportion d'hybrides dans les lots de semences (Chollet 1986). Chaque clone de mélèzes d'Europe a préalablement été sélectionné pour son aptitude générale à la combinaison[11] avec le mélèze du Japon (Häcker et Bergmann 1991). Cette méthode permet d'obtenir des familles relativement uniformes (un seul clone femelle par famille) et au potentiel génétique variable (essentiellement gouverné par le choix du clone femelle). La diversité génétique au sein de chaque famille se trouve par ailleurs maximisée grâce à l'utilisation de nombreux parents mâles. Toutefois, les possibilités d'autofécondations et de croisements entre arbres d'une même espèce (du même clone ou non) ne sont pas négligeables et la réglementation interdit la vente de lots de graines ou de plants certifiés "hybrides" lorsque leur proportion est inférieure à 60 %. Ce problème peut être très important dans certains vergers à graines destinés à produire aussi bien des hybrides *L. x eurolepis* (récolte des graines sur mélèzes d'Europe) que des hybrides *L. x leptoeuropea* (récolte de graines sur mélèzes du Japon) et qui comportent autant d'arbres de chaque espèce (même nombre de clones) en mélange pied à pied. L'identification des hybrides interspécifiques est donc une préoccupation majeure des améliorateurs dans ce programme.

Les paramètres morphologiques permettant de distinguer efficacement le mélèze d'Europe et le mélèze du Japon sont peu nombreux et seulement apparents au stade adulte (**Tableau 28**). Par ailleurs, les individus hybrides *L. x eurolepis* peuvent présenter non seulement des caractères

[11] L'aptitude générale à la combinaison (AGC) représente ce qu'un individu transmet en moyenne à ses descendants pour un caractère donné.

intermédiaires mais aussi des caractères plus ou moins similaires à l'une des deux espèces parentales. La sélection précoce des hybrides interspécifiques en pépinière sur la base de critères phénotypiques fiables est donc impossible et reste très difficile à l'âge mature. Au niveau cytologique, les caryotypes de *L. decidua* et de *L. kaempferi* sont similaires (**Tableau 29**). Selon Nkongolo et Klimaszewska (1995) cependant, certains chromosomes porteurs d'une constriction secondaire seraient suffisamment différenciés pour envisager leur utilisation comme marqueurs d'espèces. Il s'agit de la paire de chromosomes submétacentriques N°7 dans le cas de *L. decidua* (rapport bras court/bras long = 0,39) et de la paire de chromosomes métacentriques à submétacentriques N°9 concernant *L. kaempferi* (rapport bras court/bras long = 0,65).

**Tableau 28.** Les quelques caractères morphologiques distinctifs de *L. decidua* et *L. kaempferi* au stade adulte.

| Paramètre | *L. decidua* | *L. kaempferi* |
|---|---|---|
| Couleur des rameaux de l'année | Gris jaunâtre clair | Brun rougeâtre |
| Couleur des aiguilles | Vert clair | Vert bleuté |
| Forme du cône | Ovale | Globuleux |
| Aspect du bord supérieur des écailles ovulifères du cône | Arrondi ou tronqué | Emarginé et réfléchi |

D'après Debazac (1977).

**Tableau 29.** Structure et formule chromosomiques de *L. decidua* et *L. kaempferi*.

| Paramètre | *L. decidua* | *L. kaempferi* | Référence |
|---|---|---|---|
| Taille du génome (pg/2C)[a] | 22,9 (Fe)<br>19,7 (Fe) | nd | Greilhuber (1986)<br>Dhillon (1987) |
| Nombre de chromosomes | 24 | 24 | Khoshoo (1961) |
| Taille des chromosomes (µm) | 5,0-9,5 | 5,0-9,5 | Hizume (1988) |
| Formule chromosomique[b] | 6m+4m$^{cs}$+<br>2m/sm+2sm$^{cs}$+<br>8sm+2st/sm | 8m+2m$^{cs}$+<br>2m/sm+<br>4m/sm$^{cs}$+8sm | Hizume *et al.* (1993)<br>Nkongolo et Klimaszewska (1995)<br>Lubaretz *et al.* (1996) |

nd : non déterminé.

[a] Fe : microdensitométrie par la technique de Feulgen.

[b] m : métacentrique ; sm : submétacentrique ; cs : constriction secondaire ; st : subtélomérique

Toutefois, des problèmes de mesures de longueur et donc de caractérisation des chromosomes par différents auteurs ont été relevés (voir Lubaretz *et al.* 1996) et incitent à la prudence quant à l'identification des espèces sur cette base. Au niveau moléculaire, seuls deux loci isoenzymatiques (SKDH : shikimate déshydrogénase et NDH : NADH déshydrogénase) ont fourni des allèles utiles (spécifiques ou à fréquence spécifique) pour différencier plusieurs clones de mélèzes d'Europe et du Japon. Ils sont à l'origine des premières estimations de la proportion d'hybrides dans des lots de semences issus de vergers à graines allemands (68,6-84,6 % d'hybrides *L. x eurolepis*, Häcker et Bergmann 1991) et anglais (65 % d'hybrides *L. x eurolepis* et 16 % d'hybrides *L. x leptoeuropea*, Ennos et Qian 1994). Le petit nombre de clones analysés ne permet pas d'affirmer que ces allèles

constituent des marqueurs d'espèces, mais la récente étude des variations allozymiques dans les populations naturelles de plusieurs mélèzes eurasiens (Semerikov *et al.* 1999) confirme la présence d'allèles spécifiques chez *L. decidua* au locus SKDH. L'utilisation de sondes d'ADN génomique total révèle par ailleurs une forte homologie de séquence entre *L. decidua* et *L. kaempferi* (Nkongolo et Klimaszewska 1995). Treize marqueurs RAPD discriminants (*a priori* spécifiques) ont néanmoins pu être détectés par Arcade *et al.* (1996) sur la base d'un échantillon de 12 arbres représentatifs de chaque espèce. De même, Scheepers *et al.* (2000) ont récemment détectés 11 marqueurs RAPD qui permettent de différencier les deux espèces dont 4 apparaissent spécifiques et monomorphiques après avoir examiné 37 mélèzes d'Europe et 21 mélèzes du Japon de différentes origines géographiques. Trois ont été transformés en SCAR et séquencés. Pour deux d'entre eux, il s'agit de fragments du génome mitochondrial (à transmission maternelle, De Verno *et al.* 1993) de fonction inconnue. Le polymorphisme interspécifique révélé par le marqueur SCAR mitochondrial F-13 a été confirmé sur une centaine d'individus de différentes provenances pour chaque espèce (Acheré 1999). Cette étude a aussi permis de détecter sur ce même échantillon un marqueur chloroplastique diagnostique (à transmission paternelle, Szmidt *et al.* 1987). Il s'agit d'un fragment (LL) d'environ 600 pb correspondant à une partie du gène *rbc*L (qui code pour la grande sous-unité de la ribulose 1,5 bisphosphate carboxylase/oxydase) présentant un site de restriction reconnu par l'enzyme *Taq* I uniquement dans le cas du mélèze d'Europe. Les marqueurs cytoplasmiques F-13 et LL permettent de caractériser les individus hybrides et de déterminer également le sens du croisement. Des taux d'environ 40 % d'hybrides ont pu être déterminés dans plusieurs lots de semences par cette méthode (Acheré 1999). De même, Scheepers *et al.* (2000) ont pu détecter jusqu'à 84 % d'hybrides dans certains lots de graines en utilisant les 4 marqueurs RAPD spécifiques identifiés pour ces deux espèces.

Plusieurs autres régions du génome chloroplastique (9 fragments) et du génome mitochondrial (2 fragments) ont été testées par Acheré (1999) et ne révèlent aucun polymorphisme interspécifique, même au niveau nucléotidique. Des résultats comparables ont été obtenus par Vendramin *et al.* (1996) à l'aide de microsatellites chloroplastiques. Seul un locus (sur une vingtaine) pourrait permettre de différencier les deux espèces sur la base d'une différence de l'ordre d'une base. Le nombre d'arbres examinés (5 par espèce) est cependant trop faible pour conclure de manière définitive. Le polymorphisme interspécifique semble donc relativement faible entre mélèzes d'Europe et du Japon. En particulier, la recherche de marqueurs nucléaires informatifs de type ADN demeure infructueuse. L'élaboration de cartes génétiques a été récemment entreprise (Arcade *et al.* 2000). Elles permettront probablement de mieux comprendre la structure du génome de ces deux espèces phylogénétiquement très proches.

## II- Structure et organisation des unités d'ADNr 5S chez les mélèzes d'Europe et du Japon

Contrairement à la situation finalement assez simple qui a été décrite chez les chênes pédonculé et sessile (**Chapitre III**), les premiers essais d'amplification des espaceurs de l'ADNr 5S de différentes espèces de mélèzes (dont *L. decidua* et *Larix kaempferi*) réalisés par Trontin (1993) et Grandemange (1994) en utilisant les amorces universelles F et B (voir la **Présentation générale** de cette **Partie expérimentale**) ont révélé des profils complexes et difficiles à interpréter. Outre des multimères, ces profils comportent plusieurs niveaux de bandes dans la zone des monomères, certains correspondant *a priori* à différentes classes d'espaceurs de taille variable et d'autres à des artefacts de PCR du type hétéroduplexes. De plus, les analyses RFLP réalisées à l'aide de sondes d'ADNr 5S homologues donnent des résultats obscurs (digestions apparemment incomplètes, fragments d'origine inconnue) qui soulèvent d'importantes interrogations quant à l'organisation des unités d'ADNr 5S dans le génome de ces espèces.

**Tableau 30**. Dispositif expérimental utilisé pour l'étude de l'ADNr 5S chez *L. decidua* et *L. kaempferi*.

| Clone/génotype | N°* | Provenance | Nature | Source |
|---|---|---|---|---|
| *Larix decidua* | | | | |
| LdJB1 | 1 | France | Plantation | Jardin botanique (Nancy) |
| LdJB2 | 2 | France | Plantation | Jardin botanique (Nancy) |
| 92-1 | 3 | France | Lignée embryogène | M.A. Lelu |
| 107 | 4 | République Tchèque (Sudètes, Olomouc) | Forêt naturelle | L.E. Pâques |
| 285 | 5 | Pologne (Sudètes, Bystrzyca Klodsko) | Forêt naturelle | L.E. Pâques |
| 475 | 6 | Autriche (Alpes, Wienerwald) | Forêt naturelle | L.E. Pâques |
| 494 | 7 | Autriche (Alpes, Wienerwald) | Forêt naturelle | L.E. Pâques |
| 500 | 8 | Autriche (Alpes, Wienerwald) | Forêt naturelle | L.E. Pâques |
| *Larix kaempferi* | | | | |
| 3005 | 9 | Japon (Forêt de Eu) | Forêt naturelle | L.E. Pâques |
| 3009 | 10 | Japon (Forêt de Eu) | Forêt naturelle | L.E. Pâques |
| 3211 | 11 | Japon (Mont Fuji) | Forêt naturelle | L.E. Pâques |
| 3212 | 12 | Japon (Ina) | Forêt naturelle | L.E. Pâques |
| 3217 | 13 | Japon (Kamikochi) | Forêt naturelle | L.E. Pâques |
| 3242 | 14 | Belgique (Vivierfays) | Forêt artificielle | L.E. Pâques |
| 3247 | 15 | Japon (Okunikko) | Forêt naturelle | L.E. Pâques |
| 3249 | 16 | Japon (Okunikko) | Forêt naturelle | L.E. Pâques |
| LkJB1 | 17 | France | Plantation | Jardin botanique (Nancy) |
| 3076 | / | Danemark | Forêt artificielle | D. Prat |

* Numérotation utilisée dans la **Figure 41**.

L'objectif principal de cette partie est donc de préciser et de comparer la structure et l'organisation des unités d'ADNr 5S chez les mélèzes d'Europe et du Japon. Pour ce faire, nous avons

utilisé un dispositif expérimental constitué, comme celui exploité par Arcade *et al.* (1996) pour rechercher des marqueurs moléculaires de différenciation interspécifique, d'une dizaine d'arbres représentatifs de chaque espèce (**Tableau 30**).

Comme dans la **Partie II** du **Chapitre III**, nous avons procédé dans un premier temps au tri des fragments amplifiés chez les deux espèces à l'aide des amorces universelles P1 et P2 (espaceurs, hétéroduplexes, multimères). Nous avons ensuite cloné et séquencé les différentes classes d'espaceurs détectées. Ces données ont été utilisées pour décrire l'évolution structurale intragénomique des espaceurs et synthétiser un nouveau couple d'amorces spécifiques permettant d'amplifier le gène 5S et de le séquencer. Une comparaison-bilan avec les quelques séquences de gènes 5S et d'espaceurs disponibles chez les gymnospermes a été réalisée afin de préciser la topographie et l'importance des variations des unités d'ADNr 5S dans ce groupe de plantes. Nous avons en particulier cherché à identifier les zones homologues entre espèces qui présentent un intérêt phylogénétique potentiel. Enfin, l'utilisation comme sondes de différents fragments amplifiés et clonés chez un mélèze d'Europe de référence nous a permis i) d'obtenir une estimation du nombre d'unités et ii) d'explorer leur organisation par analyse RFLP de l'ADN génomique.

Ces résultats sont en grande partie présentés et discutés dans l'article qui suit (**Partie A**), publié dans la revue *Genome* en octobre 1999 [42 (5) : 837-848]. Nous en présenterons ensuite une première extension (**Partie B**) basée sur la mise en évidence d'un motif conservé au sein des espaceurs de mélèzes et d'autres gymnospermes.

## A- Two highly divergent 5S rDNA unit size classes occur in composite tandem array in European larch (*Larix decidua* Mill.) and Japanese larch (*Larix kaempferi* (Lamb.) Carr.)

Jean-François Trontin, Catherine Grandemange, and Jean-Michel Favre

### 1- Abstract

The 5S ribosomal DNA unit structure and organization have been investigated in *Larix decidua* and *Larix kaempferi* using selective amplification of gene and spacer, sequence analysis and homologous probe hybridization. Two highly divergent unit size classes of approximately 650 bp and 870 bp were detected in both species. Sequence analysis in *Larix decidua* revealed that length variations occur in the middle spacer region and are the result of duplications (in the long spacers) and considerable sequence heterogeneity. Conversely, the transcribed region is of uniform length (120 bp) and the nucleotide sequence of one *Larix decidua* clone is similar to that reported for other gymnosperms. Sequence comparison of the larch spacers with two other *Pinaceae* species (*Pinus radiata* and *Picea glauca*) showed that the 5' and 3' regions flanking the gene (40 and 60 bp, respectively) are quite conserved, suggesting a regulatory role. Moreover, a small element of about 70 bp located in the middle spacer region was found to be common to the larch long units and the six *Pinus radiata* spacer clones previously sequenced (64% sequence identity). The short and long unit size classes are mainly organized in composite tandem array(s) with evidence of extensive zones of strict alternation in both species. Mechanisms underlying this unusual association of divergent units in larch 5S rDNA arrays are discussed.

***Key words***: 5S rRNA genes, spacer variations, cluster organization, sequence comparison, Gymnosperms.

**2- Introduction**

In higher plants, the nuclear 5S ribosomal RNA genes (5S rDNA) are usually organized into clusters of tandem repeats (Lapitan 1992) that occur at one [e.g. *Beta vulgaris* (Schondelmaier *et al.* 1997)] or several [e.g. members of the tribe *Triticeae* (Scoles *et al.* 1988; Dvorák *et al.* 1989)] chromosomal loci. Each repeat unit consists of a 120 base pairs (bp) coding region (5S gene) and a non-transcribed spacer region (NTS) of about 100-700 bp. The total number of units per haploid genome ranges from 1000 (e.g. *Arabidopsis thaliana*, Campell *et al.* 1992) to over 50 000 (e.g. *Linum usitatissimum*, Goldsbrough *et al.* 1981).

The 5S rDNA unit structure and organization have been extensively studied in dozens of angiosperm species belonging to several important families such as *Brassicaceae*, *Fabaceae*, *Linaceae*, *Poaceae*, *Malvaceae*, etc (Sastri *et al.* 1992 for a review; Kellogg and Appels 1995; Cronn *et al.* 1996). This accumulating database demonstrates that the 5S genes are highly conserved across species, presumably reflecting the function of the small 120 nucleotides (nt) 5S rRNA as a component of the large subunit (60S) of the ribosomes, whereas the intergenic spacer is more variable. However, within individual array, the repeat units usually exhibit high homogeneity even though moderate levels of length and sequence variations have been reported from several plant groups (Gottlob-McHugh *et al.* 1990; Kellogg and Appels 1995; Cronn *et al.* 1996). As in 18S-5.8S-26S rDNA and other tandemly repeated multigene families, such a maintenance of array homogeneity results from intralocus concerted evolution of repeats (Arnheim 1983) and involves various molecular processes such as repeated cycles of unequal crossing-over and gene conversion (Smith 1976; Dover 1982; Nagylaki 1984).

The 5S rDNA structure and organization have been comparatively less investigated in gymnosperms. Detailed reports, including sequence data, are only available for the two *Pinaceae* species *Pinus radiata* (Cullis *et al.* 1988; Gorman *et al.* 1992; Moran *et al.* 1992) and *Picea glauca* (Brown and Carlson, 1997; Brown *et al.* 1998). As previously found in the angiosperms, the 5S rDNA units are clustered in tandem arrays and the coding region is 120 bp long.

In *Pinus radiata*, two 5S rDNA unit classes of about 525 and 850 bp in length were identified (Moran *et al.* 1992). The sequence identity between these two classes is high (near 78%) and the main source of size difference is a 330-bp insertion in the spacer region of the long units. With respect to units organization, Gorman *et al.* (1992) showed that tandem arrays of various sizes, containing relatively few 5S repeats per array, are widely dispersed across the chromosomal complement. Such a scattered organization has only been previously observed in rye (Reddy and Appels 1989) and flax (Schneeberger *et al.* 1989). However, the respective organization of the short and long unit classes remains to be elucidated.

In *Picea glauca*, the 5S rDNA units also occur in two size classes of 221 bp and approximately 600 bp in length (Brown and Carlson 1997). The relationship between these two size classes is unresolved since sequence information about the 600-bp class is lacking. In contrast with *Pinus radiata*, only one 5S rDNA locus has been detected (Brown and Carlson 1997; Brown *et al.* 1998) by means of fluorescence *in situ* hybridization (FISH). According to these authors, both 5S rDNA unit classes identified in *Picea glauca* are probably clustered at this unique chromosomal locus, either in adjacent arrays or, to some extent, interspersed.

Such a highly clustered organization of 5S rRNA genes appeared to be prevalent in the *Pinaceae*, the exception being *Pinus radiata*. Recent FISH studies indeed reported only one 5S rDNA locus in *Picea sitchensis* (Brown and Carlson 1997; Brown *et al.* 1998), *Picea abies* and *Larix decidua* (Lubaretz *et al.* 1996), 3 loci in *Pinus elliottii* (Doudrick *et al.* 1995) and 4 loci in the chromosomal complement of *Pinus sylvestris* (Lubaretz *et al.* 1996).

Additional data in other gymnosperms and related plants are still limited to the size of 5S repeats in *Juniperus virginiana* (378 bp) and *Cycas revoluta* (1572 bp; Gottlob-McHugh *et al.* 1990). Some sequence information concerning the 5S gene product (5S rRNA) is also available in *Ginkgo biloba*, *Cycas revoluta* and *Metasequoia glyptostroboides* (Hori *et al.* 1985), *Ephedra kokanica* and *Gnetum gnemon* (Melekhovets *et al.* 1988), *Pinus sylvestris* (Mashkova *et al.* 1990), *Pinus contorta*, *Taxus baccata* and *Juniperus media* (Van de Peer *et al.* 1990).

The present study is part of an ongoing effort to provide new informations about the 5S rDNA structure and organization in gymnosperms, specifically in *Pinaceae*. We focused on another genus, *Larix* Mill., which includes about ten species distributed throughout the colder regions of the Northern Hemisphere, chiefly in mountainous areas. The European larch (*Larix decidua* Mill.) is the unique *Larix* species represented in western Europe, particularly in the Alps, Sudetans, Tatras, central Poland and Carpathians. This valuable forest tree, fast growing with high timber quality, is concerned with both classical (Pâques 1996a) and molecular breeding programs (Shin *et al.* 1994), including interspecific hybridization with the Japanese larch (*Larix kaempferi* (Lamb.) Carr.). The superiority with regard to growth performances of the resultant hybrid larch, *Larix x eurolepis* Henry, has been widely documented (Pâques 1989 for a review) and correlated with high levels of heterozygosity (Arcade *et al.* 1996).

Concerning the larch genome, progress in constructing the karyotypes and studying the molecular relationships between *Larix decidua*, *Larix kaempferi* and their hybrid have recently been reported (Nkongolo and Klimaszewska 1995; Arcade *et al.* 1996).

In this paper, the 5S rDNA units structure and organization in *Larix decidua* and *Larix kaempferi* are described in terms of amplification patterns, nucleotide sequence, and composition of tandem arrays.

## 3- Materials and methods

### a- Plant material and DNA isolation

5S rDNA analyses were performed on 8 European larch genotypes and 9 Japanese larch genotypes obtained from the Institut National de la Recherche Agronomique (INRA) station of Orléans or collected in the Botanical Garden of Nancy (France). Genomic DNA was isolated from needles according to Doyle and Doyle (1987) with 1% 2-mercaptoethanol (v/v) and 1% polyvinylpyrrolidone 40 (w/v) added to the extraction buffer.

### b- DNA amplification, electrophoresis and cloning

Based on a comparison of compiled nuclear 5S rRNA and 5S gene sequences (5S rRNA Data Bank, Szymanski *et al.* 1998), two universal plant primers (P1 and P2) were designed to selectively amplify the spacer region (**Figure 39**). The sequences and positions of these primers in the 5S gene region (ca. 120 pb) were as follows:

P1: 5'-TGGGAAT*TCCTCGTGTTGCA-3' (nt 96 to 115)

P2: 5'-ATTAGA*GCTC*GTATGATCGC-3' (nt 24 to 5)

Base changes (*) were introduced in the consensus sequences in order to generate the *Eco*R I and *Sac* I restriction sites (underlined) in P1 and P2, respectively.

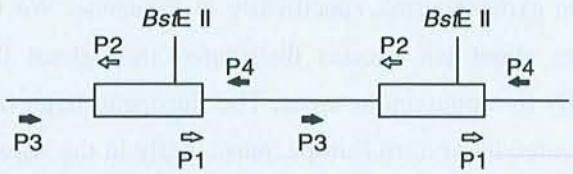



**Figure 39.** Location of the primers used for amplification of larch 5S rDNA from genomic DNA.
The 5S gene region is shown as a box and the spacer as a line. The 5S repeat unit is defined with *Bst*E II. This enzyme has a single recognition site (GGTGACC) within the gene region in larch (nucleotides 86-92).

Two additional primers (P3 and P4) were designed in the spacer region to amplify the 5S gene (**Figure 39**). The following sequences of these primers were deduced from the short spacer sequences previously amplified with the P1 and P2 primers:

P3: 5'-ACGGTCGAC*CCTCTTGGTTG-3' (nt -81 to -62 upstream from 5S genes)

P4: 5'-GCCCCGCGG*TTTCCCTCATC-3' (nt +45 to +26 downstream from 5S genes)

As in P1/P2, two restriction sites *Sal* I and *Sac* II were, respectively, introduced in P3 and P4.

PCR amplifications were carried out in a 25 µl reaction volume containing 0.1-1 ng genomic DNA, 0.2 µM primers P1 and P2 or 0.4 µM primers P3 and P4, 200 µM dNTPs, 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2 mM $MgCl_2$, and 0.6 unit Taq DNA polymerase (Gibco BRL). After a denaturing step at 96°C for 3 min, 30 amplification cycles were performed on a PTC-100 thermal cycler (MJ Research, Inc.), each consisting of denaturation at 94°C for 40 sec, annealing at 53°C for 1 min and primer extension at 72°C for 1 min (P1/P2) or for 30 sec (P3/P4). Amplification was ended by a final extension step at 72°C for 10 min. The amplified fragments were electrophoresed for 6 h at 3 $V.cm^{-1}$ through 1.6% agarose gels made up in 0.5x TBE (45 mM Tris-borate, 1 mM EDTA).

For cloning purposes, total PCR mixtures were purified by ethanol precipitation and digested with the appropriate restriction enzymes. The restricted fragments were then electrophoresed, gel purified and ligated to the phagemid vector pBluescript KS +/- (Stratagene). DH5α line cells of *E. coli* (Life Technologies Inc.) were transformed using the calcium chloride protocol of Sambrook *et al.* (1989) and plasmid extractions from recombinant clones were performed using the Qiagen kits. Dideoxy chain-terminating sequencing reactions were carried out using T3 and T7 dye primers (protocol from Applied Biosystems) and analyzed on an ABI 373 automated sequencer (Genome express SA; Grenoble, France).

c- Sequence alignment and analysis

The larch 5S rDNA sequences were subjected to a similarity search against the standard sequence databases (Genbank, EMBL and DDBJ) using the NCBI (National Center for Biotechnology Information) BLAST network service and programs (Altschul *et al.* 1990).

Sequence alignments were performed either by manual adjustment using SeqApp version 1.9a157 (Don Gilbert, Indiana University) or using the CLUSTAL V program (Higgins *et al.* 1992).

Phenetic analyses were carried out on the aligned data using PHYLIP (Felsenstein 1995). Distances were first computed between all pairs of sequences using the 2-parameter method of Kimura (1980). The resulting distance matrix was then subjected to the Neighbor-Joining method of clustering (Saitou and Nei 1987).

d- Hybridization and copy number estimation

Standard genomic DNA preparations were digested with *Bgl* II or *Bst*E II restriction enzymes and electrophoresed through 1.6% agarose gels at 3 $V.cm^{-1}$ for 6 h. Capillary transfer of restricted DNA to positively charged nylon membrane (Euroblot) was performed for 12 h in 0.4 M NaOH. Two 5S rDNA spacer clones amplified from *Larix decidua* using primers P1/P2 (5SLd-S2 and 5SLd-L1) were excised from the pBluescript vectors pB5SLd-S2 and pB5SLd-L1, gel purified and digoxigenin-labeled (DIG DNA labeling and detection kit, Boehringer Mannheim). Hybridizations with these two probes (named p5SLdS2 and p5SLdL1) were carried out at 42°C overnight in standard formamide

buffer [5x SSC (0.75 M NaCl, 0.075 M sodium citrate), 50% formamide (v/v), 0.1% N-lauroylsarcosine (w/v), 0.02% sodium dodecyl sulfate (w/v), 2% blocking reagent (w/v) from Boehringer Mannheim]. Filters were subsequently washed twice for 5 min at room temperature in 2x SSC, 0.01% SDS and twice for 15 min at 65°C in 0.1x SSC, 0.01% SDS. Immunodetection and visualization with the colorimetric substrates NBT/X-phosphate were performed according to the manufacturer's instructions.

To estimate the 5S rDNA copy number, pure genomic DNA was extracted using the DNeasy plant mini kit (Qiagen). Precise DNA concentrations were measured by UV absorption (Genequant, Pharmacia) and fluorometry (TKO 100, Hoefer) using calf thymus DNA as internal standard (Sigma). Aliquots of genomic DNA (0.5 and 1.0 µg) isolated from *Larix decidua* were digested with *Bst*E II (only one restriction site in the 5S gene), then electrophoresed together with dilutions of linearized vectors pB5SLd-S2 and pB5SLd-L1 corresponding to the expected weights of 500, 1000, 2000, 4000, 6000, 8000 and 10000 copies and transferred on a nylon membrane as described above. One set of these filters was first hybridized with the p5SLdS2 probe and another set with the p5SLdL1 probe. After scanning, the filters were analyzed using the NIH image software (V1.61 by Wayne Rasband, National Institutes of Health, USA).

**4- Results**

a- 5S gene

PCR amplification using the primer combination P3/P4 produced the expected 250 bp fragment containing the 5S gene in both European and Japanese larches (data not shown). One clone from *Larix decidua* (5SLd-G) was sequenced (EMBL accession number: AJ010579). By comparison with the 9 published gymnosperm 5S rRNA sequences (Hori et al 1985, Melekhovets *et al.* 1988, Mashkova *et al.* 1990; Van de Peer *et al.* 1990) we assumed the coding region to be 120 bp long with "GGG" as the start and "CTC" as the end. The inferred 5S gene sequence could be folded into the generalized secondary structure model proposed for plant 5S rRNAs by Barciszewska *et al.* (1994) and the tertiary interactions between loops C/D and between loops B/E proposed by these authors could also be verified (**Figure 40**). Therefore, it can be concluded that the analyzed 5S gene clone probably represents a real transcribed gene.

Sequence comparison of the larch 5S gene with all the previously sequenced gymnosperm 5S rRNA or 5S genes (16 sequences, 11 species) revealed high sequence identity (89% to 98%) and 35 variable positions (**Table 31**). Most of them were transitions (68.5%). An explanation for the low frequency of transversions (20.0%) at these positions may be that once propagated and homogenized across the 5S array(s), transversions may alter the ribosome function and reduce relative fitness (Cronn *et al.* 1996). **Figure 40** shows that these variable positions are more often localized in the double stranded stem regions (19 sites) than in the single stranded loop regions (12 sites). According to Barciszewska *et al.* (1994), substitutions in the stems may represent a potential source of alteration

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY



of the 5S rRNA secondary structure. However, it can be easily observed (**Table 31**) that a number of stem positions (20 and 58 in stem II; 28 and 51 in stem III; 80 and 96 in stem IV; 67-69 and 106-108 in stem V) probably undergo compensating substitutions as mentioned by Steele *et al.* (1991). This phenomenon is likely to limit the impact of stem mutations on the secondary structure.

**Table 31.** Sequence comparison of the *Larix decidua* 5S gene cloned (first line) with the 16 gymnosperm 5S rRNAs or 5S genes published to date.

| Species | Sequence identity (%) | Nucleotide position | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 2 | 10 | 17 | 20 | 22 | 23 | 24 | 26 | 28 | 31 | 36 | 49 | 51 | 54 | 58 | 67 | 68 | 69 | 73 | 77 | 80 | 83 | 84 | 85 | 87 | 95 | 96 | 106 | 107 | 108 | 110 | 112 | 119 | 120 | 121 |
| *Larix decidua* 5SLd-G | | G | C | G | T | A | A | T | C | C | G | C | C | G | A | A | C | T | A | T | G | G | A | T | G | G | C | C | T | A | G | G | T | T | C | - |
| *Pinus radiata* 5SPr6[a] | 98.3 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | g | . | . | . | . | a | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *Pinus radiata* 5SPr7[a] | 98.3 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | g | . | . | . | . | a | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *Pinus radiata* 5SPr9[a] | 98.3 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | g | . | . | . | . | a | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *Ephedra kokanica*[b] | 98.3 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | g | . | . | . | . | a | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *Picea glauca*[c] | 97.5 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | g | . | . | . | . | a | . | . | . | . | . | . | . | . | . | . | . | . | . | t | . |
| *Pinus contorta*[d] | 97.5 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | g | . | . | . | . | a | . | . | . | . | . | . | . | . | . | . | . | . | . | . | c |
| *Pinus radiata*[e] | 96.6 | . | . | . | . | . | . | . | . | . | . | . | t | . | . | g | . | . | . | c | a | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *Pinus radiata* 5SPr5[a] | 96.6 | . | . | . | . | . | . | . | . | . | . | g | . | . | . | g | . | . | . | . | a | c | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *Pinus radiata* 5SPr10[a] | 96.6 | . | . | . | . | . | . | . | . | . | a | . | . | . | t | g | . | . | . | . | a | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *Pinus sylvestris*[f] | 96.6 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | g | . | . | . | . | a | . | . | . | . | . | . | . | . | . | . | . | . | c | - | . |
| *Ginkgo biloba*[g] | 95.8 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | g | . | a | a | . | . | . | . | . | u | c | . | . | . | . | . | . | . |
| *Metasequoia glyptostroboides*[g] | 92.5 | . | . | . | . | . | g | . | . | . | . | . | . | . | . | g | . | c | g | . | a | . | . | . | . | . | . | . | c | g | . | a | . | . | u | . |
| *Cycas revoluta*[g] | 92.5 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | g | u | . | g | . | a | a | . | . | . | . | . | u | . | . | a | a | . | . | u | . |
| *Taxus baccata*[d] | 90.9 | a | . | . | . | u | g | . | . | . | . | . | . | . | . | g | . | c | . | . | a | . | . | . | . | . | . | . | c | g | . | . | c | . | u | u |
| *Juniperus media*[d] | 90.0 | . | . | . | . | . | g | . | . | . | . | . | . | . | . | g | . | c | g | . | a | . | g | g | a | u | . | . | c | g | . | . | . | . | u | . |
| *Gnetum gnemon*[b] | 89.1 | . | a | c | c | . | . | c | u | u | . | g | u | a | . | g | . | . | . | . | a | . | . | . | . | u | . | . | . | . | . | . | . | a | . | . |

**Note**: Only the 35 variable nucleotide positions with the corresponding substitutions (lower-case letters) are indicated for each species. Periods represent residues identical to those of the reference 120 bp sequence from *Larix decidua* and dashes indicate alignment gaps.

[a]Moran *et al.* (1992) ; [b]Melekhovets *et al.* (1988) ; [c]Brown and Carlson (1997) ; [d]Van de Peer *et al.* (1990) ; [e]Gorman *et al.* (1992) ; [f]Mashkova *et al.* (1990) ; [g]Hori *et al.* (1985).

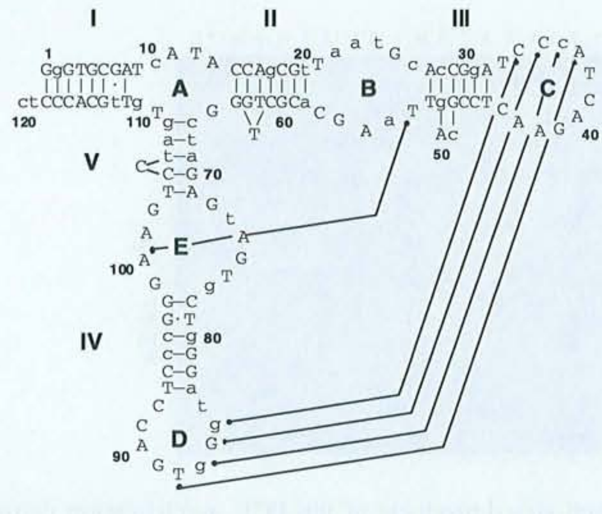



**Figure 40.** Nucleotide sequence of the *Larix decidua* 5S gene (clone 5SLd-G) arranged into the secondary structure model of plant 5S rRNA described in Barciszewska *et al.* (1994). A, B, C, D and E: loop regions; I, II, III, IV and V: stem regions. The tertiary interactions between loops C and D, and between loops B and E are shown as thick lines. The variable nucleotide positions identified by comparison with the 16 gymnosperm 5S rRNAs or 5S rRNA genes published to date (11 species) are indicated in lower-case letters.

Sequence identity was lower with angiosperms. Interestingly, the highest score (91%) was found with two *Magnolia* species (Nalaskowska *et al.* 1990) belonging to the order Magnoliales which is classically recognized as one of the most primitive group of angiosperms (Loconte, 1996).

b- 5S rDNA spacer

α- Amplification

Primers P1 and P2 were used to amplify the 5S rDNA spacers from *Larix decidua* (8 individuals) and *Larix kaempferi* (9 individuals). Results are shown in **Figure 41**. Both species exhibited the same pattern of bands. Three major size classes of PCR products, approximately 580 bp (S), 800 bp (L) and 1450 bp (dSL), were observed for all individuals. Within each size class, two bands were clearly visible for most individuals. In addition to the S, L and dSL main products, lesser amount of 650-bp (hS) and 850-bp fragments (hL) were also observed. Re-amplification of the hS and hL fragments produced the S and L fragments respectively, consistent with hS and hL resulting from heteroduplexe associations within each of the S and L size classes. Re-amplification of the dSL fragments produced both S and L fragments. Cloning and subsequent restriction analysis clearly identified these products as dimers between the S and L fragments separated by the 5S gene (data not shown).

Finally, the S and L PCR products were identified as two putative 5S rDNA spacer classes. After correction for primer length, the short (S) and long (L) spacer classes were estimated to be approximately 530 bp and 750 bp long in European and Japanese larches.

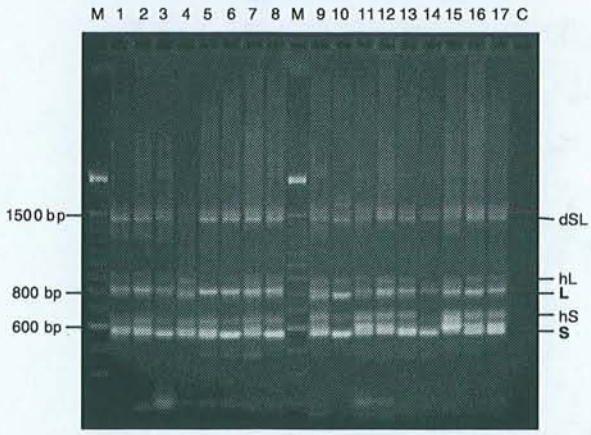



**Figure 41.** Gel-fractionated products of the PCR amplification using the P1/P2 primer combination. Lanes M, 100-bp DNA ladder; lanes 1-8: *Larix decidua*; lanes 9-17: *Larix kaempferi*; lane C, negative control (no DNA template).

β- Sequence analysis in *Larix decidua*

Three clones of the short spacer class ("5SLd-S") and two clones of the long spacer class ("5SLd-L") were sequenced from genotype 92-1 (5SLd-S1; see **Figure 41**, lane 3) and genotype 107 (5SLd-S2, 5SLd-S3, 5SLd-L1 and 5SLd-L2; see **Figure 41**, lane 4). The EMBL accession numbers, lengths, GC-contents and pairwise comparisons are indicated in **Table 32**.

**Table 32.** EMBL accession number, length, GC-content and pairwise comparison of the short (5SLd-S) and long (5SLd-L) spacer clones sequenced in *Larix decidua*.

| Accession number | Length (bp) | GC-content (%) | Clone designation | Similarity score (%)* | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | 5SLd-S1 | 5SLd-S2 | 5SLd-S3 | 5SLd-L1 | 5SLd-L2 |
| AJ010574 | 514 | 45.1 | 5SLd-S1 | 100 | | | | |
| AJ010575 | 517 | 49.1 | 5SLd-S2 | 75.6 | 100 | | | |
| AJ010576 | 539 | 49.5 | 5SLd-S3 | 80.9 | 88.3 | 100 | | |
| AJ010577 | 738 | 61.8 | 5SLd-L1 | 25.5 | 32.2 | 34.7 | 100 | |
| AJ010578 | 767 | 61.9 | 5SLd-L2 | 27.4 | 32.8 | 34.3 | 91.5 | 100 |

* As implemented in the CLUSTAL V program (Higgins *et al.* 1992) using the default options

The percentage of GC content (GC%) of the long larch spacer clones is high (near 62% on average) and similar to values obtained from all other published gymnosperm 5S rDNA spacer sequences, namely 6 clones from radiata pine (62% on average; Gorman *et al.* 1992; Moran *et al.* 1992) and one clone from white spruce (65%; Brown and Carlson 1997). In contrast, the GC% of the short larch spacer clones is lower (near 48% on average). Within each size class, alignment was straightforward and pairwise comparison resulted in high similarity scores, i.e. 81% on average in the short spacer class and 91% in the long spacer class. Differences consisted mainly of substitutions and the transition/transversion ratio as well as the number of polymorphic sites were found to be higher in the short spacer class than in the long (2.33 and 40 vs 0.88 and 17 in tree 107, respectively).

A diagrammatic representation of the two spacer classes is shown in **Figure 42A**. Small alignment gaps 8 bp to 21 bp-long are observed in the middle spacer region of both short ($d_8$, $d_{13}$ and $d_{21}$) and long spacer classes ($d_{15}$ and $i_{17}$). It is noteworthy that both sides of 4 out of these 5 alignment gaps ($d_8$, $d_{13}$, $d_{21}$ and $d_{15}$) are just preceded by the same 3 bp to 8 bp motif (TTAT, GAG, TATTCTTT and GGC, respectively). As described in Scoles *et al.* (1988) such small regions of similarity could occasionally be involved in "slippage" and looping out of parental strand during DNA replication. Therefore, it can be concluded that these 4 alignment gaps resulted from deletion events. Conversely, due to its specific position in a repeated element (see below and **Figures 42A** and **42B**), the 17-bp alignment gap in the long spacer class ($i_{17}$) is most readily accounted for by an insertion.

The highly divergent status of the short and long spacer classes was especially borne out by the low global similarity scores observed (31% on average, **Table 32**). In fact, we were unable to find

significant sequence identity (> 50%) in the middle spacer regions and as can be seen in **Figure 42A**, alignment between the two classes was reduced to the 5' (89 bp with 85% mean sequence identity) and 3' spacer regions (47 bp with 81% mean sequence identity). Length differences between the two classes partly originated in the duplication events that occurred in the long spacer class. One large duplication (48 bp) was observed near the 5' end (α and β elements) and five tandemly arranged repeated elements, approximately 50 bp-long, were identified 80 bp upstream the 3' end (A, B, C, D and E). Sequence comparison clearly demonstrated that the repeated elements are "old" enough to have accumulated mutations such as an insertion (in the A element) and numerous base-pair changes, particularly in the A-E elements near their 3' end (**Figure 42B**).

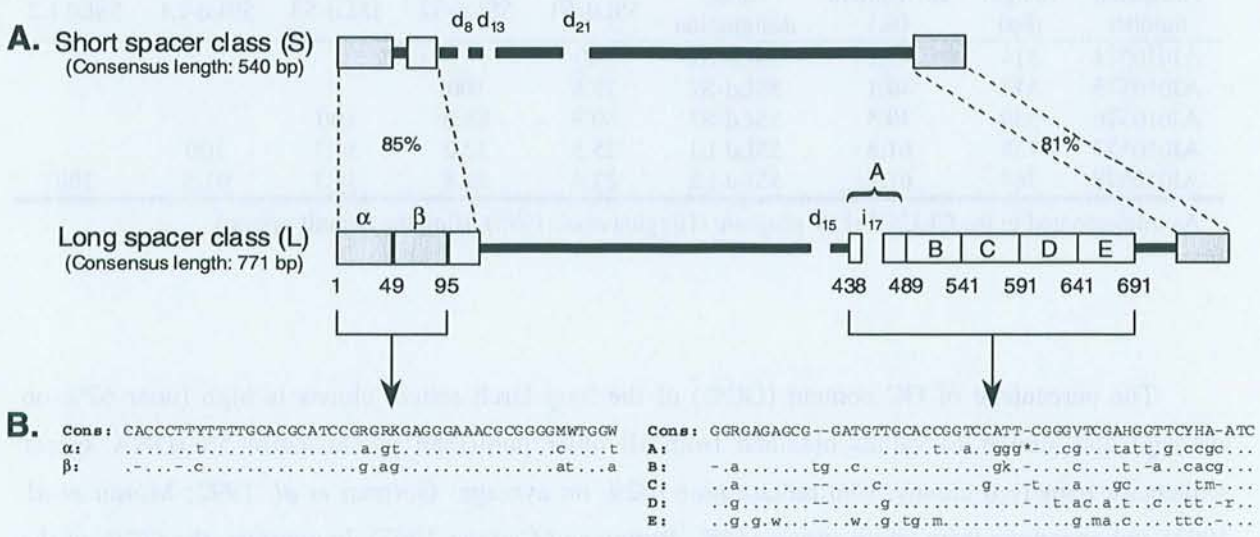



**Figure 42.** Structure of the short and long spacer classes in *Larix decidua*.
***A.*** Diagrammatic representation. Small alignment gaps within each class are represented as interruptions in the diagrams and their origin (i.e. insertion or deletion) indicated by the initials followed by the length in bp (e.g. $d_8$, $i_{17}$). The small regions of sequence identity (%) between the two spacer classes are shown as patterned boxes. The tandemly duplicated elements identified in the long 5S rDNA spacer class are shown as labeled boxes (α/β and A/B/C/D/E) and their relative position in the consensus diagram indicated (in bp).
***B.*** Sequence comparisons of the tandemly duplicated elements identified in the long spacer class. Only those bases differing from the majority rule consensus sequence (Cons) are shown. Polymorphisms are indicated by standard IUPAC symbols and the 17 bp insertion located in the A element was replaced by an asterisk for clarity.

### χ- Sequence comparison with higher plant 5S rDNA spacers

The spacer sequences of *Larix decidua* (5 clones) and higher plants (~ 275 accessions) were compared. Only the spacers from the two conifers *Pinus radiata* (6 clones; Gorman *et al.* 1992; Moran *et al.* 1992) and *Picea glauca* (1 clone; Brown and Carlson 1997) produced high-scoring

segment pairs. Further investigation of these 12 coniferous spacer clones revealed that sequence identity is restricted to 3 small sections located immediately downstream and upstream the 5S gene (5' and 3' spacer regions, respectively) and in the middle spacer region.

The mean similarity score observed in the 5' spacer region (40 bp) was only 47% whereas in the 3' spacer region (60 bp) it was 58%, reaching 79% for the last 34 bp. As all the sequences were unambiguously aligned with a few gaps (**Figure 43A**), we used these two spacer regions as a whole sequence data (100 bp) to investigate the relationship between the European larch, radiata pine and white spruce spacers. The unrooted phenetic tree obtained is shown in **Figure 43B**. On the basis of these two 5S gene flanking regions, clustering of the spacer sequences from a given species is evident. As expected, the short and long larch spacers clearly appeared as two independent groups which have different rate of evolution (see the branch lengths).

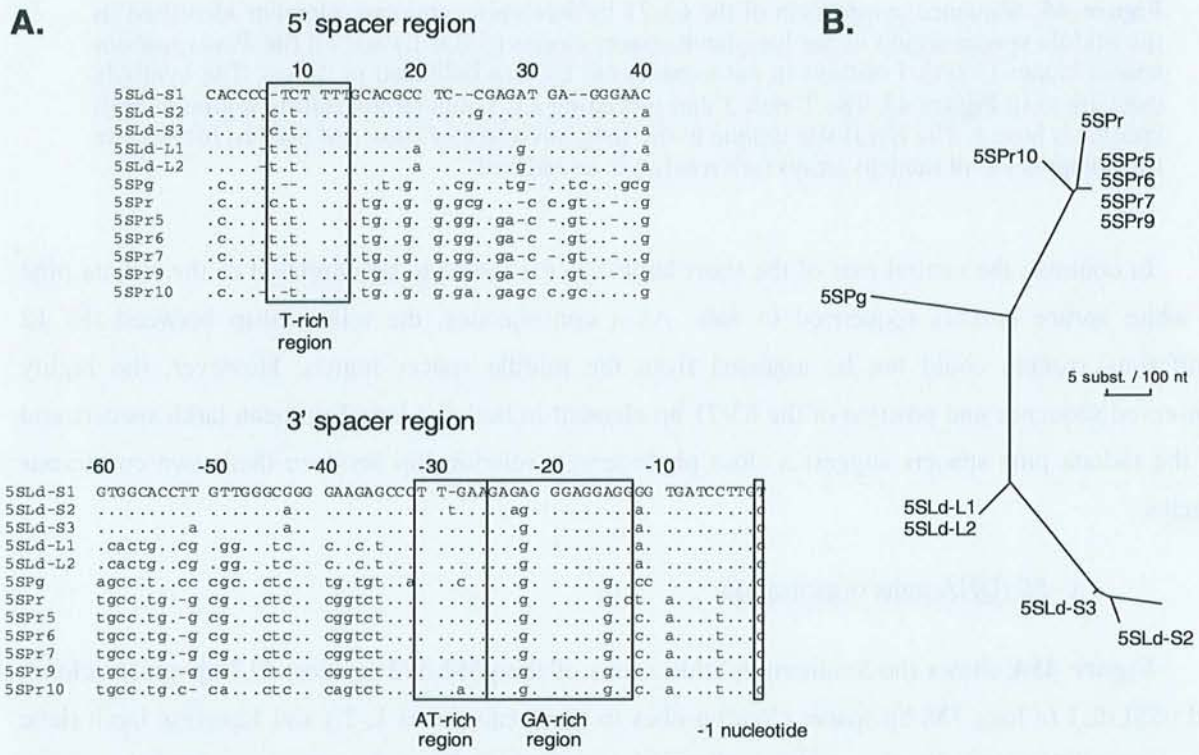



**Figure 43.** Comparison of the 5' (40 bp) and 3' (60 bp) spacer regions of *Larix decidua* (5SLd), *Picea glauca* (5SPg) and *Pinus radiata* (5SPr).

***A.*** Sequence analysis. The *Picea glauca* isolate was taken from Brown and Carlson (1997) and the *Pinus radiata* isolates from Gorman *et al.* (1992) and Moran *et al.* (1992). The nucleotides are numbered relatively to the 5S gene region. Nucleotide substitutions are indicated by lowercase letters. Periods represent residues identical to those of the reference sequence from *Larix decidua* and dashes indicate alignment gaps. The putative regulatory elements identified in plants and referred to in the Discussion are boxed.

***B.*** Neighbor-Joining tree generated from the 5' and 3' spacer regions defined in *A*. The two spacer regions were assembled and the resultant sequence data was used to produce the unrooted tree.

Concerning the middle spacer region, a short 63-71 bp element was found to be common to the long larch spacer clones and both the short and long spacer clones from radiata pine (**Figure 44**). This larch/pine common element was 63% conserved on average (more than 85% for the last 41 bp) and characterized by a T-rich 3' end. It is noteworthy that immediately downstream this element we found a CT-rich region in both species (data not shown), 46 bp long in larch (91% CT-rich) and 54-78 bp long in radiata pine (87-94% CT-rich). This microsatellite region was the main source of size differences in the short radiata pine spacers (Moran *et al.* 1992).

```
5SLd-L1   281 CCGGGAATCC CTCGACGAG- ---ATCT--- ACGCAGCGCC CTCGGTGGCC GCGGACCTTA CGCCTTTTTT T 344
5SLd-L2   277 .......... ..g....... .......... .......... .......... .......... ........... . 340
5SPr      122 .ac....agg .cgcg...cg c.ac..caaa .......... g......... .....a.... ........... c 192
5SPr5     122 .ac.-..agg .cgcg..g.. ...c..c... ...g...... g......... .....a.... ...a.....c . 184
5SPr6     123 .ac....agg .cgcg..g.. ...c..c... ...g...... g......... .....a.... ...a.....c . 186
5SPr7     123 .ac....agg .cgcg..g.. ...c..c... ...g...... g......... .....a.... ...a.....c . 186
5SPr9     119 .ac....agg .cgcg..... ...c..c... .......... g......... .....a.... ....      c . 182
5SPr10    451 .ac....agg .ag.g..... ...c..c... .......... g.......t. .....a.... ....      c 514
```

**Figure 44.** Sequence comparison of the 63-71 bp larch/pine common element identified in the middle spacer region of the long larch spacer clones (5SLd-L) and all the *Pinus radiata* spacer clones (5SPr). Positions in each spacer (in bp) are indicated in italics. The symbols used are as in **Figure 43**. The T-rich 3'end preceding a CT-rich microsatellite region in both species is boxed. The *Bgl* II site unique to the long larch spacer class and used to investigate the composition of tandem arrays (see results) is underlined.

In contrast, the central part of the short larch spacers seems to be unrelated to the radiata pine or white spruce spacers sequenced to date. As a consequence, the relationship between the 12 coniferous spacers could not be assessed from the middle spacer region. However, the highly conserved sequence and position of the 63-71 bp element in both the long European larch spacers and all the radiata pine spacers suggest a close phylogenetic relationship between these two coniferous species.

c- 5S rDNA units organization

**Figure 45A** shows the Southern hybridizations of the p5SLdS2 (a short 517 bp spacer clone) and p5SLdL1 (a long 738 bp spacer clone) probes to European (lanes 1, 2,) and Japanese larch (lane 3) genomic DNA partially to completely digested with *Bst*E II. This restriction enzyme has a single recognition site within the 5S rDNA coding sequence in larch (**Figure 39**).

The intense digest products observed in lanes 2 and 3 confirmed that two highly divergent short (S) and long (L) 5S rDNA unit classes of about 650 bp and 870 bp occur in both species. As expected from the sequence data, the p5SLdS2 and p5SLdL1 probes were specific to the short and long units, respectively, with very low cross-hybridization.

Partial digestions of European larch genomic DNA (lane 1) clearly generated a ladder of hybridizing fragments (ladders A and B as revealed by p5SLdS2 and p5SLdL1, respectively) characteristic of DNA sequences organized in tandem array(s). Remarkably, the two ladders are composed of rather identical bands (except the 1740 bp and 2610 bp fragments in ladder B) which are

not multimers of either short or long units. This observation suggests that both short and long units occur in each of these ladder fragments.

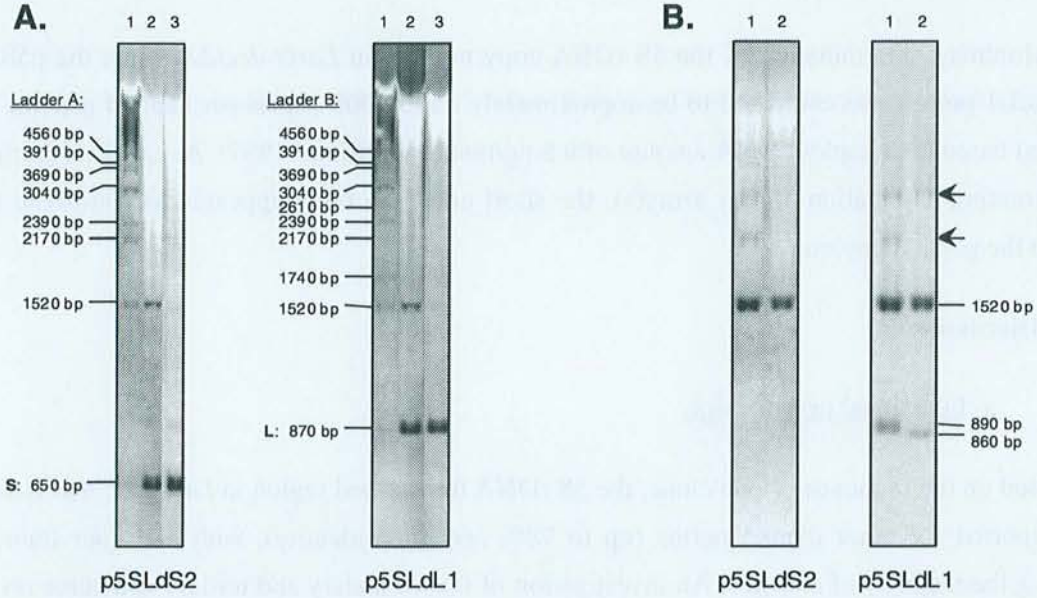



**Figure 45.** Organization of the larch 5S rDNA units.
***A.*** Genomic DNA (2 g) from *Larix decidua* (lanes 1,2) or *Larix kaempferi* (lane 3) digested with 0.5 U (lane 1), 5 U (lane 2) and 10 U (lane 3) of *Bst*E II enzyme and hybridized with the p5SLdS2 or p5SLdL1 probes.
***B.*** Genomic DNA (2 g) from *Larix decidua* (lane 1) or *Larix kaempferi* (lane 2) digested with 10 U of *Bgl* II enzyme and hybridized with the p5SLdS2 or p5SLdL1 probes. Two bands of approximately 2200 bp and 3000 bp and indicated by the arrows could either represent partial digests or DNA flanking the main 5S array.

A more precise investigation of the relative organization of the two 5S rDNA unit classes within array(s) was carried out using the *Bgl* II enzyme since a unique restriction site was only found in the long spacer class of both European and Japanese larches (in the 63-71 bp larch/pine common element, see **Figure 44**). Genomic DNA was digested to completion and subsequently hybridized with p5SLdS2 or p5SLdL1. A prominent 1520 bp fragment was revealed with both probes in both species (**Figure 45B**) demonstrating its composite structure (short + long units). This result, as well as the occurrence of a 1450 bp dimer fragment after amplification with P1/P2 (dSL, see **Figure 41**) and the fact that the size of most of the ladder fragments corresponded to multimeric combinations of short and long units (**Figure 45A**), are all consistent with a 5S rDNA organization in composite tandem array(s) with extensive zones of alternating short and long units.

However, the faint hybridization of the p5SLdL1 probe to a 890 bp fragment in *Larix decidua* (lane 1) and 860 bp fragment in *Larix kaempferi* (lane 2) clearly demonstrated the additional occurrence of small clusters of tandemly arranged long units. This result was also borne out by the 1740 bp and 2610 bp fragments observed in ladder B which were interpreted as multimers of the

870 bp unit class (**Figure 45A**, lane 1). In contrast, we failed to detect any hybridization signal in the high molecular weight zone with the p5SLdS2 probe, revealing the absence in both species of tandemly arranged short units.

Preliminary determination of the 5S rDNA copy number in *Larix decidua* using the p5SLdS2 and p5SLdL1 probes was estimated to be approximately 1500-3000 copies per haploid genome (data not shown) based on a haploid DNA amount of 9.8 pg/nucleus (Dhillon 1987). As expected from their main alternating association within array(s), the short and long units appeared to contribute about equally to the global copy number.

**5- Discussion**

a- Functional organization

Based on the sequence of one clone, the 5S rDNA transcribed region in *Larix decidua* is similar to that reported for other gymnosperms (up to 98% sequence identity), with base pair transitions comprising the majority of changes. An investigation of the secondary and tertiary structures revealed that the larch 5S gene sequenced is likely to represent a functional gene.

The spacer regions immediately upstream and downstream of the 5S gene were found to be well conserved in *Larix decidua* and two other gymnosperm species belonging to the genera *Pinus* and *Picea*. Sequence elements in these two regions were shown to be involved in transcriptional regulation of 5S genes in *Xenopus*, several fungi (Korn 1982; Selker *et al.* 1986; Challice and Segall 1989) and, by analogy, in angiosperm plants as well (Hemleben and Werts 1988; Venkateswarlu *et al.* 1991). Likewise, we revealed similar putative regulatory elements in gymnosperms (shown in **Figure 43A**). The AT-rich region more probably related in function as an initial site for DNA unwinding was identified as TT(T/-)(G/C/A)AA from nucleotides -26 to -31 upstream the 5S gene. As mentioned by Brown and Carlson (1997), the strict AT composition reported among angiosperms for this AT-rich region has been rarely observed in gymnosperms. The GC-rich region potentially involved in transcriptional regulation seems to be a GA-rich region in gymnosperms (nucleotides -13 to -25) with a highly conserved hexanucleotide core sequence GGGAGG from nucleotides -16 to -21. Eleven of the 12 sequences have a C nucleotide at position -1, that is very likely to play a role in the initiation of transcription. Finally, the well-known T-rich region, possibly functioning as an RNA polymerase III termination signal, was located 7 to 13 bp downstream of the 5S gene.

b- The two highly divergent unit size classes in larch

Two major size classes of 5S rDNA units, approximately 650 and 870 bp long were detected in *Larix decidua* and *Larix kaempferi*. This situation seems to be a common feature in the *Pinaceae,* since several size classes were also revealed in the larch species *Larix x eurolepis*, *Larix russica*, *Larix laricina* and *Larix occidentalis* (Trontin and Favre, unpublished results, voir **Partie II-B** de ce

**Chapitre IV**) as well as in *Pinus radiata* (525 and 850 bp, Moran *et al.* 1992) and *Picea glauca* (221 and 600 bp, Brown and Carlson 1997). Multiple 5S rDNA unit size classes have also been reported for many flowering plants including flax (Goldsbrough *et al.* 1981), pea (Ellis *et al.* 1988) and members of the *Triticeae* (Baum and Appels 1992).

Insertions-deletions (indels) and duplications located in the spacer region have been largely invoked as the main source of 5S rDNA units size variation (Scoles *et al.* 1988; Sastri *et al.* 1992). In *Pinus radiata* for example, the difference between the short and long units is mainly due to one 330 bp insertion in the middle spacer region (Moran *et al.* 1992). In *Triticum aestivum* (Cox *et al.* 1992) and *Hordeum vulgare* (Kanazin *et al.* 1993) the shorter spacers may have derived from the longer ones by deletions (27-46 bp long and 130 bp long, respectively). In *Australopyrum pectinatum* (Kellogg and Appels 1995) and *Pisum sativum* (Ellis *et al.* 1988) some long spacers result from one large duplication in the short spacers (110 bp and 50 bp, respectively). Similarly, in larch, the large duplications observed in the long spacers, particularly in the middle spacer region, account for most of the length difference between the short and long spacer classes. However and unlike other plant species, which also have multiple 5S rDNA repeat length classes, the comparison between the two spacer classes revealed that sequence identity is dramatically low in the middle spacer region. As a consequence, it can be concluded that neither the long nor the short units derived simply from each other (i.e. as a result of a few simple events such as indels). Interestingly, sequence comparison between the larch and other gymnosperm spacers revealed that one 63-71 bp middle spacer region is highly conserved in the long larch spacers and all the radiata pine spacers (short and long). The most parsimonious interpretation for the occurrence of this shared element is that it reflects the existence of an ancestral sequence in an hypothetical common ancestor for larch and pine. As this region was not found in the short larch spacers, this result is consistent with, but does not prove, that there are two main independent evolutionary 5S rDNA lineages in larch. In this hypothesis, the numerous base pair changes observed in the tandemly duplicated sequences of the long spacers suggest that they could be the more ancient. This larch/pine common element was not found in the spacer region of the 221-bp unit class sequenced in *Picea glauca* (Brown and Carlson 1997). However, in these units, the spacer region is particularly short (101 bp) compared with those of most other plants and almost entirely composed of the 5' and 3' flanking regions involved in transcriptional regulation (see **Figure 43A**). This clearly suggests that the middle spacer region was eliminated in these units. In this context, it will be of interest to test the 600-bp long unit class identified in *Picea glauca* for the presence/absence of this region.

c- The composite tandem array

As in most other tandemly repeated multigene families subjected to intralocus concerted evolution of repeats (Arnheim 1983), our results also revealed 5S rDNA array homogeneity, but the elementary repeat which is homogenized in both European and Japanese larch consists of an unusual 1520 bp dimeric association of the two highly divergent short (ca. 650 bp) and long units (ca. 870 bp).

This results in composite tandem array(s) with extensive zones of strict alternation of the two unit size classes.

Comparable situations have only been previously reported in pea (Ellis *et al.* 1988) and flax (Schneeberger and Cullis 1992). In pea, two unit size classes (279 bp and 329 bp) differing by a 50 bp duplication in the spacer region were found to be largely interspersed with each other, but without strict alternation as in larch. In flax, the 5S rDNA arrays are dispersed over many chromosomes and the majority of them have a typical repeat unit (5S gene + spacer) of 341 bp to 344 bp. However, one small cluster approximately composed of 500 copies/C and named *Flp-1* is characterized by an unusual arrangement of dimer repeats 706 bp long. As a difference with larch, the two alternating units (5S1 and 5S2) are moderately divergent, showing the same length (353 bp) and only 23% sequence variation.

In larch, two scenarios can be proposed to account for the establishment of dimer repeats within 5S array(s), depending on whether or not the two alternating unit size classes represent independent evolutionary lineages.

In the first scenario, the interspersion pattern of short and long units may result from 5S rDNA transfer between ancestral independent arrays of short and long units. The occurrence in both species of only one (or several) small cluster(s) of tandemly arranged long units in addition to the main composite tandem array(s) suggests that the long units could have been transposed to an array of short units. Such recombination of 5S rRNA genes to other loci has not yet been demonstrated in plants. However, in *Gossypium*, Wendel *et al.* (1995) showed that the 18S-5.8S-26S rDNA repeats evolve under strong interlocus concerted evolution. Most models proposed to explain such interlocus interactions invoke mechanisms such as unequal crossing-over and gene conversion when arrays are on homologous chromosome positions (Ohta and Dover 1983; Nagylaki 1984, 1990). When barriers prevent such direct exchanges (arrays distributed at different chromosomal locations), it has been suggested that functional copies of the 5S rRNA genes can be transposed through covalently closed circular DNA (cccDNA) or RNA intermediates (Drouin and Moniz de Sá 1995).

The alternative scenario (i.e. two dependent length classes) supposes that variant units would have diverged within homogeneous individual array despite intralocus concerted evolution. Such a phenomenon has been described in several species, but restricted to the ends of the 5S arrays (Smith 1976; Mc Mahon *et al.* 1984; Tyler-Smith and Brown 1987; Ellis *et al.* 1988; Linares *et al.* 1994). In these terminal regions the degree of possible mis-alignment required for unequal crossing-over is indeed more limited than in the centre. In such a scenario, the fact that the long spacer class is necessarily the more ancient one supports the hypothesis that the short units are derived from the ends of an ancestral array of tandemly arranged long units.

The fact that only one 5S rDNA locus has been detected in *Larix decidua* by means of FISH (chromosome pair III; Lubaretz *et al.* 1996) rather supports the second scenario. However, these

authors used a heterologous 5S rDNA probe amplified from genomic DNA of the angiosperm *Vicia faba*. FISH experiments using the p5SLdS2 and p5SLdL1 homologous larch probes would be more sensitive to locate 5S rDNA loci, especially small cluster of tandemly arranged long units, and thus provide additional insights to further investigate the 5S rDNA unit evolution in the larch genome.

The observation of composite tandem array(s) with alternating short and long units in both European and Japanese larch clearly suggests that the establishment and subsequent homogenization to such a fairly great extent of the 1520 bp dimer repeat probably occurred before divergence of these two species (Eocene time; Le Page and Basinger 1995). The fixation of such a structure may be due to the relief of selective pressure for a functional 5S gene product and the simpler hypothesis is that the dimer repeat represents the functional entity. A more detailed structural and functional analysis of the 1520 bp dimer repeat would be helpful in addressing this unusual situation.

***Acknowledgements:*** Authors thank Drs. L.E. Pâques and M.A. Lelu (INRA Orléans, France) for supplying most of the plant material used in this study.

**6- Conclusion**

a- Rappel des principaux résultats obtenus

**Les espaceurs non transcrits (NTS)** de l'ADNr 5S de 8 mélèzes d'Europe et 9 mélèzes du Japon ont été sélectivement amplifiés par PCR en utilisant des amorces "universelles", c'est-à-dire s'hybridant sur les gènes qui alternent avec les espaceurs. Les profils électrophorétiques obtenus sont de même type pour tous les arbres testés et permettent d'identifier deux classes de taille d'espaceurs, l'une d'environ 530 pb (classe d'espaceurs courts) et l'autre d'environ 750 pb (classe d'espaceurs longs). Trois clones d'espaceurs courts (514-539 pb) et 2 clones d'espaceurs longs (738-767 pb) ont été isolés et séquencés à partir de 2 mélèzes d'Europe de référence (107 et 92-1).

L'homologie de séquence au sein de chaque classe est élevée (81 % pour la classe courte et 91 % pour la classe longue) et les petites variations de longueur observées peuvent être expliquées par des événements du type insertion ou délétion, probablement induits par des phénomènes de glissements de brins au cours de la réplication de l'ADN (présence de petits motifs dupliqués de part et d'autre des insertions et/ou délétions).

En revanche, la différence de longueur entre les deux classes d'espaceurs ne peut être que partiellement expliquée par la présence de motifs répétés en tandem au sein des espaceurs longs (2 familles de 2 et 5 éléments répétés d'environs 50 pb). L'homologie de séquence entre les deux classes est en effet globalement très faible (31 %). Aucune homologie n'a pu être détectée dans la zone centrale des espaceurs et seules les extrémités 5' (89 pb) et 3' (47 pb) bordant le gène sont relativement bien conservées (plus de 80 % de similarité de séquence). Cette importante divergence de séquence entre les classes d'espaceurs courts et longs est également vérifiée par la mesure de divers

paramètres comme la composition en base GC (48 % VS 62 %), le nombre de sites polymorphes (40 VS 17) et le taux de transitions/transversions (2,33 VS 0,88).

Les recherches effectuées dans les banques de données internationales n'ont permis de détecter des similarités de séquence qu'entre les espaceurs de mélèze d'Europe et ceux des deux seules autres gymnospermes disponibles lors de la rédaction de l'article : *Pinus radiata* (Gorman *et al.* 1992, Moran *et al.* 1992) et *Picea glauca* (Brown et Carlson 1997). La similarité apparaît de nouveau limitée aux extrémités 5' (40 pb avec 47 % d'homologie) et 3' (60 pb avec 58 % d'homologie) bordant le gène, probablement en raison de la présence dans ces deux régions d'éléments régulateurs de la transcription communs aux trois espèces (région riche en T de l'extrémité 5' ; régions riches en AT et en GA de l'extrémité 3' ; base C précédant le gène). Ces deux zones d'homologie (une centaine de pb au total) constituent donc une cible potentielle d'étude des relations phylogénétiques entre espèces de gymnospermes.

Une similarité de séquence significative (> 50 %) affectant la région centrale des espaceurs a aussi été détectée, mais uniquement entre les espaceurs longs de *Larix decidua* et l'ensemble des espaceurs séquencés (courts ou longs) de *Pinus radiata*. La présence de cette région conservée suggère qu'il existe chez les mélèzes d'Europe et du Japon deux lignées indépendantes d'espaceurs dont l'une, la classe d'unités longues, est probablement très ancienne. Selon les clones, sa taille varie de 63 à 71 pb. L'homologie globale est de 63 % et elle atteint plus de 85 % au niveau des 40 dernières bases. De manière remarquable, elle est caractérisée par une extrémité 3' riche en T qui précède une région riche en CT.

**La région transcrite (gène 5S)** a été amplifiée grâce à un couple d'amorces spécifiques choisies à partir de la séquence des extrémités 5' et 3' des espaceurs. Aucun variant de longueur n'a pu être détecté parmi les arbres testés. Un clone issu d'un mélèze d'Europe a été isolé et séquencé. La séquence obtenue (120 pb) s'est avérée peu différente de celles publiées chez d'autres gymnospermes (89 à 98 % d'homologie, 11 espèces). Les rares mutations observées sont souvent compensatrices, c'est-à-dire qu'elles n'ont pas d'impact sur la structure secondaire des ARNr 5S produits. En d'autres termes, une mutation à un site (par exemple "A") implique la sélection d'une mutation complémentaire à un autre site (par exemple "G"). Le nombre de sites informatifs et indépendants est donc faible et l'intérêt du gène 5S pour les études phylogénétiques chez les gymnospermes apparaît donc très réduit.

**L'organisation relative des deux classes d'unités (gène + espaceur)**, courtes (environ 650 pb) et longues (environ 870 pb), ainsi détectées dans le génome des mélèzes d'Europe et du Japon a été étudiée par analyse RFLP. Un clone d'espaceur court (517 pb) et un clone d'espaceur long (738 pb) d'un mélèze d'Europe ont été marqués à la digoxigénine et utilisés comme sondes.

| Espèce | Classe d'unités | Structure des unités | Nombre d'unités/2C | Nombre de loci | Organisation |
|---|---|---|---|---|---|
| *Larix decidua*[1] | Courte (634-659 pb) | Gène 5S; $d_8$ $d_{13}$ $d_{21}$ | 1500-3000 | 1 | - 1 série majeure d'unités courtes et longues alternées (série composite) |
| | Longue (858-887 pb) | Gène 5S; α β $d_{15}$ $i_{17}$ | 1500-3000 | | - 1 série mineure d'unités longues |
| *Pinus radiata*[2] | Courte (502-538 pb) | Gène 5S | 7700-15000 | Nombreux loci dispersés sur l'ensemnle des chromosomes | Séries hétérogènes de taille et de composition variables |
| | Longue (851 pb) | Gène 5S; Insertion de 330 pb | 300-3500 | | |
| *Picea glauca*[3] | Courte (221 pb) | Gène 5S | 2340 | 1 | 2 séries homogènes adjacentes ou 1 série hétérogène |
| | Longue (~ 600 pb) | ? | ? | | |
| *Pseudotsuga menziesii*[4] | 871-888 pb | Gène 5S; $d_{17}$ a b | ? | 1 | 1 série hétérogène |

**Figure 46**. Bilan synthétique comparatif des connaissances acquises sur la structure et l'organisation des unités d'ADNr 5S (gène 5S + espaceur) chez les gymnospermes.
Les zones conservées (plus de 50 % de similarité de séquence) entre les classes d'unités au sein d'une espèce ou entre les espèces sont représentées par des rectangles de même motif et les zones non conservées par des lignes épaisses et noires. Les petites insertions ou délétions principalement responsables des variations de longueur au sein de chaque classe sont figurées par des interruptions de lignes et notées par une abréviation (i ou d, respectivement) suivie de leur taille (en pb). Les éléments dupliqués sont étiquetés (α/β et a/b).
Données extraites et adaptées de : [1] Trontin *et al.* (1999), [2] Moran *et al.* (1992) et Gorman *et al.* (1992), [3] Brown et Carlson (1997), [4] Amarasinghe et Carlson (1998).

Région de l'espaceur en aval du gène (3' ds)
Région de l'espaceur en amont du gène (5' us)
Motif conservé de 35 à 71 pb (motif 5SmsCR1)
Région riche en TC de taille variable
100 pb

Les hybridations réalisées sur des digestions partielles ou totales d'ADN génomique des deux espèces à l'aide d'enzymes ne coupant qu'une seule fois dans les gènes (*Bst*E II) ou dans les espaceurs longs (*Bgl* II) ont démontré que la plupart des deux types d'unités (chacune au nombre approximatif de 1500 à 3000 par génome 2C) alternent régulièrement au sein d'une ou plusieurs séries composites. Une telle alternance d'unités très divergentes au sein d'un même locus n'a été décrite que chez *Linum usitatissimum* et *Pisum sativum* et à beaucoup plus petite échelle (alternance partielle et/ou séries de taille réduite). Cette situation originale peut être expliquée soit par des échanges entre deux loci ancestraux d'unités courtes et longues (scénario 1), soit par la divergence d'unités variantes directement au sein des séries (scénario 2). L'évolution concertée aurait ensuite paradoxalement conduit à l'amplification au hasard d'un motif dimérique d'environ 1520 pb composé de deux gènes (2 x 120 pb), d'un espaceur court (~ 530 pb) et d'un espaceur long (~ 750 pb). Les informations disponibles concernant la localisation chromosomique des séries d'ADNr 5S chez le mélèze d'Europe (un seul locus sur la paire de chromosomes III, Lubaretz *et al.* 1996) et l'apparente absence de séries homogènes d'unités courtes dans le génome des deux mélèzes étudiés plaide en faveur du second scénario. Si telle est bien la situation, l'émergence d'unités courtes au sein d'une série d'unités longues, puis l'amplification du motif dimérique chez les mélèzes d'Europe et du Japon datent d'une époque très ancienne, précédant la divergence de ces deux espèces (Eocène, il y a au moins 34 à 52 Ma, Le Page et Basinger 1995) et peut-être même la diversification du genre (voir la **Partie II-B** de ce **Chapitre IV**). Le maintien à grande échelle d'une telle unité répétée dimérique au cours de l'évolution est probablement à mettre en relation avec l'organisation fonctionnelle de cette famille de gène.

b- Analyse comparative des données récemment publiées chez les gymnospermes

Depuis les travaux réalisés chez *Pinus radiata*, *Picea glauca* et *Larix decidua* déjà cités, de nouveaux résultats concernant la caractérisation moléculaire (incluant des données de séquence) et la localisation physique des unités d'ADNr 5S (hybridation *in situ*) chez le sapin de Douglas (*Pseudotsuga menziesii*) ont été publiés (Amarasinghe et Carlson 1998). Un bilan synthétique des connaissances acquises chez ces quatre *Pinaceae* est proposé dans la **Figure 46.**

Les séquences nucléotidiques des 4 clones de gènes 5S isolés chez *Pseudotsuga menziesii* (chacun de 120 pb) sont identiques entre elles et à celles de différents clones obtenus chez *Pinus radiata* (5SPr6, 5SPr7 et 5SPr9) et *Picea glauca*, mais aussi chez *Pinus contorta* et *Ephedra kokanica*. Elles ne se distinguent du clone séquencé de mélèze d'Europe que par deux transitions aux positions 58 et 77 (voir **Tableau 31**).

A la différence des autres espèces, une seule classe d'espaceurs de tailles comprises entre 751 pb et 768 pb a été détectée chez le sapin de Douglas. Au sein de cette classe, l'homologie de séquence est excellente (98,8-100 % entre 4 clones). La plus petite taille de l'un des clones résulte d'une petite délétion de 17 pb localisée 80 pb en aval du gène ($d_{17}$, **Figure 46**). La recherche de similarités de séquence entre les espaceurs de *Pseudotsuga menziesii* et ceux de *Larix decidua*, *Picea*

*glauca* et *Pinus radiata* confirme la bonne conservation des extrémités 5' (46-51 pb) et 3' (59-60 pb), en particulier au niveau des motifs impliqués dans la régulation de la transcription (**Figure 47**). Les similarités moyennes non corrigées entre les 4 espèces pour ces deux régions sont indiquées dans le **Tableau 33**. Côté 5', la similarité est excellente entre *Pseudotsuga mensiezii* et *Larix decidua* (76 % environ), tandis qu'elle est relativement faible dans les autres cas (45,7-51,4 %). Une bonne homologie entre *Pseudotsuga menziesii* et *Larix decidua*, espèces phylogénétiquement proches (voir Gernandt et Liston 1999), est également observée côté 3' (66,7 %) mais, étrangement, cette homologie apparaît meilleure entre *Picea glauca* et *Pinus radiata* (76,3 %). Les autres comparaisons donnent des similarités plus faibles et du même ordre de grandeur (57,6-61,6 %). Si l'on considère que ces 4 gymnospermes ont un ancêtre commun, alors il devient évident que la région 5' évolue plus vite que la région 3' (similarité moyenne globale entre les 4 espèces de 53,7 % et 64,1 %, respectivement), probablement en raison d'une pression de sélection plus importante côté 3', où se trouvent des motifs régulateurs de l'initiation de la transcription. Les 34 pb en amont du gène sont d'ailleurs remarquablement conservées entre ces 4 espèces (**Figure 47**). La bonne homologie entre *Pseudotsuga menziesii* et *Larix decidua* côté 5' d'une part, et entre *Picea glauca* et *Pinus radiata* côté 3' d'autre part, tend donc à montrer que les espaceurs de *Larix decidua* et *Pseudotsuga menziesii* possèdent un ancêtre commun plus récent que l'ancêtre commun aux espaceurs de *Pinus radiata* et *Picea glauca*. Ces deux groupes d'espaceurs dériveraient donc d'un ancêtre commun, probablement encore plus ancien. Ces observations sont tout à fait compatibles avec les phylogenèses moléculaires les plus récentes établies pour les gymnospermes (voir Chaw *et al.* 1997). Cependant, la similarité plus faible, côté 3', entre les espaceurs de *Pseudotsuga menziesii* et *Larix decidua* par rapport aux espaceurs de *Picea glauca* et *Pinus radiata* va apparemment à l'encontre de cette hypothèse évolutive, ce qui pourrait s'expliquer par des différences de vitesses d'évolution de cette région selon les groupes taxinomiques.

**Tableau 33.** Similarité de séquence moyenne non corrigée (%) au niveau des extrémités 5' (sous la diagonale) et 3' (au dessus de la diagonale) des espaceurs de l'ADNr 5S de *Larix decidua* (5SLd), *Picea glauca* (5SPg), *Pinus radiata* (5SPr) et *Pseudotsuga menziesii* (5SPm).

| | **5SLd** | **5SPg** | **5SPr** | **5SPm** |
|---|---|---|---|---|
| **5SLd** | / | 61,6 | 61,0 | 66,7 |
| **5SPg** | 48,7 | / | 76,3 | 57,6 |
| **5SPr** | 50,2 | 45,7 | / | 61,6 |
| **5SPm** | 76,0 | 50,0 | 51,4 | / |

La comparaison des séquences d'espaceur montre aussi que le motif de 63 à 71 pb identifié dans la région centrale des espaceurs longs de *Larix decidua* et dans les espaceurs courts et longs de *Pinus radiata* est également présent au sein des espaceurs de *Pseudotsuga menziesii*, en deux copies séparées d'environ 180 pb (a et b, **Figure 46**). Ces deux copies ont toutefois sans doute été tronquées

lors de l'événement (ou des événements) de duplication qui a conduit à leur formation, car l'homologie avec *Larix decidua* et *Pinus radiata*, bien qu'excellente (plus de 80 % de similarité en moyenne), est réduite aux 37 dernières paires de bases (**Figure 48**). Une autre différence majeure avec *Larix decidua* et *Pinus radiata* est l'absence de région riche en TC immédiatement en aval du motif de 63-71 pb, après la région riche en T (poly-T).

**A.**

```
                 10         20         30         40         50
5SLd-S1     CACCCC-TCT TTTGCACGCC TC--CGAGAT GA--GGGAAC A-TGCGGGGCC -TGG
5SLd-S2     ......c.t. .......... ..--g..... ..--......a .-c.......t -c..
5SLd-S3     ......c.t. .......... ..--...... ..--......a .-c.......t -c..
5SLd-L1     .....-t.t. .......a.. ..--....g. ..--......- .-c.......a -...
5SLd-L2     .....-t.t. .......a.. ..--....g. ..--......- .-c.......t -...
5SPg        .c....---. ...t.g.... ..cg...tg- ..tc...gcg c-.tttaa... -ctc
5SPr        .c....c.t. ....tg..g. g.gcg...-c c.gt..-..g .-..ac.c.ag -ga.
5SPr5,6,7   .c....t.t. ....tg..g. g.gg..ga-c -.gt..-..g g-..acac.ag -ga.
5SPr9       .c....-.t. ....tg..g. g.gg..ga-c ..gt..-..g .-..ac.c.ag -ga.
5SPr10      .c...--.t. ....tg..g. g.ga..gagc c.gc.....g .-..ac.c.gg gga.
5SPm4g,3e   .g.a..-.t. .....t.... ..--..g.gc ..--......a .a....c...- -...
5SPm2e,4f   .g.a..-.t. .....t.... ..--..g.gc ..--......a .a........- -...
```

Région riche en T

**B.**

```
                -60        -50        -40        -30        -20        -10
5SLd-S1         GTGGCACCTT GTTGGGCGGG GAAGAGCCCT T-GAAGAGAG GGAGGAGGGG TGATCCTTGT
5SLd-S2         .......... .......a.. .......... .t....ag.. ........a. .........c
5SLd-S3         .........a .......a.. .......... .-....g... ........a. .........c
5SLd-L1,L2      .cactg..cg .gg...tc.. c..c.t.... .-....g... ........a. .........c
5SPg            agcc.t..cc cgc..ctc.. tg.tgt..a. .-c...g... .....g..cc .........c
5SPr            tgcc.tg.-g cg...ctc.. cggtct.... .-....g... .....g.ct. a....t...c
5SPr5,6,7,9     tgcc.tg.-g cg...ctc.. cggtct.... .-....g... .....g..c. a....t...c
5SPr10          tgcc.tg.c- ca...ctc.. c.gtct.... .-a...g... .....g..c. a....t...c
5SPm2e,3e,4f,4g .gtctc.tcg c...ct.tcc ccgc.c.... .-....g... ........a. .........c
```

Région riche en AT — Région riche en GA — Base -1

**Figure 47**. Comparaison des extrémités 5' (**A**) et 3' (**B**) des espaceurs de l'ADNr 5S isolés chez *Larix decidua*
(5 clones "5SLd", Trontin *et al.* 1999), *Picea glauca* (1 clone "5SPg", Brown et Carlson 1997), *Pinus radiata* (6 clones "5SPr", Gorman *et al.* 1992, Moran *et al.* 1992) et *Pseudotsuga menziesii* (4 clones "5SPm", Amarasinghe et Carlson 1998).
Les positions nucléotidiques sont numérotées par rapport à la région transcrite (gène 5S) située immédiatement en amont (**A**) ou en aval (**B**). Les substitutions figurent en lettres minuscules. Les points représentent les bases identiques à celles de la séquence de référence 5SLd-S1 et les tirets, les bases manquantes. Les régions probablement impliquées dans la régulation de la transcription (régions riches en T, en AT, en GA et base -1) sont encadrées.

Cette conservation d'une région centrale des espaceurs des unités d'ADNr 5S chez les *Pinaceae* nous a conduit à essayer d'approfondir la connaissance de ce motif (**Partie II-B** suivante de ce **Chapitre IV**) que nous nommerons désormais 5SmsCR1 pour "5S rDNA middle spacer Conserved Region N°1".

```
                                        10         20         30         40         50         60         70
5SLd-L1                  281  CCGGGAATCC CTCGACGAG- ---ATCT--- ACGCAGCGCC CTCGGTGGCC GCGGACCTTA CGCCTTTTTT T
5SLd-L2                  277  .......... ..g......- ---....--- .......... .......... .......... ..........  .
5SPr                     122  .ac....agg .cgcg...cg c-ac..caaa .......... g......... .....a.... ..........  c
5SPr5                    122  .ac.-..agg .cgcg..g.- ---c..c--- ...g...... g......... .....a.... ...a.....c  .
5SPr6,7                  123  .ac....agg .cgcg..g.- ---c..c--- ...g...... g......... .....a.... ...a.....c  .
5SPr9                    119  .ac....agg .cgcg....- ---c..c--- .......... g......... .....a.... ....     c  .
5SPr10                   451  .ac....agg .ag.g....- ---c..c--- .......... g.......t. .....a.... ....     c  c
5SPm4g,3e,4f,4g (a)      160  .ttctccga. gcga.aagag tgycgtcggc ga.accta.. .ag.....g. .....a.... ....   ...g..  .
5SPm4g,3e,4f,4g (b)      339  ggt....gga taaccgttcg agcgagrcaa cgcgg..c.. .......... .....a.... ....     c  .
```

**Figure 48.** Comparaison nucléotidique du motif 5SmsCR1 présent dans la région centrale de deux clones d'espaceurs longs de *Larix decidua* (5SLd-L, Trontin *et al.* 1999), 6 clones d'espaceurs de *Pinus radiata* (5SPr, Moran *et al.* 1992; Gorman *et al.* 1992) et 4 clones d'espaceurs de *Pseudotsuga menziesii* (5SPm, Amarasinghe et Carlson 1998). Ces derniers renferment chacun deux motifs 5SmsCR1 (**a** et **b**). La position au sein de l'espaceur de la première base du motif 5SmsCR1 est donnée en pb (en italique). Seules les bases qui diffèrent de la séquence 5SLd-L1 sont indiquées en minuscules. Les bases identiques sont indiquées par des points et les bases manquantes par des tirets. Pour plus de lisibilité, les polymorphismes entre les 4 clones 5SPm sont indiqués par les symbols standards. La séquence riche en T (poly-T) caractéristique de l'extrémité 5' du motif est encadrée.

## B- Etude préliminaire d'un motif conservé (5SmsCR1) au sein de la région centrale des espaceurs de l'ADNr 5S des mélèzes et d'autres gymnospermes

Jean-François Trontin, Damien Chanteranne et Jean-Michel Favre

### 1- Introduction

La région centrale de certaines classes d'espaceurs de l'ADNr 5S de *Larix decidua* (unités longues), *Pinus radiata* (unités courtes et longues) et *Pseudotsuga menziesii*, comporte un motif conservé de 37 à 71 pb (5SmsCR1). La présence de ce motif, apparemment absent chez les angiospermes (consulter la 5S rRNA data bank, Szymanski *et al.* 1998), est remarquable car cette région centrale des espaceurs est par ailleurs très divergente (**Figure 46**). Il pourrait donc constituer une cible intéressante pour l'étude des relations phylogénétiques chez les *Pinaceae* et peut-être plus largement chez les gymnospermes dont la taxinomie peut varier selon les auteurs (voir Chaw *et al.* 1997 pour une discussion).

Après une brève recherche de sa fonction éventuelle, nous avons cherché à développer un outil de détection rapide et spécifique du motif 5SmsCR1 qui permettrait ensuite d'évaluer sa distribution et sa conservation chez les gymnospermes. Ce travail a été réalisé en grande partie par Damien Chanteranne (1999) au cours d'un stage de Maîtrise (2 mois) sous ma responsabilité. Afin d'éviter le séquençage systématique, nous avons choisi une approche par PCR basée sur l'utilisation de couples d'amorces se fixant d'une part sur le gène 5S (très conservé entre les espèces) et d'autre part sur l'élément 5SmsCR1 lui-même. Dans un premier temps, l'outil a été optimisé à partir de 6 espèces de mélèze qui constituent donc nos espèces de référence puis, dans un deuxième temps, nous avons élargi notre recherche à 4 espèces de pin et 3 espèces d'épicéa pour lesquels nous disposions déjà d'informations sur la présence (*Pinus radiata*, Moran *et al.* 1992) ou l'absence du motif (*Picea glauca*, Brown et Carlson 1997). Nous avons enfin inclu pour comparaison un autre membre des *Pinaceae* (le sapin *Abies alba*) et également *Ginkgo biloba* (*Ginkgoacea*), l'un des plus anciens représentant actuel des gymnospermes (Chaw *et al.* 1997).

Parallèlement aux informations apportées concernant le motif 5SmsCR1, les données recueillies nous ont aussi permis d'approfondir notre connaissance de la structure et de l'organisation des unités d'ADNr 5S chez ces gymnospermes.

## 2- Matériel et méthodes

### a- Matériel végétal et extraction de l'ADN génomique

Les espèces testées, d'origine eurasienne ou américaine, sont indiquées dans le **Tableau 34**. Les prélèvements (aiguilles de l'année) ont été effectués sur des arbres adultes de l'Arboretum d'Amance (INRA de Nancy, Champenoux) ou localisés sur le campus de l'Université Henri Poincaré (Vandoeuvre-lès-Nancy). Les échantillons ont été congelés dans l'azote liquide et conservés à - 80°C. Généralement, deux individus par espèce ont été analysés sauf dans le cas de *Larix gmelinii*, pour lequel un seul individu a pu être prélevé. La faible variation intraspécifique (entre les individus) de la taille des espaceurs de l'ADNr 5S observée non seulement chez *Larix decidua* et *Larix kaempferi* (**Partie II-A** de ce **Chapitre IV**) mais aussi chez *Pinus radiata* (Moran *et al.* 1992) justifie ce choix qui privilégie largement le nombre d'espèces par rapport au nombre d'individus.

L'ADN génomique a été extrait et purifié à partir de 100 mg d'aiguilles à l'aide du kit commercial "DNeasy plant mini kit" (Qiagen) en suivant les recommandations du fabricant.

**Tableau 34**. Origine géographique et liste des espèces testées (abréviations entre parenthèses).

| Origine | *Larix* | *Pinus* | *Picea* | *Abies* | *Gingko* |
|---|---|---|---|---|---|
| Eurasie | *L. decidua* (Ld)<br>*L. kaempferi* (Lk)<br>*L. russica* (Lr)<br>*L. gmelinii* (Lg) | *P. sylvestris* (Ps)<br>*P. nigra* (Pn)<br>*P. pinaster* (Pp) | *P. abies* (Pa) | *A. alba* (Aa) | *G. biloba* (Gb) |
| Amérique | *L. occidentalis* (Lo)<br>*L. laricina* (Ll) | *P. radiata* (Pr) | *P. glauca* (Pg)<br>*P. mariana* (Pm) | | |

### b- Amplification

#### α- Sélection des amorces

La détection du motif 5SmsCR1 a été réalisée à l'aide de deux couples d'amorces, chacun constitué d'une amorce complémentaire d'une extrémité du gène (amorces universelles P1 ou P2) et d'une amorce s'hybridant sur le motif 5SmsCR1 (amorces spécifiques M1 ou M2). Une représentation schématique des fragments amplifiés à l'aide de ces couples d'amorces est donnée dans la **Figure 49**. Le couple P1/M1 permet d'amplifier la partie 5' de l'espaceur située entre l'extrémité 3' du gène et le motif 5SmsCR1 (5' sr pour "5' spacer region"). Le couple M2/P2 permet d'amplifier la partie 3' de l'espaceur située entre le motif 5SmsCR1 et l'extrémité 5' du gène (3' sr). Enfin, le couple P1/P2 permet d'amplifier et de caractériser les différentes classes d'espaceurs de chacune des espèces testées.

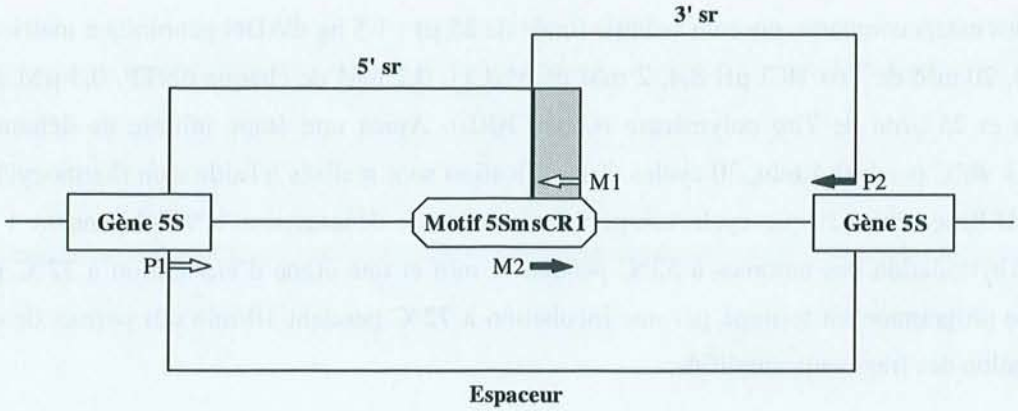



**Figure 49**. Représentation schématique des fragments d'ADNr 5S amplifiés à l'aide des couples d'amorces P1/M1 (5' sr), M2/P2 (3' sr) et P1/P2 (espaceur). La position et l'orientation des amorces sont indiquées par des flèches. La zone grisée (33 pb) représente le recouvrement des fragments 5' sr et 3' sr.

La présence du motif 5SmsCR1 au sein des espaceurs des différentes espèces testées a été considérée comme démontrée lorsque la somme des tailles des fragments amplifiés à l'aide des couples d'amorces P1/M1 (5' sr) et M2/P2 (3' sr) correspondait à la taille des fragments amplifiés par le couple d'amorces P1/P2 (espaceur). La taille des fragments amplifiés est corrigée avant calcul afin de tenir compte d'une part de la taille et de la position des amorces P1 et P2 sur le gène 5S (49 pb à déduire) et, d'autre part, de la zone de recouvrement entre les fragments 5' sr et 3' sr (33 pb à déduire, **Figure 49**).

Les séquences et positions des amorces universelles P1 (Tm = 63 °C) et P2 (Tm = 55°C) sont données dans la **Partie II-A** de ce **Chapitre IV**. Les deux amorces M1 et M2 complémentaires du motif 5SmsCR1 ont été choisies dans sa région 3' très conservée entre *Pinus radiata* et *Larix decidua* (41 pb, voir **Figure 48**). Nous n'avions pas alors connaissance des travaux concernant *Pseudotsuga menziesii* (Amarasinghe et Carlson 1998). Le choix des amorces a été réalisé à l'aide du logiciel Primer 1.2 (Andersen et Bristol, Université de Californie) avec comme exigence un Tm suffisamment élevé pour assurer robustesse et spécificité de l'amplification, l'absence de complémentarité entre amorces et au sein des amorces (structures secondaires) et enfin une distance suffisante entre l'extrémité 3' des amorces (à partir de laquelle la *Taq* polymérase initie la synthèse) et les sites polymorphes observés entre *Larix decidua* et *Pinus radiata* (**Figure 48**). La séquence et la position des amorces M1 (indirecte) et M2 (directe) sélectionnées sont les suivantes :

M1 : 5'-GCGTAAGGTCCGCGGCCACC-3' (20 nt, Tm = 70°C, positions 63-44 de la **Figure 48**)
M2 : 5'-ACGCAGCGCCCTCGGTGGCCG-3' (21 nt, Tm = 74°C, positions 31-51 de la **Figure 48**)

β- Conditions d'amplification

Nous avons cherché à définir un protocole de base susceptible de permettre l'amplification simultanée de l'espaceur et des fragments 5' sr et 3' sr. Le mélange réactionnel retenu après de

multiples essais comporte, pour un volume finale de 25 µl : 1-5 ng d'ADN génomique matrice, 50 mM de KCl, 20 mM de Tris HCl pH 8,4, 2 mM de $MgCl_2$, 0,2 mM de chaque dNTP, 0,3 µM de chaque amorce et 25 U/ml de *Taq* polymérase (Gibco BRL). Après une étape initiale de dénaturation de l'ADN à 96°C pendant 4 min, 30 cycles d'amplification sont réalisés à l'aide d'un thermocycleur PTC-100 (MJ Research). Chaque cycle comporte une étape de dénaturation à 94°C pendant 1 min, une étape d'hybridation des amorces à 53°C pendant 1 min et une étape d'élongation à 72°C pendant 1 min. Le programme est terminé par une incubation à 72°C pendant 10 min qui permet de compléter l'élongation des fragments amplifiés.

Ce protocole, optimisé pour les mélèzes, donne des résultats contrastés pour les autres espèces, probablement en raison d'une hybridation des amorces moins efficace. En effet, le Tm de l'amorce universelle P2 est nettement plus faible (55°C) que celui des amorces P1 (63°C), M1 (70°C) ou M2 (74°C). Le risque de mésappariements avec la séquence cible est donc probablement important. L'utilisation de protocoles d'amplification par "touchdown" au cours desquels la température d'hybridation diminue à chaque cycle de 0,5 à 1°C et qui permettent théoriquement de travailler à des températures d'hybridation plus faibles tout en conservant une bonne spécificité n'ont pas permis d'améliorer significativement les rendements (deux plages de températures testées : 60-50°C et 55-45°C). Le protocole de base a donc été conservé et appliqué à toutes les espèces.

c- Electrophorèse et estimation de la taille des fragments amplifiés

Les échantillons amplifiés sont mélangés avec 10 % (v/v) de tampon de charge 10X [50 % (v/v) de glycérol, 0,4 % (p/v) de bleu de bromophénol, 0,4 % (p/v) de xylène cyanol]. Cinq µl du mélange sont ensuite déposés sur un gel d'agarose à 1,8 % confectionné dans du TBE 0,5X (45 mM Tris-borate, 1 mM EDTA). La migration des fragments est effectuée sous un champ électrique de 5 $V.cm^{-1}$ pendant 6 heures. L'ADN est révélé par immersion du gel dans une solution de bromure d'éthidium à 0,5 µg/ml pendant 30 min. La préparation est enfin photographiée sous ultraviolets (312 nm).

La taille des fragments amplifiés est estimée à partir d'une courbe étalon reportant le logarithme de la taille (en pb) des bandes du marqueur de taille moléculaire "Ladder 100 pb" (Gibco BRL) en fonction de leur distance de migration sur gel d'agarose. Afin d'améliorer l'estimation de la taille des plus grands et des plus petits fragments, nous avons utilisé une régression polynomiale de degré 3 (au lieu d'une régression linéaire). La précision obtenue (généralement à partir de 4 répétitions) après 6 h de migration sur un gel d'agarose à 1,8 % est de l'ordre de 5 %.

d- Recherche de similarités

Les similarités entre le fragment 5SmsCR1 et d'autres séquences nucléotidiques présentes dans les banques de données internationales GenBank, EMBL data library et DDBJ (DNA DataBase of Japan) ont été recherchées à l'aide de l'algorithme de comparaison BLASTn (Altschul *et al.* 1990)

accessible par E-mail (blast@ncbi.nlm.nih.gov) au National Center For Biotechnology Information (NCBI).

### 3- Résultats et discussion

a- Fonction du motif 5SmsCR1

Le **Tableau 35** donne la fonction présumée des 15 séquences offrant les meilleures possibilités d'alignement avec le motif 5SmsCR1 du clone 5SPr9 de *Pinus radiata* choisi comme référence. Comme attendu, les 4 autres clones d'ADNr 5S de *Pinus radiata* affichent les scores d'alignement les plus élevés sur l'ensemble du motif (63-71 pb). Ensuite, plusieurs séquences d'origine très variable (plantes, bactéries, homme, oiseau) présentent des scores plus faibles en raison d'une similarité peu élevée ou n'impliquant qu'une partie seulement du motif. Parmi elles, les séquences d'ARNm codants pour certaines protéines de la sous-unité α du complexe protéasomique 20S des deux angiospermes *Arabidopsis thaliana* (accessions Y13179 et Y13180) et *Lycopersicon esculentum* (accession Y14339) apparaissent particulièrement intéressantes.

**Tableau 35**. Séquences offrant les meilleurs scores d'alignement avec 5SmsCR1 après recherche de similarité "BLAST" dans les banques de données[a].

| Banque | Accession | Description partielle | Score[b] | EO[c] | S (%)[d] |
|---|---|---|---|---|---|
| EMBL | Z33608 | *P. radiata* (5spr9) 5S ribosomal RNA gene | 305 | 1-71 | 100 |
| EMBL | Z33606 | *P. radiata* (5spr6) 5S ribosomal RNA gene | 278 | 1-71 | 95 |
| EMBL | Z33607 | *P. radiata* (5spr7) 5S ribosomal RNA gene | 278 | 1-71 | 95 |
| EMBL | Z33605 | *P. radiata* (5spr5) 5S ribosomal RNA gene | 273 | 1-71 | 93 |
| EMBL | Z33604 | *P. radiata* (5spr10) 5S ribosomal RNA gene | 268 | 1-71 | 92 |
| EMBL | Y13180 | *A. thaliana* mRNA for proteasome subunit Prc6c | 107 | 31-64 | 79 |
| EMBL | Y13179 | *A. thaliana* mRNA for proteasome subunit Prc6b | 107 | 31-64 | 79 |
| DDBJ | D14157 | Rabbit mRNA for brain calcium... | 102 | 49-10 | 78 |
| EMBL | X73308 | *B. caldolyticus* pyrimidine biosynthetase... | 101 | 14-60 | 72 |
| Genbank | U62403 | *Rhodobacter sphaeroides* Nitrite and... | 100 | 7-51 | 73 |
| Genbank | L07300 | Chicken glucose transporter protein... | 100 | 46-2 | 73 |
| Genbank | AF016258 | *Rhodobacter sphaeroides* NnrS... | 99 | 9-51 | 75 |
| Genbank | M31628 | *S. coelicolor* histidine biosynthesis... | 99 | 52-1 | 68 |
| Genbank | M97560 | Human (clone LA171) chromosome 21... | 99 | 66-6 | 64 |
| EMBL | Y14339 | *L. esculentum* mRNA for proteasome, α subunit | 98 | 31-64 | 76 |

[a] Le motif utilisé pour la recherche est issu du clone 5SPr9 de *Pinus radiata*

[b] Score d'alignement HSPs (High scoring segment pairs) donnée par le logiciel BLASTn

[c] Etendue et orientation de la région de similarité. Les positions nucléotidiques sont indiquées selon la numérotation utilisée dans la **Figure 48**.

[d] Pourcentage de similarité pour la région concernée

En effet, l'importante similarité de séquence avec 5SmsCR1 détectée dans ces 3 cas (76-79 %) concerne très précisément son extrémité 3' (environ 36 pb) dont nous avons vu qu'elle était très conservée entre *Larix decidua*, *Pinus radiata* et *Pseudotsuga menziesii* (**Figure 48**). Bien que nettement plus faible, on retrouve également une bonne similarité au niveau des séquences déduites en

acides aminés. Elle atteint par exemple 50 % entre *Pinus radiata* et *Arabidopsis thaliana* ou *Lycopersicon esculentum* (**Figure 50**).Le complexe protéique multicatalytique 20S est composé de deux sous-unités protéasomiques (α et β) impliquées dans la dégradation de diverses protéines régulatrices et de protéines endommagées (Parmentier *et al.* 1997). Il est codé par une super-famille de gènes qui occupe généralement un grand nombre de loci (17 localisés chez *Arabidopsis thaliana*). L'ADNr 5S étant parfois physiquement associé à d'autres familles de gènes répétés comme par exemple l'ADNr, les gènes du TSL (Trans-Spliced leader) ou les gènes des histones (Drouin et Moniz de Sá 1995), les similarités de séquence observées suggèrent que le motif 5SmsCR1 représente peut-être le vestige d'une situation ancienne où l'ADNr 5S se trouvait associé à certains membres de la famille des gènes protéasomiques 20S chez *Larix decidua*, *Pinus radiata* et *Pseudotsuga menziesii*.

Une autre possibilité peut cependant être envisagée. En effet, la présence d'un poly T (signal de terminaison de la transcription par la Pol III) à l'extrémité 3' du motif 5SmsCR1 évoque un gène de la classe III. Or, les gènes de protéines protéasomiques sont transcrits par la Pol II. Le motif 5SmsCR1 pourrait donc correspondre à des petits gènes de la classe III, comme les gènes d'ARNt, d'ARNsn ou d'ARNsc. L'hypothèse selon laquelle il résulterait de la duplication d'une autre région de l'unité d'ADNr 5S, en particulier l'extrémité 3' du gène 5S et la région de l'espaceur immédiatement en aval (région riche en T), peut être exclue (comparer les **Figures 47A** et **48**).

```
A.  Arabidopsis thaliana        gcc gcc gtc ggt gtc cgc ggt acc gac acc gtt gcc
                                 A   A   V   G   V   R   G   T   D   T   V   V
    Lycopersicon esculentum     ... ... ... ... ..a ..t ..a ..t ... ..t ... ...
                                 =   =   =   =   =   =   =   =   =   =   =   =

B.  Pinus radiata               ag. ... ... ... .g. ... ..a ..t t.. g.m t.. t.y
                                 S   =   =   =   G   =   =   =   Y   A   F   F
    Larix decidua               ag. ... c.. ... .g. ... ..a c.t t.. g.. t.. t.t
                                 S   =   L   =   G   =   =   P   Y   A   F   F
    Pseudotsuga menziesii       ssy m.. cws ... .gs ... ..a ..t t.. g.. t.. k.t
                                P/G T/G G/L  =   G   =   =   =   Y   A   F  V/F
```

**Figure 50**. Alignement des séquences nucléotidiques (minuscules) et en acides aminés correspondantes (majuscules) d'une petite région des gènes codant pour certaines protéines de la sous-unité α du complexe protéasomique 20S d'*Arabidopsis thaliana* et *Lycopersicon esculentum* (**A**) avec l'extrémité 3' du motif 5SmsCR1 (36 pb) détecté chez *Pinus radiata*, *Larix decidua* et *Pseudotsuga menziesii* (**B**).
Les séquences nucléotidiques et en acides aminés sont figurées selon les conventions IUPAC standards. Seuls les bases et acides aminés qui diffèrent de la séquence d'*Arabidopsis thaliana* sont indiqués. Un point ou le symbol "=" représentent respectivement une base ou un acide aminé identique à la séquence de référence.

Jusqu'à présent, un seul exemple de présence d'un gène de classe III au sein des espaceurs de l'ADNr 5S a été rapporté chez les plantes. Il s'agit d'un gène inactif d'ARNt accepteur de l'histidine (probablement un rétropseudogène) observé chez le riz (Hariharan *et al.* 1987). La taille du motif 5SmsCR1 (63-71 pb si l'on considère la partie homologue entre *Larix decidua* et *Pinus radiata*) est remarquablement proche de celle des gènes d'ARNt (75-85 pb, Singer et Berg 1992). Il existe d'autre part une similarité de séquence intrigante entre l'extrémité 3' du pseudogène d'ARNt accepteur de

l'histidine du pseudogène d'ARNt présent dans les espaceurs de riz (TTACGCCATTTT) et l'extrémité 3' du motif 5SmsCR1 (TTACGCCtTTTT). Il est même possible de donner au motif 5SmsCR1 une structure secondaire en forme de "feuille de trèfle" très similaire à celle d'un ARNt et présentant les bases typiquement conservées à certaines positions (voir Singer et Berg 1992 et **Figure 51**). La présence dans la boucle anticodon (boucle II) du triplet ACG conduit même à envisager que le motif 5SmsCR1 puisse être apparenté à un gène d'ARNt accepteur de la cystéine. Si c'est bien le cas, ce gène d'ARNt$_{cys}$ serait toutefois probablement inactif (pseudogène) car il apparaît tronqué à son extrémité 5', les nucléotides situés en amont du motif 5SmsCR1 n'étant pas conservés entre *Larix decidua* et *Pinus radiata* et peu complémentaires de l'extrémité 3'. Il devrait de plus comporter en amont du poly T le motif CCA commun et indispensable à la fonction de tous les ARNt (CMT est observé à la place). Enfin, la structure de l'ARN déduit apparaît sensiblement modifiée par rapport à celle d'un ARNt standard, en particulier au niveau des 3 boucles (Singer et Berg 1992).

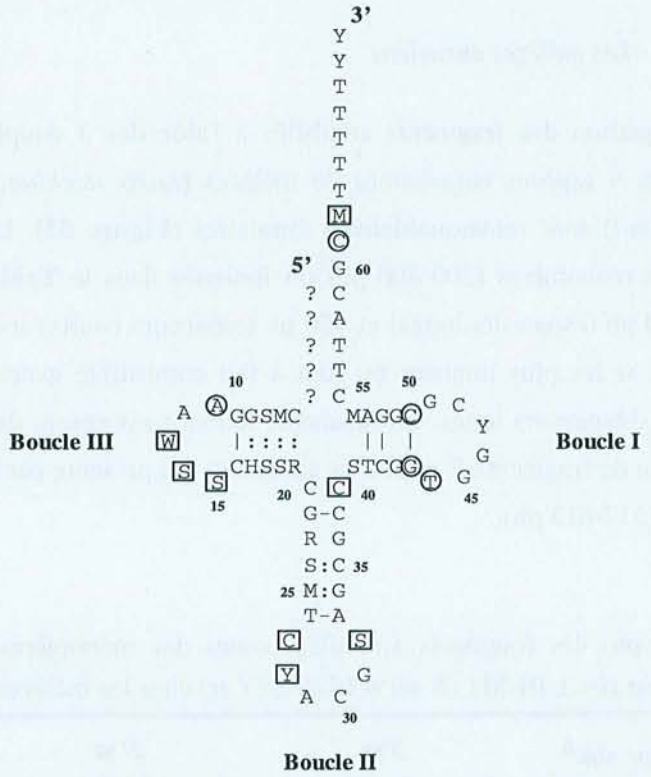



**Figure 51**. Structure secondaire "en feuille de trèfle" proche de celle d'un ARNt déduite des séquences du motif 5SmsCR1 de *Larix decidua* (clones 5SLd-L1 et 5SLd-L2) et *Pinus radiata* (clones 5SPr5, 5Spr6, 5SPr7, 5SPr9 et 5SPr10). Les appariements de bases observés dans les régions tiges pour toutes les séquences ou seulement pour certaines séquences sont respectivement indiqués par un tiret ou par deux points. Les positions variables sont représentées par les symbols IUPAC standards. Les bases encerclées sont invariantes dans tous les ARNt. Les positions nucléotidiques variables encadrées répondent partiellement aux caractéristiques structurales conservées des ARNt. L'extrémité 5' apparaît incomplète (?).

La boucle N°1 (correspondant à la boucle dite "TψC") ne comporte ainsi que 6 nucléotides au lieu de 7 et un motif TGG au lieu de TTC. La boucle N°2 (potentiellement la boucle de l'anticodon) est également plus courte (6 nucléotides au lieu de 7) et on devrait trouver, en position adjacente à l'anticodon présumé ACG, un T côté 5' (T ou C sont observées) et une purine (A ou G) côté 3' (nous observons G ou C). Enfin, la boucle N°3 (dite boucle DU de taille variable selon les ARNt) devrait comporter plusieurs bases G et T en plus de la base A observée en position 11. Malgré ces différences, l'analogie de structure du motif 5SmsCR1 avec les gènes d'ARNt reste remarquable. Une telle association du gène 5S avec un autre gène de la classe III au sein d'une même série est *a priori* plus crédible qu'une association avec un gène de classe II (gène protéasomique).

b- Recherche du motif 5SmsCR1 chez divers gymnospermes

α- Les mélèzes

*- Les mélèzes eurasiens*

Les profils de migration des fragments amplifiés à l'aide des 3 couples d'amorces P1/M1, M2/P2 et P1/P2 chez les 4 espèces eurasiennes de mélèzes (*Larix decidua*, *Larix russica*, *Larix kaempferi* et *Larix gmelinii*) sont remarquablement similaires (**Figure 52**). La taille moyenne des fragments de la zone des monomères (300-800 pb) est indiquée dans le **Tableau 36**. Deux classes d'espaceurs d'environ 740 pb (espaceurs longs) et 550 pb (espaceurs courts) sont observées. La taille des fragments 5' sr et 3' sr les plus intenses est tout à fait compatible avec la présence du motif 5SmsCR1 dans la classe d'espaceurs longs. En revanche, le motif est absent de la classe d'espaceurs courts (pas d'amplification de fragments 5' sr et 3' sr adéquats) qui présente par ailleurs des variations de longueur importantes (517-613 pb).

**Tableau 36**. Taille (en pb) des fragments amplifiés (zones des monomères) à l'aide des couples d'amorces P1/P2 (espaceur obs.), P1/M1 (5' sr) et M2/P2 (3' sr) chez les mélèzes eurasiens[a].

| Classes | Espaceur obs.[b] | | | | 5'sr | | | | 3' sr | | | | Espaceur cal.[c] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Ld | Lr | Lk | Lg | Ld | Lr | Lk | Lg | Ld | Lr | Lk | Lg | Ld | Lr | Lk | Lg |
| **Espaceurs longs** | 748 | 741 | 752 | 731 | 344 | 352 | 358 | 364 | 479 | 475 | 459 | 455 | 741 | 745 | 735 | 737 |
| | | | | | | | 341 | 396 | 498 | 490 | | | | | | |
| | | | | | 394 | 404 | 413 | 425 | 554 | 559 | 546 | 541 | | | | |
| **Espaceurs Courts** | 518 | 517 | 525 | 531 | | | | | | | | | | | | |
| | 539 | 550 | 605 | 613 | | | | | | | | | | | | |

[a] Taille des fragments majeurs, communs aux deux individus testés et reproductibles (amplifiés à 4 reprises). Les abréviations d'espèces utilisées (Ld, Lr, Lk et Lg) sont indiquées dans le **Tableau 34**. Pour une même classe d'espaceur, les fragments sont ordonnés selon une échelle d'intensité décroissante

[b] Taille observée de l'espaceur (corrigée de 49 pb afin de tenir compte des amorces P1 et P2)

[c] Taille d'espaceur calculée à partir des fragments 5' sr et 3' sr les plus intenses (5' sr + 3' sr - 49 pb - 33 pb)

Une certaine hétérogénéité de taille est également révélée dans la classe d'espaceurs longs (présence de fragments supplémentaires 5' sr et 3' sr moins intenses) mais elle est insuffisante pour être clairement détectée lors des amplifications P1/P2. Les fragments intenses de taille comprise entre 1000 et 1200 pb générés par chacun des 3 couples d'amorces (**Figure 52**) ont clairement une nature dimérique (artefact de PCR, **Figure 53**) révélatrice de l'organisation alternée (au moins partiellement) des unités courtes et longues, non seulement chez *Larix decidua* et *Larix kaempferi* (voir **Partie II-A** de ce **Chapitre IV**), mais aussi chez *Larix russica* et *Larix gmelinii* (**Tableau 37**). Dans le cas de *Larix russica* par exemple, la taille du fragment dimérique issu de l'amplification P1/P2 (1449 pb) est approximativement égale à la somme des tailles d'un gène (120 pb), d'un espaceur court (517-550 pb) et d'un espaceur long (741 pb) augmentée de 49 pb (contribution des amorces P1 et P2). La taille du dimère résultant de l'amplification P1/M1 (1016 pb) correspond quant à elle à l'addition d'un espaceur court (517-550 pb), d'un gène (120 pb) et d'un fragment majeur 5' sr (352 pb). Enfin, la taille du dimère observé lors de l'amplification M2/P2 (1138 pb) correspond à la somme des tailles du fragment majeur 3' sr (475 pb), d'un gène (120 pb) et d'un espaceur court (517-550 pb). Les calculs réalisés sous l'hypothèse alternative d'une organisation en séries d'unités courtes ou longues homogènes conduisent à des résultats incompatibles avec les observations.

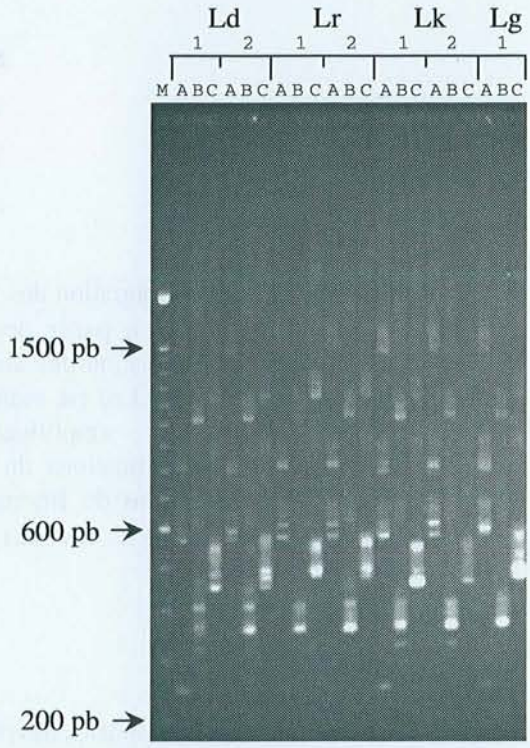



**Figure 52.** Profils de migration des fragments d'ADNr 5S amplifiés à partir des mélèzes eurasiens. La signification des abréviations d'espèces utilisées (Ld, Lr, Lk, Lg) est indiquée dans le **Tableau 34**. A : amplifications des espaceurs ; B : amplifications du fragment 5'sr ; C : amplifications du fragment 3'sr ; M : ladder 100 pb ; 1 et 2 : les deux individus testés par espèce.

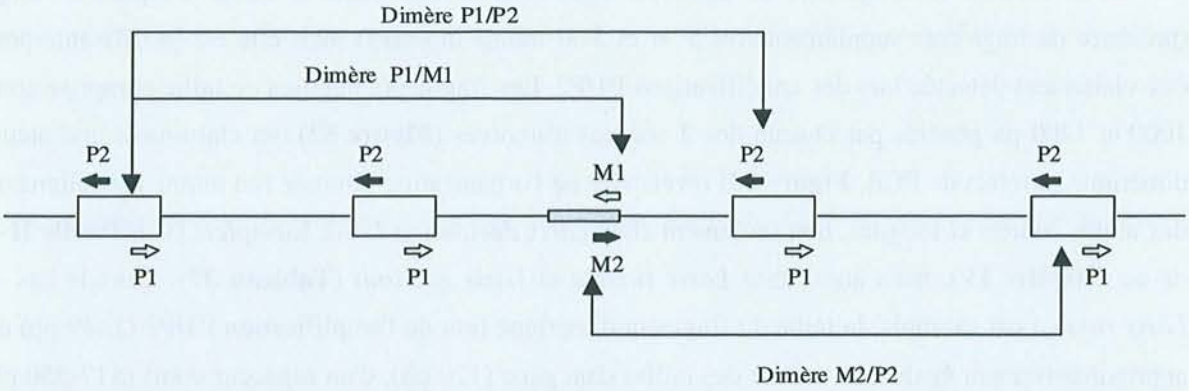



**Figure 53**. Représentation schématique des dimères amplifiés à l'aide des couples d'amorces P1/P2, P1/M1 et M2/P2. Les gènes 5S sont représentés par des rectangles vides tandis que le motif 5SmsCR1 est symbolisé par un rectangle quadrillé au sein des espaceurs figurés par des lignes. La position et l'orientation des amorces sont représentées par des flèches. Les dimères sont produits lorsque les sites de fixation des amorces sur l'ADN cible ne sont pas tous saturés (déficit en amorce, inaccessibilité ou modification des séquences cibles).

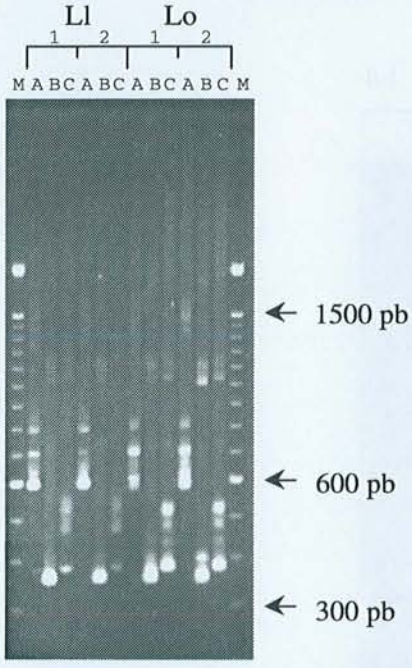



**Figure 54**. Profils de migration des fragments d'ADNr 5S amplifiés à partir des mélèzes Américains. La signification des abréviations d'espèces utilisées (Ll, Lo) est indiquée dans le **Tableau 34**. A : amplifications des espaceurs ; B : amplifications du fragment 5'sr ; C : amplifications du fragment 3'sr ; M : ladder 100 pb ; 1 et 2 : les deux individus testés par espèce.

Si l'on utilise de nouveau l'exemple de *L. russica*, les dimères amplifiés devraient en effet avoir une taille proche de 1651 pb (au lieu de 1449 pb, amorces P1/P2), 1213 pb (au lieu de 1016 pb, amorces P1/M1) ou 1336 pb (au lieu de 1138 pb, amorces M2/P2). Les calculs similaires réalisés pour les 3 autres espèces, et en particulier *Larix gmelinii* aboutissent au même constat d'alternance des unités courtes et longues (calculs non détaillés).

*- Les mélèzes américains*

Les deux espèces américaines testées (*Larix occidentalis* et *Larix laricina*) présentent des profils d'amplification semblables à ceux des espèces eurasiennes (**Figure 54**). Toutefois, une classe d'espaceurs supplémentaire, de taille intermédiaire (650 pb) entre les classes d'espaceurs courts (~ 540 pb) et longs (744 pb), est observée dans la zone des monomères (**Tableau 38**). La taille des fragments 5' sr et 3' sr obtenus indiquent la présence du motif 5SmsCR1, non seulement dans les espaceurs longs, mais aussi dans ces espaceurs de taille intermédiaire. L'analyse des dimères (**Tableau 39**) révèle, comme chez les mélèzes eurasiens (comparer avec le **Tableau 37**) une organisation alternée des unités courtes et longues au sein de séries composites. La taille de certains dimères obtenus avec les couples d'amorces P1/P2 (~ 1370 pb) et M2/P2 (~ 1040 pb) montre aussi l'existence de séries composites où alternent des unités courtes et intermédiaires. L'obtention d'un unique fragment dimérique avec les amorces P1/M1 (~ 1015 pb, **Tableau 39**) indique que les régions 5' sr des espaceurs longs et des espaceurs intermédiaires ont une taille similaire et que la différence de longueur entre ces deux classes de taille est principalement localisée au sein de la région 3' sr (environ 100 pb de différence, **Tableau 38**).

**Tableau 37**. Taille (en pb) des fragments amplifiés (zones des dimères) à l'aide des couples d'amorces P1/P2, P1/M1 et M2/P2 chez les mélèzes eurasiens[a].

| Nature de l'amplification | Espèce[b] | | | | Organisation déduite[c] |
|---|---|---|---|---|---|
| | Ld | Lr | Lk | Lg | |
| **P1/P2** | 1436 | 1449 | 1451 | 1464 | Unités courtes et longues alternées |
| **P1/M1** | 1007 | 1016 | 1012 | 1014 | Unités courtes et longues alternées |
| **M2/P2** | 1129 | 1138 | 1136 | 1121 | Unités courtes et longues alternées |

[a] Taille des fragments majeurs, communs aux deux individus testés et reproductibles (amplifiés à 4 reprises)

[b] Les abréviations utilisées sont indiquées dans le **Tableau 34**

[c] Déduction sur la base des calculs de tailles de dimères à partir des fragments monomères (**Tableau 36**)

**Tableau 38**. Taille (en pb) des fragments amplifiés (zones des monomères) à l'aide des couples d'amorces P1/P2 (espaceur obs.), P1/M1 (5' sr) et M2/P2 (3' sr) chez les mélèzes américains*.

| Classes | Espaceur obs. | | 5'sr | | 3' sr | | Espaceur cal. | |
|---|---|---|---|---|---|---|---|---|
| | Lo | Ll | Lo | Ll | Lo | Ll | Lo | Ll |
| **Espaceurs longs** | 744 | 744 | 350 | 353 | 482 | 488 | 749 | 758 |
| | | | 400 | | 435 | 468 | | |
| | | | | | 516 | 525 | | |
| | | | | | 558 | 558 | | |
| **Espaceurs intermédiaires** | 650 | 651 | 350 | 353 | 375 | 380 | 642 | 650 |
| | | | | | 389 | | | |
| **Espaceurs courts** | 536 | 544 | | | | | | |

* Les abréviations utilisées sont données dans les **Tableaux 34 et 36**

Sur le plan de l'organisation et de l'origine des unités d'ADNr 5S chez ces mélèzes américains, la présence du motif 5SmsCR1 à la fois dans les unités longues et intermédiaires, l'alternance de celles-ci avec les unités courtes ainsi que les tailles similaires des fragments 5' sr suggèrent qu'elles sont fortement apparentées. Une étude de séquence permettrait de le démontrer.

**Tableau 39**. Taille (en pb) des fragments amplifiés (zones des dimères) à l'aide des couples d'amorces P1/P2, P1/M1 et M2/P2 chez les mélèzes américains.

| Nature de l'amplification | Espèce[a] | | Organisation déduite[b] |
|---|---|---|---|
| | Lo | Ll | |
| **P1/P2** | 1373 | 1376 | Unités courtes et intermédiaires alternées |
| | 1478 | 1485 | Unités courtes et longues alternées |
| **P1/M1** | 1013 | 1020 | Unités courtes et longues ou intermédiaires alternées |
| **M2/P2** | 1043 | 1045 | Unités courtes et intermédiaires alternées |
| | 1141 | 1159 | Unités courtes et longues alternées |

[a] Les abréviations utilisées sont données dans le **Tableau 34**

[b] Déduction sur la base des calculs de tailles de dimères à partir des fragments monomères (**Tableau 38**)

**Tableau 40**. Taille (en pb) des fragments amplifiés (zones des monomères) à l'aide des couples d'amorces P1/P2 (espaceur obs.), P1/M1 (5' sr) et M2/P2 (3' sr) chez les épicéas*.

| Classes | Espaceur obs. | | | 5'sr | | | 3' sr | | | Espaceur cal. | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Pg | Pm | Pa | Pg | Pm | Pa | Pg | Pm | Pa | Pg | Pm | Pa |
| **Espaceurs longs hypothétiques** | | | | 242 | 254 | 254 | 299 | 305 | 306 | 459 | 477 | 478 |
| | | | | 269 | 278 | | | 247 | 335 | | | |
| | | | | 189 | | 196 | | 339 | | | | |
| | | | | 509 | | 495 | | 539 | | | | |
| **Espaceurs courts** | 117 | 116 | 113 | | | | | | | | | |

* Les abréviations utilisées sont données dans les **Tableaux 34 et 36**

β- Les épicéas

Les amplifications réalisées à l'aide des amorces P1/P2 révèlent des profils similaires pour les trois espèces testées (**Figure 55**) : *Picea glauca*, *Picea mariana* (espèces américaines) et *Picea abies* (espèce européenne). On observe une seule classe d'espaceurs très courts (~ 115 pb, **Tableau 40**), de taille comparable à celle des espaceurs de *Picea glauca* (101 pb) séquencés par Brown et Carlson (1997).

L'examen des fragments 5' sr et 3' sr (**Tableau 40**) permet de conclure que ces espaceurs très courts ne possèdent pas le motif 5SmsCR1 comme le laissait déjà entrevoir l'analyse de la séquence de *Picea glauca* (voir la **Partie II-A-5** précédente). La taille des espaceurs déduits de ces fragments est en effet largement supérieure à celle des espaceurs courts. Cependant, la possibilité d'obtenir une amplification avec les amorces P1/M1 et M2/P2 suggère l'existence d'une hypothétique classe d'espaceurs d'environ 470 pb qui reste indécelable par amplification directe à l'aide des amorces P1/P2 et qui contiendrait tout ou partie du motif 5SmsCR1. Les unités correspondantes d'environ 600 pb

(470 + 120) ont pourtant été détectées récemment chez *Picea glauca* (Brown et Carlson 1997) par analyse RFLP de l'ADN génomique à l'aide d'une sonde d'ADNr 5S homologue. Les diverses tentatives d'amplification directe de cette classe d'unités par ces auteurs à l'aide d'amorces universelles s'hybridant sur le gène 5S (sites d'hybridation différents de P1 et P2) se sont aussi soldées par des échecs. Ces unités de 600 pb représentent donc probablement les vestiges d'un ancien locus fonctionnel qui renferme des pseudogènes dont la séquence très perturbée limite l'appariement avec les amorces universelles. Le succès des amplifications P1/M1 (5' sr) et M2/P2 (3' sr) pourrait alors résulter du fait que les amorces M1 et M2 ont des Tm beaucoup plus élevés que les amorces P1 et P2. Leur hybridation avec les séquences matrices est donc plus stable, même en cas de mutations au niveau du site cible, ce qui permettrait une amplification correcte des fragments 5' sr et 3' sr. L'absence de produits d'amplification P1/P2 issus de ces unités de 600 pb complique l'interprétation des profils complexes obtenus à l'aide des amorces P1/M1 et M2/P2. De plus, certains fragments 5'sr et 3' sr ne sont pas communs entre les deux individus analysés (**Figure 55**). Aucune information fiable sur l'organisation relative des unités courtes (221 pb) et longues (600 pb) n'a donc pu être dégagée de l'analyse des fragments pouvant correspondre à des dimères.

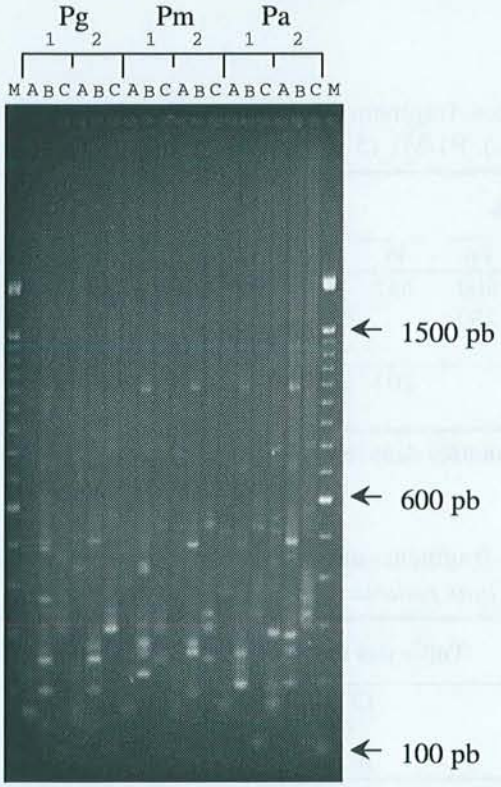



**Figure 55.** Profils de migration des fragments d'ADNr 5S amplifiés à partir des épicéas. La signification des abréviations d'espèces utilisées (Pg, Pm, Pa) est indiquée dans le **Tableau 34**. A : amplifications des espaceurs ; B : amplifications du fragment 5'sr ; C : amplifications du fragment 3'sr ; M : ladder 100 pb ; 1 et 2 : les deux individus testés par espèce.

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY



χ- Les pins

Contrairement aux espèces des deux genres précédents, les 4 espèces de pins étudiées (*Pinus radiata*, *Pinus sylvestris*, *Pinus nigra* et *Pinus pinaster*) présentent des profils d'amplification P1/P2 relativement différents (**Figure 56**). Les tailles moyennes des fragments observés dans la zone des monomères sont données dans le **Tableau 41**. Deux ou trois classes de taille sont détectées chez *Pinus radiata* (694 pb, 736 pb et 405 pb) et *Pinus sylvestris* (588 pb, 489 pb et un fragment de 351 pb peu intense), une seulement chez *Pinus nigra* (592 pb) et une ou deux chez *Pinus pinaster* (730 pb et 800 pb). Dans le cas de *Pinus radiata*, la présence des deux classes d'unités d'environ 850 pb et 525 pb caractérisées par Moran *et al.* (1992) est confirmée, même si la classe la plus courte est amplifiée avec une intensité relativement faible qui rend parfois sa détection difficile (**Figure 56**). Cette faiblesse d'amplification pourrait résulter d'une mauvaise fixation des amorces P1 et P2 ou être la conséquence de la sous-représentation des unités courtes par rapport aux unités longues dans le génome des deux pins radiata étudiés. Cependant, Moran *et al.* (1992), ont le plus souvent observé la situation inverse (environ 10 fois plus d'unités courtes que de longues) en étudiant une trentaine d'individus issus de plusieurs populations naturelles.

**Tableau 41**. Taille (en pb) des fragments amplifiés (zones des monomères) à l'aide des couples d'amorces P1/P2 (espaceur obs.), P1/M1 (5' sr) et M2/P2 (3' sr) chez les pins*.

| Classes | Espaceur obs. | | | | 5'sr | | | | 3' sr | | | | Espaceur cal. | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Pr | Ps | Pn | Pp | Pr | Ps | Pn | Pp | Pr | Ps | Pn | Pp | Pr | Ps | Pn | Pp |
| **Espaceurs longs** | 694<br>736 | 588 | 592 | 800<br>730 | 552 | 443<br>540 | 441<br>520<br>215 | 590<br>380<br>170 | 229<br>274 | 237 | 237<br>224<br>280 | 580<br>430 | 698<br>743 | 597 | 595 | ?<br>728 |
| **Espaceurs courts** | 405 | 489<br>351 | | | 203 | 270 | | | 274 | 280 | | | 394 | 467 | | |

* Les abréviations utilisées sont données dans les **Tableaux 34 et 36**

**Tableau 42**. Taille (en pb) des fragments amplifiés (zones des dimères) à l'aide des couples d'amorces P1/P2, P1/M1 et M2/P2 chez *Pinus radiata*.

| Nature de l'amplification | Taille des fragments | Organisation déduite* |
|---|---|---|
| **P1/P2** | 1321 | Unités courtes et longues alternées |
| **P1/M1** | 751 | Série homogène d'unités courtes |
| **M2/P2** | 1038 | Série homogène d'unités longues |

*Déduction sur la base des calculs de tailles de dimères à partir des fragments monomères (**Tableau 41**)

La taille des fragments majeurs 5' sr et 3' sr dont la somme correspond à la taille des espaceurs amplifiés par les amorces P1/P2 atteste clairement de la présence du motif 5SmsCR1 au sein de toutes les classes d'espaceurs révélées chez *Pinus radiata*, *Pinus nigra* et *Pinus sylvestris* (**Tableau 41**). Ce n'est pas le cas en revanche pour *Pinus pinaster*. La somme des fragments majeurs 5' sr (590 pb) et 3'

sr (580 pb) dépasse en effet largement celle des différents espaceurs obtenus par amplification à l'aide du couple d'amorces P1/P2. Une combinaison associant les fragments mineurs 5' sr de 380 pb et 3' sr de 580 pb est toutefois compatible avec la présence du motif 5SmsCR1 au sein des espaceurs de 730 pb. L'obtention de fragments majeurs 5' sr et 3' sr reproductibles dont les tailles additionnées ne correspondent pas à celle des espaceurs révélés par les amorces P1/P2 pourrait être expliquée par la présence de plusieurs motifs 5SmsCR1 au sein des espaceurs de cette espèce comme c'est le cas chez *Pseudotsuga menziesii* (voir **Figures 46 et 48**). Les données obtenues chez *Pinus pinaster* ne permettent toutefois pas de soupçonner la présence d'unités de grande taille (plus de 1000 pb) susceptibles de contenir plusieurs copies du motif 5SmsCR1.

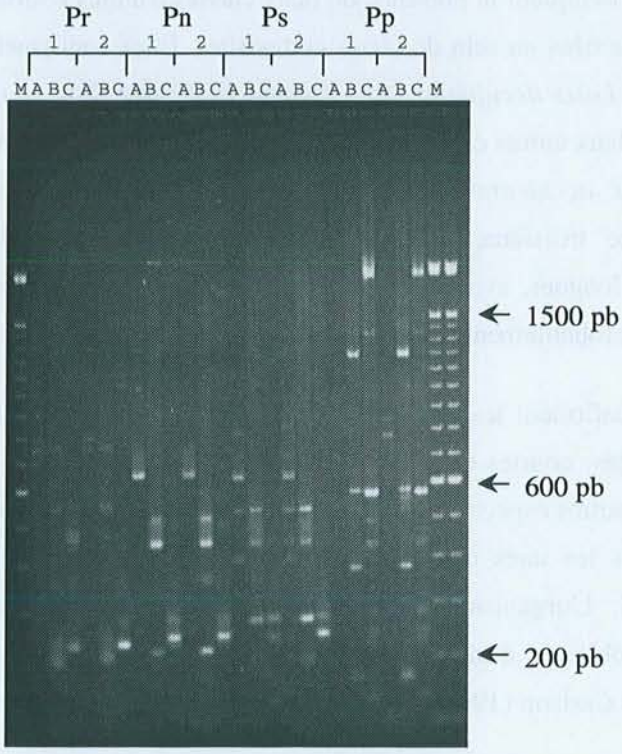



**Figure 56**. Profils de migration des fragments d'ADNr 5S amplifiés à partir des pins. La signification des abréviations d'espèces utilisées (Pr, Pn, Ps, Pp) est indiquée dans le **Tableau 34**. A : amplifications des espaceurs ; B : amplifications du fragment 5'sr ; C : amplifications du fragment 3'sr ; M : ladder 100 pb ; 1 et 2 : les deux individus testés par espèce.

L'analyse des dimères n'a été possible que chez *Pinus radiata* (**Tableau 42**) en raison de l'absence d'amplification chez *Pinus nigra* et *Pinus sylvestris* ou de la taille "exotique" des fragments obtenus chez *Pinus pinaster*. Bizarrement, la taille du dimère P1/P2 de 1321 pb est compatible avec une organisation alternée des unités courtes et longues, tandis que les tailles des dimères P1/M1 (751 pb) et M2/P2 (1038 pb) indiquent plutôt des séries homogènes d'unités courtes ou d'unités longues, respectivement. Cette situation suggère que dans le génome de *Pinus radiata*, où l'ADNr 5S est dispersé à de multiples loci (Gorman *et al.* 1992), il n'y aurait pas d'organisation préférentielle des deux classes d'unités qui pourraient être ou non associées au sein de séries composites. Dans cette hypothèse, certains dimères ne sont donc pas détectés (3 catégories de dimères devraient être observées pour chaque couple d'amorces).

δ- *Abies alba* et *Ginkgo biloba*

Deux classes d'espaceurs ont été décelées chez *Abies alba* (environ 630 pb et 420 pb) et *Ginkgo biloba* (environ 370 pb 175 pb) à l'aide des amorces P1/P2 (données non montrées). En revanche, les amplifications des fragments 5' sr et 3' sr se sont avérées très faibles et non reproductibles, indiquant une probable absence du motif 5SmsCR1 chez ces deux espèces.

**4- Conclusions**

a- Organisation des unités d'ADNr 5S

Les données recueillies chez les mélèzes indiquent la présence de deux classes d'unités courtes (~ 650 pb) et longues (~ 870 pb) largement alternées au sein de séries composites. Elles confirment ainsi les résultats obtenus antérieurement chez *Larix decidua* et *Larix kaempferi* (**Partie II-A** de ce **Chapitre IV**) et élargissent cette conclusion à deux autres espèces eurasiennes, *Larix russica* et *Larix gmelinii*, et deux espèces américaines, *Larix occidentalis* et *Larix laricina*. Ces dernières se distinguent cependant par la présence d'une troisième classe d'unités (~ 770 pb, qualifiée d'intermédiaire) qui alterne, comme les unités longues, avec des unités courtes. L'établissement des séries composites est donc ancienne et précède probablement la diversification du genre *Larix*.

Les résultats obtenus pour les épicéas confirment les travaux de Brown et Carlson (1997) qui font état de la présence de deux classes d'unités, courtes (~ 230 pb) et longues (~ 600 pb) dans le génome de *Picea glauca* et les étendent à deux autres espèces, *Picea mariana* et *Picea abies*. Ces trois espèces ne peuvent donc pas être distinguées les unes des autres sur la seule base des profils d'amplification des espaceurs de l'ADNr 5S. L'organisation relative des deux classes d'unités observées n'a pu être élucidée en raison de problèmes d'amplification de la classe d'unités longues, problèmes signalés précédemment par Brown et Carlson (1997).

Chez *Pinus radiata*, nous avons retrouvé les résultats de Moran *et al.* (1992) qui indiquent la présence d'unités courtes (~ 525 pb) et longues (~ 850 pb). En outre, nous avons également observé la présence d'unités de taille différente chez *Pinus sylvestris* (~ 470 pb, 610 pb et 710 pb) et *Pinus pinaster* (~ 850 pb et 920 pb). En revanche, une seule classe d'unités a été détectée chez *Pinus nigra* (710 pb). Cette plus grande hétérogénéité de taille des unités d'ADNr 5S relevée dans le genre *Pinus*, comparé aux genres *Larix* et *Picea* pourrait être utilisée pour l'identification des espèces, après validation définitive sur des échantillons représentatifs des populations naturelles.

b- Présence du motif 5SmsCR1

Le motif 5SmsCR1, qui pourrait correspondre à une partie de gènes protéasomiques ou plus probablement à un pseudogène d'ARNt, a été détecté chez la plupart des espèces de *Pinaceae* testées. Il est présent au sein des espaceurs courts ou longs de différents pins tels que *Pinus radiata*, *Pinus sylvestris*, *Pinus nigra* et *Pinus pinaster*, bien que chez cette dernière espèce l'interprétation des

profils d'amplification n'ait pas pu être complètement résolue. Chez les mélèzes, en revanche, il est seulement présent au sein des espaceurs longs (*Larix decidua*, *Larix kaempferi*, *Larix russica*, *Larix gmelinii*, espèces eurasiennes) et intermédiaires (*Larix laricina*, *Larix occidentalis*, espèces américaines). Chez *Picea glauca*, *Picea mariana* et *Picea abies*, la présence du motif semble liée à l'existence d'une hypothétique classe d'unités longues d'environ 600 pb qui demeure très difficile à mettre en évidence par les techniques classiques de PCR. Enfin, le motif 5SmsCR1 semble absent chez *Abies alba* et *Ginkgo biloba*.

Les phylogénies moléculaires récemment publiées tendent à montrer que les gymnospermes ont une origine monophylétique (voir Chaw *et al.* 1997 ainsi que les références citées dans cet article) et que le groupe formé par les Ginkgoales (*Ginkgoaceae*) et les Cycadales (*Cycaceae*, *Zamiaceae*) est probablement celui qui s'est différencié le plus tôt. Sont apparues ensuite les Gnetales (*Ephedraceae*, *Welwitschiaceae*), puis le groupe plus vaste des Coniferales[12] qui contient la famille des *Pinaceae*. L'absence du motif 5SmsCR1 chez les angiospermes et chez *Gingko biloba* ainsi que sa présence irrégulière chez les *Pinaceae* pourraient donc lui conférer un intérêt phylogénétique à un niveau taxinomique assez élevé au sein de cette famille (entre des genres ou entre des groupes de genres). Dans cette optique, il serait judicieux d'élargir la recherche du motif à d'autres membres de cette famille (*Cathaya*, *Cedrus*, *Hesperopeuce*, *Keteleeria*, *Nothotsuga*, *Pseudolarix*, *Tsuga*), et réaliser quelques études de séquences.

La mise en évidence du motif 5SmsCR1 par PCR à l'aide des couples d'amorces P1/P2, P1/M1 et M2/P2 semble bien fonctionner, en particulier chez les mélèzes à partir desquels nous avons optimisé le protocole. Elle doit cependant être encore améliorée car les amplifications réalisées sur les autres espèces sont globalement moins efficaces, probablement en raison d'une hybridation plus difficile des amorces. De plus, les amorces M1 et M2 ne peuvent pas être utilisées en couple car elles sont complémentaires au niveau de leurs extrémités 3' (voir **Figure 49**). L'utilisation d'un quatrième couple d'amorces complémentaires de M1 et M2 permettant d'amplifier le fragment d'ADN compris entre deux motifs 5SmsCR1 (en plus de l'espaceur et des fragments 5' sr et 3' sr) pourrait ainsi déboucher sur un test de détection plus robuste et plus informatif sur l'organisation des unités.

---

[12] Les coniferales comportent les 9 familles suivantes : *Araucariaceae*, *Cephalotaxaceae*, *Cupressaceae*, *Phyllocladaceae*, *Pinaceae*, *Podocarpaceae*, *Sciadopityaceae*, *Taxaceae* et *Taxodiaceae*.

## III- Héritabilité, diversité et divergence nucléotidique des unités d'ADNr 5S chez les mélèzes d'Europe (*Larix decidua* Mill.) et du Japon [*Larix kaempferi* (Lamb.) Carr.]

### 1- Introduction

Comme chez la plupart des plantes supérieures (Sastri *et al.* 1992), l'ADNr 5S des mélèzes d'Europe (*Larix decidua* Mill.) et du Japon [*Larix kaempferi* (Lamb.) Carr.] est organisé en longues séries d'unités répétées en tandem constituées d'un gène codant pour l'ARNr 5S (gène 5S) d'environ 120 pb et d'un espaceur non transcrit de taille variable. Deux classes d'unités d'environ 650 pb et 870 pb, présentes chacune à 1500-3000 copies par génome haploïde, ont été détectées par amplification *in vitro* à l'aide d'amorces universelles. L'analyse RFLP de l'ADN génomique que nous avons menée a permis de révéler la présence chez ces deux espèces de séries hétérogènes ("composites") au sein desquelles les unités courtes (S) et longues (L) présentent une organisation majoritairement alternée (**Partie II-A** du **Chapitre IV**). Seule une petite fraction des unités longues forment des séries homogènes. Celles-ci sont sans doute adjacentes ou intégrées aux séries composites car elles occupent apparemment avec elles un unique locus chromosomique (Lubaretz *et al.* 1996).

L'organisation alternée prépondérante des unités courtes et longues semble générale, et donc ancienne et stable, au sein du genre *Larix*. Elle a en effet été retrouvée chez quatre autres espèces, eurasiennes (*Larix russica*, *Larix gmelinii*) ou américaines (*Larix laricina*, *Larix occidentalis*) (**Partie II-B** précédente). Cette organisation apparaît d'autant plus originale que les quelques séquençages réalisés chez le mélèze d'Europe indiquent une forte divergence nucléotidique des espaceurs courts et des espaceurs longs (homologie réduite aux extrémités 5' et 3'). Chez les plantes supérieures, les unités d'ADNr 5S très divergentes se trouvent en effet habituellement localisées à des loci différents (Scoles *et al.* 1988, Kellogg et Appels 1995) car les mécanismes de l'évolution concertée s'exercent surtout au sein des loci et très peu entre loci (Cronn *et al.* 1996). Les analyses des régions ITS de l'ADNr (1759-1770 pb) réalisées récemment chez différents taxa de mélèzes confirment cette situation (Gernandt et Liston 1999). Au niveau de l'ADNr 5S des mélèzes, il semble donc que l'évolution concertée intralocus ait conduit au cours de l'évolution à l'homogénéisation d'un motif dimérique de taille importante (environ 1450 pb) constitué d'une unité courte et d'une unité longue qui alternent donc au sein des séries.

Une telle situation ouvre *a priori* des possibilités intéressantes dans l'optique d'un approfondissement des études taxinomiques au sein du genre *Larix*. La structure et la taille du motif dimérique semblent en particulier tout à fait adaptées à la recherche de marqueurs nucléaires de différenciation intra- et interspécifique entre mélèzes d'Europe et du Japon, espèces économiquement importantes pour lesquelles un programme d'amélioration génétique a été mis en place (Pâques 1996a, 1996b, Arcade *et al.* 1996 ; voir également la **Partie I** de ce **Chapitre IV**). Le motif dimérique offre d'autre part des facilités d'investigation au moins comparables à celles des ITS de l'ADNr, régions souvents exploitées dans les études phylogénétiques, mais qui s'avèrent insuffisantes dans le cas des

mélèzes (Gernandt et Liston 1999). Enfin, d'un point de vue plus fondamental, l'organisation alternée d'unités divergentes constitue une situation intéressante pour analyser et comprendre les mécanismes de l'évolution concertée des familles multigéniques.

Les informations disponibles concernant la diversité moléculaire (deux classes d'unités) et le déterminisme génétique de l'héritabilité des unités d'ADNr 5S (*a priori* un seul locus) chez les mélèzes ne sont cependant pas suffisantes pour permettre l'analyse de la variabilité génétique ou des relations phylogénétiques en raison des problèmes de paralogie/orthologie soulevés par l'évolution concertée (**Figure 19**). Dans cette étude préalable, nous nous proposons donc, comme nous l'avions fait précédemment pour les chênes pédonculés et sessiles (voir le **Chapitre III**), d'étudier l'héritabilité et d'évaluer la diversité moléculaire des unités d'ADNr 5S chez les mélèzes d'Europe et du Japon. Cette étude d'héritabilité a été entreprise afin de vérifier le caractère mendélien des marqueurs de l'ADNr 5S et de préciser le nombre de loci impliqués. Elle a été réalisée à partir d'un croisement contrôlé interspécifique qui avait déjà été utilisé par Arcade *et al.* (2000) pour construire une carte génétique des mélèzes d'Europe et du Japon. L'estimation de la diversité moléculaire, quant à elle, a pour objectifs essentiels d'inventorier les séquences paralogues (et définir ainsi les relations d'orthologie), d'apprécier l'efficacité et le mode d'action de l'évolution concertée chez ces espèces à long cycle et, finalement, de déterminer le degré de divergence nucléotidique de l'ADNr 5S entre ces deux espèces. Les profils d'amplification des espaceurs obtenus chez différentes provenances de mélèzes d'Europe et du Japon étant très similaires (**Figure 41**), nous avons abordé cette étude de diversité moléculaire directement au niveau nucléotidique à partir d'un effectif limité d'individus de différentes provenances de chaque espèce. Le plan de séquençage retenu nous a permis d'estimer indépendamment la diversité nucléotidique dans les gènes 5S, les espaceurs courts et les espaceurs longs. En conclusion, l'origine et l'évolution du motif dimérique amplifié chez les mélèzes ont été discutées sur la base de ces données nouvelles.

## 2- Matériel et méthodes

### a- Matériel végétal

L'étude de la diversité nucléotidique au sein des unités de l'ADNr 5S a été réalisée à partir de 4 mélèzes d'Europe (génotypes 92-1, 107, 285 et 475) et de 3 mélèzes du Japon (génotypes 3211, 3076 et 3247). Il n'existe aucun lien de parenté directe entre ces arbres. Les génotypes européens 107 (Olomouc, Sudètes tchèques), 285 (Bystrzyca Klodsko, Sudètes polonaises) et 475 (Wienerwald, Alpes autrichiennes), sélectionnés en forêt dans l'aire naturelle de l'espèce, ainsi que les génotypes japonais 3211 (Japon, Mont Fuji) et 3247 (Japon, Okunikko), sélectionnés en test de provenances (IUFRO), nous ont été fournis par L.E. Pâques (INRA Orléans) sous forme d'aiguilles. Le génotype européen 92-1 (origine française) est quant à lui multiplié au laboratoire sous forme d'une lignée embryogène obtenue par M.A. Lelu (INRA Orléans) en 1992 à partir d'un embryon zygotique immature. Enfin, le génotype japonais 3076 a été mis à notre disposition par D. Prat (INRA d'Orléans)

sous forme d'un extrait d'ADN génomique. Il est issu d'un croisement entre arbres plein-frères sélectionnés en forêt artificielle au Danemark.

L'étude de l'héritabilité des unités d'ADNr 5S a été réalisée à partir des extraits d'ADN génomique de 100 individus de la famille hybride 179 (*Larix x eurolepis*) obtenue à l'INRA d'Orléans en 1982 par croisement contrôlé entre le mélèze d'Europe 107 (utilisé comme femelle) et le mélèze du Japon 3076 (utilisé comme mâle). Ce matériel nous a été procuré par D. Prat.

b- Extraction d'ADN génomique, PCR et électrophorèse

L'ADN génomique a été extrait et purifié à partir d'aiguilles (génotypes 92-1, 107, 285, 475, 3211 et 3247) ou de masses cellulaires embryogènes en cours de multiplication (génotype 92-1) selon la méthode de Doyle et Doyle (1987) avec des modifications mineures (**Partie II-A** de ce **Chapitre IV**). Les extraits d'ADN génomique de la famille hybride 179 ont été préparés à l'INRA d'Orléans selon la méthode au CTAB décrite par Arcade *et al.* (1996).

L'amplification par PCR des espaceurs et/ou de la région transcrite des unités d'ADNr 5S a été réalisée à l'aide des couples d'amorces respectifs P1/P2 et P3/P4. Les séquences et positions des amorces ainsi que les conditions d'amplification utilisées ont été décrites dans la **Partie II-A** de ce **Chapitre IV**. Après mélange avec 10 % (v/v) de tampon de charge 10X [50 % (v/v) de glycérol, 0,4 % (p/v) de bleu de bromophénol, 0,4 % (p/v) de xylène cyanol], les produits d'amplification (10 µl) sont séparés pendant 4 h sur gel d'agarose à 1,5 % confectionné dans du TBE 0,5X (45 mM de Tris-borate, 1 mM d'EDTA). La révélation des fragments a été effectuée par immersion du gel dans une solution de bromure d'éthidium à 0,5 µg/ml pendant 30 min. Pour l'étude de la famille 179, nous avons en outre utilisé des gels d'acrylamide non dénaturants (8 % T, 3,3 % C) de 30 cm de longueur et 0,75 mm d'épaisseur réalisés dans du TBE 1X (90 mM de Tris borate, 2 mM d'EDTA). Dans ce cas, seulement 2 µl de chaque échantillon sont déposés sur le gel et l'électrophorèse est effectuée sous une tension de 450 V pendant 12 h (distance parcourue par les fragments de 600 pb : environ 20 cm). La révélation des profils d'amplification est réalisée par coloration au nitrate d'argent (Bassam *et al.* 1991). La taille des fragments est estimée par comparaison de leur mobilité électrophorétique avec celle des bandes du marqueur de taille moléculaire "Ladder 100 pb" (15 fragments de 100 à 1500 pb avec un pas de 100 pb, Gibco BRL).

c- Clonage et séquençage

Afin d'éviter les problèmes posés par le séquençage direct de produits de PCR hétérogènes, probablement aggravés ici par les variations nucléotidiques *cis* et *trans* au sein des séries d'ADNr 5S, nous avons procédé par clonage selon le protocole suivant : les produits de PCR ont été purifiés et concentrés par extraction au phénol/chloroforme, précipitation à l'éthanol absolu/acétate d'ammonium, lavage à l'éthanol 70 %, séchage et redissolution dans un volume approprié de TE (10 mM Tris-HCl pH 8,0, 1 mM EDTA). Les fragments purifiés ont été ensuite digérés par les enzymes de restriction

appropriées (*Eco*R I et *Sac* I pour les produits d'amplification P1/P2 ; *Sal* I et *Sac* II pour les produits d'amplification P3/P4) avant d'être séparés sur gel d'agarose à 1,5 %. Après coloration au bromure d'éthidium, les bandes correspondant aux fractions de 200-300 pb (P3/P4), 600 pb et 800 pb (P1/P2) ont été découpées individuellement sous UV (312 nm) puis extraites du gel d'agarose en utilisant le kit commercial QIAquick (Qiagen). Les fragments extraits ont alors été insérés dans le plasmide pBluescript KS +/- (Stratagene), préalablement linéarisé (*Eco*R I/*Sac* I ou *Sal* I/*Sac* II), par ligation à l'aide de l'enzyme T4 DNA ligase (Boehringer) selon les recommandations du fournisseur. Les vecteurs recombinants ont été multipliés par la souche DH5α d'*E. coli* (Gibco BRL) transformée selon le protocole standard basé sur un traitement au chlorure de calcium (Sambrook *et al.* 1989). La taille des inserts est estimée par PCR (1 µl de culture bactérienne à saturation pour 20 µl de mélange réactionnel) en utilisant les couples d'amorces P1/P2 (espaceur) ou P3/P4 (région transcrite) à une température d'hybridation inférieure de 3°C à celle du protocole standard (50°C au lieu de 53°C). L'extraction et la purification des plasmides à partir des clones bactériens sélectionnés a été obtenue en utilisant le "plasmid extraction minikit" (Qiagen). Les séquences nucléotidiques ont été déterminées par la société Genome express SA (Grenoble, France) à l'aide d'un séquenceur automatique ABI 373 (méthode de Sanger, protocoles fournis par Applied Biosystems).

d- Alignement et analyse des séquences

Les alignements des séquences de gènes 5S, d'espaceurs courts ou d'espaceurs longs ont d'abord été générés par le logiciel Clustal V (Higgins *et al.* 1992) et ensuite optimisés par ajustement manuel à l'aide de la version 1.9a157 de l'éditeur de séquence SeqApp (D. Gilbert, Indiana University).

Afin de quantifier au mieux les variations de séquence, nous avons calculé les différences nucléotidiques moyennes par site entre deux séquences choisies au hasard, soit dans chacune des deux espèces (diversité nucléotidique notée $\pi$), soit entre les espèces (notée $d$). Une estimation de $\pi$ est donnée par :

$$\pi = \sum_{i<j} \pi_{ij}/n_c$$

où $\pi_{ij}$ représente la proportion de nucléotides différents entre les séquences i et j (distance brute ou non corrigée) et $n_c$, le nombre total de comparaisons de séquences effectuées ($n_c = n(n-1)/2$ où n est le nombre de séquences examinées). Les distances brutes ont été calculées manuellement (gènes 5S) ou en utilisant l'option "showdist" du logiciel PAUP (Swofford 1993). La variance de chaque estimation de $\pi$ a été calculée selon Nei (1987, équation 10.7, p. 256).

De manière similaire, $d$ est définie par :

$$d = \sum_{ij} d_{ij}/(n_1 \times n_2)$$

où $d_{ij}$ représente la proportion de nucléotides différents entre la séquence i de la population X ($n_1$ séquences examinées) et la séquence j de la population Y ($n_2$ séquences examinées).

Ces deux paramètres de diversité présentent l'avantage d'être moins sensibles à la taille de l'échantillon de séquences examinées (Nei 1987) que d'autres, tels que le nombre de séquences polymorphes ou la proportion de sites polymorphes.

Les arbres de gènes ont été construits à l'aide du logiciel PAUP version 3.1.1 (Swofford 1993) par la méthode de parcimonie maximum en utilisant l'algorithme heuristique général sans pondération des états de caractères et sans spécification de la racine (valeurs par défaut). Les positions de l'alignement comportant des indels (insertions ou délétions) ont été inclues dans les analyses phylogénétiques. Les "gaps" d'alignement ont alors simplement été traités comme des données manquantes. La robustesse de chacun des groupes de séquences obtenus a été testée par la méthode du bootstrap (100 répétitions).

**3- Résultats**

a- Héritabilité des unités d'ADNr 5S

α- phénotypes des arbres 107 et 3076

Les profils d'amplification des espaceurs (couple d'amorces P1/P2) obtenus sur gel d'agarose pour les arbres parentaux 107 et 3076 sont présentés dans la **Figure 57**. Ils sont conformes aux résultats obtenus précédemment (**Partie II-A** de ce **Chapitre IV**). Deux niveaux de taille, d'environ 580 pb et 800 pb, correspondant respectivement aux deux classes d'espaceurs "S" (espaceurs courts, environ 530 pb) et "L" (espaceurs longs, environ 750 pb) sont en effet observés, ainsi que les hétéroduplexes hS (~ 650 pb) et hL (~ 850 pb) résultant de l'hétérogénéité de séquence au sein de chaque classe. On retrouve de même la présence des fragments dimériques (dSL) d'environ 1450 pb, révélateurs de l'alternance des unités courtes et longues observées chez les mélèzes.

Cependant, au-delà de cette similitude globale des profils, les deux arbres apparaissent polymorphes. Trois marqueurs basés sur des différences de taille des espaceurs ($\mathbf{S_1}$, $S_2$ et $\mathbf{L_1}$) permettent de les différencier. Les deux premiers ($\mathbf{S_1}$ et $S_2$) concernent la classe des espaceurs courts au niveau de laquelle on observe deux fragments majeurs : un fragment commun de 565 pb ($S_0$) et un autre de 580 pb chez l'arbre 107 ($\mathbf{S_1}$) et de 600 pb chez l'arbre 3076 ($S_2$). Le troisième marqueur ($\mathbf{L_1}$) appartient à la classe des espaceurs longs où l'on trouve deux fragments majeurs de 785 pb ($\mathbf{L_1}$) et 820 pb ($L_0$) chez le parent femelle 107 tandis que chez le parent mâle 3076 seul le fragment de 820 pb ($L_0$) est clairement détecté. Le fragment de 785 pb y est parfois aussi observé, mais uniquement à l'état de traces. Malgré tout, la ségrégation de ce marqueur $\mathbf{L_1}$ a pu être analysée en utilisant l'intensité des fragments plutôt que l'habituelle présence/absence. L'électrophorèse sur gel d'acrylamide à 8 % de ces mêmes produits d'amplification P1/P2 a en outre permis de détecter deux bandes discriminantes supplémentaires assez discrètes ($\mathbf{S_3}$ et $S_4$), visibles uniquement après coloration au nitrate d'argent et qui présentent une taille approximative de 380 pb ($\mathbf{S_3}$) chez l'arbre 107 et de 360 pb ($S_4$) chez l'arbre 3076. Ces fragments pourraient correspondre à une nouvelle classe d'espaceurs très courts (310-330

pb) irrégulièrement détectée sur gel d'agarose (voir individu 91, **Figure 57** et individus 3, 10 et 11, **Figure 41**) et donc probablement peu représentée dans le génome. Ils ont été inclus dans l'analyse de la ségrégation.

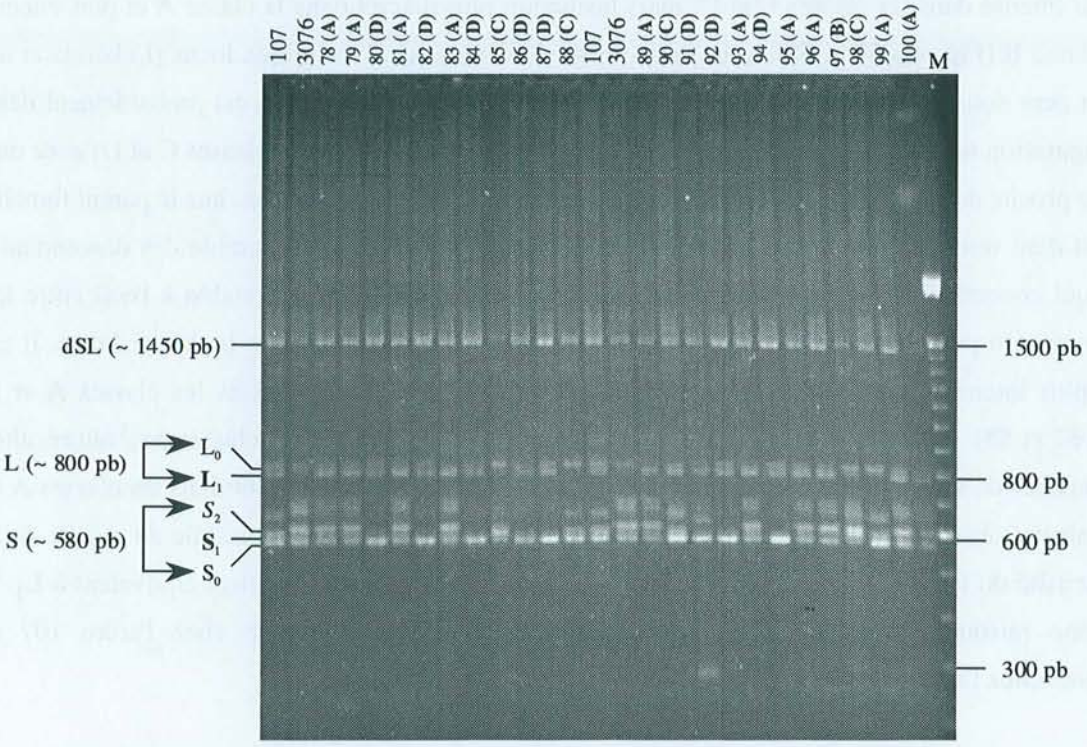



**Figure 57**. Profils d'amplification des espaceurs de l'ADNr 5S (couple d'amorces P1/P2) obtenus pour l'arbre 107 (mélèze d'Europe), l'arbre 3076 (mélèze du Japon) et un échantillon de leurs descendants hybrides F1 (individus 78-100 de la famille 179) issus du croisement contrôlé 107 x 3076. A, B, C et D : les différents profils observés dans la famille 179 (voir aussi **Figure 58**). M : marqueur de taille moléculaire «Ladder 100 pb».

β- Phénotypes des hybrides F1 et ségrégation des fragments parentaux

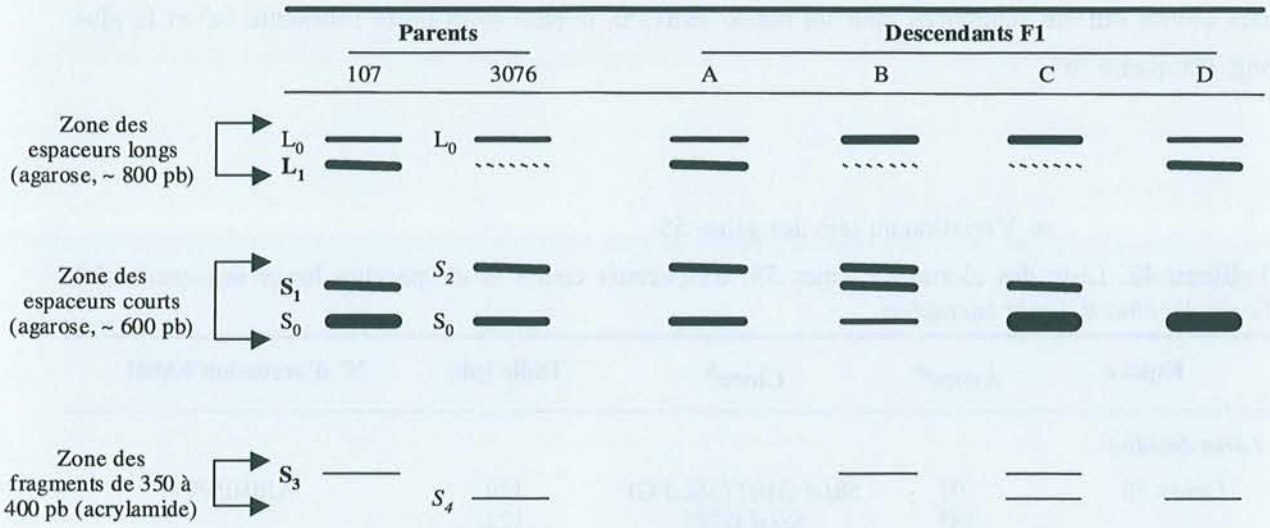
Le phénotype de chacun des membres de la famille 179 est indiqué dans la **Figure 58**. Quatre classes phénotypiques caractérisées par la présence des fragments parentaux $S_0$, $S_2$, $L_0$, $\mathbf{L_1}$ (classe A, 37 individus), $S_0$, $\mathbf{S_1}$, $S_2$, $\mathbf{S_3}$, $L_0$ (classe B, 18 individus), $S_0$, $\mathbf{S_1}$, $\mathbf{S_3}$, $S_4$, $L_0$ (classe C, 23 individus) ou $S_0$, $S_4$, $L_0$, $\mathbf{L_1}$ (classe D, 22 individus) sont distinguées. La ségrégation mendélienne selon une fréquence observée ($F_{obs}$) proche de 50 % de chacun des 5 fragments polymorphes,

- $\mathbf{S_1}$ [$F_{obs}$ = 41 % ; $\chi^2_{(1ddl)}$ = 3,24 ; $p > 0,05$],
- $S_2$ [$F_{obs}$ = 55 % ; $\chi^2_{(1ddl)}$ = 1,00 ; $p > 0,1$],
- $\mathbf{S_3}$ [$F_{obs}$ = 41 % ; $\chi^2_{(1ddl)}$ = 3,24 ; $p > 0,05$],
- $S_4$ [$F_{obs}$ = 45 % ; $\chi^2_{(1ddl)}$ = 1,00 ; $p > 0,1$],
- $\mathbf{L_1}$ [$F_{obs}$ = 59 % ; $\chi^2_{(1ddl)}$ = 3,24 ; $p > 0,05$],

révèle clairement leur configuration hétérozygote. Le cas des fragments $S_0$ et $L_0$ est en revanche plus difficile à analyser car ils sont présents chez les deux parents et l'ensemble des descendants. Quelques différences d'intensité reproductibles ont toutefois été notées pour $S_0$ qui est généralement bien marqué chez le parent 107 et plus faible chez le parent 3076 (**Figures 57** et **58**). Ce fragment $S_0$ est de même très intense dans les classes C et D, mais beaucoup plus discret dans la classe A et plus encore dans la classe B (**Figures 57** et **58**). Sous l'hypothèse de l'existence d'un unique locus (Lubaretz *et al.* 1996), on peut donc en déduire que chez le parent mâle 3076, le fragment $S_0$ est probablement dans une configuration hétérozygote et ségrège dans la descendance (individus des classes C et D) avec une fréquence proche de 50 % [$F_{obs}$ = 45 % ; $\chi^2_{(1ddl)} = 1{,}00$ ; $p > 0{,}1$], tandis que chez le parent femelle 107, il est dans une configuration homozygote et se retrouve donc dans l'ensemble des descendants. Pour ce qui concerne le fragment $L_0$, aucune différence d'intensité n'est détectable à l'oeil entre les deux parents. On peut néanmoins observer des différences reproductibles dans la descendance. Il est en effet plus intense dans les classes B et C (41 % des descendants) que dans les classes A et D (**Figures 57** et **58**). La faible intensité de ce fragment dans ces deux dernières classes pourrait résulter de phénomènes de compétition au cours de la PCR avec le fragment $\mathbf{L_1}$ (présent dans les classes A et D, mais absent dans les classes B et C). Cette hypothèse apparaît cependant fragile au regard de la faible intensité du fragment $L_0$ chez l'arbre 3076 qui ne comporte pas de fragment équivalent à $\mathbf{L_1}$. Il paraît donc raisonnable d'admettre que le marqueur $L_0$ est hétérozygote chez l'arbre 107 et homozygote chez l'arbre 3076.

### χ- Génotypes proposés pour les membres de la famille 179

Pour ce qui concerne l'arbre 107, il y a disjonction du groupe de marqueurs liés $\mathbf{S_1S_3}L_0$ et du marqueur $\mathbf{L_1}$ (**Figure 58**) au cours de la méiose dans un rapport 1:1 [$\chi^2_{(1ddl)} = 3{,}24$ ; $p > 0{,}05$]. Si l'on tient compte de l'organisation alternée de la plupart des unités courtes et longues chez les mélèzes (**Parties II-A** et **II-B** de ce **Chapitre IV**) et de la configuration homozygote ou hétérozygote des fragments $S_0$ et $L_0$, ce résultat est compatible avec l'existence d'un unique locus constitué des deux séries alléliques composites ($S_0\mathbf{S_1S_3}L_0$) et ($S_0\mathbf{L_1}$). Dans le cas de l'arbre 3076, la disjonction du marqueur $S_2$ et du groupe de marqueurs liés $S_0S_4$ dans les proportions 1:1 [$\chi^2_{(1ddl)} = 1{,}00$ ; $p > 0{,}1$] et la configuration homozygote du marqueur $L_0$ conduisent également à une interprétation génétique basée sur un seul locus composé des deux séries alléliques composites ($S_2L_0$) et ($S_0S_4L_0$). Les génotypes proposés au locus d'ADNr 5S pour les arbres 107 [($S_0\mathbf{S_1S_3}L_0$)/($S_0\mathbf{L_1}$)] et 3076 [($S_2L_0$)/($S_0S_4L_0$)] sont donc globalement hétérozygotes.




| | A | B | C | D |
|---|---|---|---|---|
| Effectif observé | 37 | 18 | 23 | 22 |
| Numéro des 100 individus hybrides F1 | 1 30 60 83<br>8 32 61 89<br>11 35 64 93<br>13 41 66 95<br>17 45 69 96<br>19 47 71 99<br>20 48 73 100<br>21 52 78<br>22 57 79<br>28 59 81 | 2 36<br>4 43<br>7 53<br>12 58<br>14 62<br>15 67<br>16 77<br>18 97<br>26<br>33 | 3 46 88<br>5 49 90<br>9 50 98<br>10 54<br>24 56<br>25 68<br>31 72<br>37 75<br>39 76<br>40 85 | 6 63 92<br>23 65 94<br>27 70<br>29 74<br>34 80<br>38 82<br>42 84<br>44 86<br>51 87<br>55 91 |

**Figure 58.** Phénotypes (A-D) observés sur gel d'agarose (zone des monomères, 550-850 pb) et sur gel d'acrylamide (zone des fragments de 350-400 pb) après amplification des espaceurs de l'ADNr 5S (amorces universelles P1 et P2) dans une population F1 de 100 descendants (famille 179) issue d'un croisement contrôlé entre le mélèze d'Europe 107 (parent femelle) et le mélèze du Japon 3076 (parent mâle). L'intensité des différents fragments amplifiés chez les parents et les descendants ($S_0$, $\mathbf{S_1}$, $S_2$, $\mathbf{S_3}$, $S_4$, $L_0$ et $\mathbf{L_1}$) est figurée par l'épaisseur variable des traits.

La combinaison lors du croisement 107 X 3076 de ces 4 séries alléliques *cis* qui constituent 4 allèles *trans* codominants, produit les classes génotypiques $[(S_0\mathbf{L_1})/(S_2L_0)]$, $[(S_0\mathbf{S_1}\mathbf{S_3}L_0)/(S_2L_0)]$, $[(S_0\mathbf{S_1}\mathbf{S_3}L_0)/(S_0S_4L_0)]$, et $[(S_0\mathbf{L_1})/(S_0S_4L_0)]$ qui correspondent respectivement aux 4 classes phénotypiques A, B, C et D détectées dans la famille 179 (**Figure 58**). Malgré un $\chi^2$ un peu élevé révélateur d'une légère distorsion de ségrégation [$\chi^2$ (3 ddl) = 8,24 ; 0,025 < p < 0,05], on peut admettre que ces 4 classes génotypiques sont produites dans un rapport 1:1:1:1 (37:18:23:22) confirmant un déterminisme génétique simple gouverné par un unique locus d'ADNr 5S (nommé 5SA) chez le mélèze d'Europe 107 et le mélèze du Japon 3076.

b- Diversité nucléotidique et divergence au sein des unités d'ADNr 5S

Nous avons analysé les séquences de 6 clones de gènes 5S, de 13 clones d'espaceurs courts et de 7 clones d'espaceurs longs pour étudier les variations nucléotidiques des unités d'ADNr 5S à la fois chez les mélèzes d'Europe (clones 5S*Ld*) et du Japon (clones 5S*Lk*). Les clones sont listés dans le **Tableau 43**, y compris les 6 clones étudiés dans la **Partie II-A** de ce **Chapitre IV**. Dans le cas où

deux clones ont été séquencés pour un même individu, le plus court porte l'étiquette "a" et le plus long, l'étiquette "b".

α- Variation au sein des gènes 5S

**Tableau 43**. Liste des clones de gènes 5S, d'espaceurs courts et d'espaceurs longs séquencés chez *Larix decidua* et *Larix kaempferi*.

| Espèce | Arbre[a] | Clone[b] | Taille (pb) | N° d'accession EMBL |
|---|---|---|---|---|
| ***Larix decidua*** | | | | |
| Gènes 5S | 107 | 5S*Ld*-G107 (5SLd-G) | 120 | AJ010579 |
| | 285 | 5S*Ld*-G285 | 122 | |
| | 475 | 5S*Ld*-G475 | 122 | |
| Espaceurs courts | 107 | 5S*Ld*-S107a (5SLd-S2) | 517 | AJ010575 |
| | | 5S*Ld*-S107b (5SLd-S3) | 539 | AJ010576 |
| | 285 | 5S*Ld*-S285a | 516 | |
| | | 5S*Ld*-S285b | 543 | |
| | 475 | 5S*Ld*-S475a | 516 | |
| | | 5S*Ld*-S475b | 522 | |
| | 92-1 | 5S*Ld*-S92.1 (5SLd-S1) | 514 | AJ010574 |
| Espaceurs longs | 107 | 5S*Ld*-L107a (5SLd-L1) | 738 | AJ010577 |
| | | 5S*Ld*-L107b (5SLd-L2) | 767 | AJ010578 |
| | 285 | 5S*Ld*-L285 | 771 | |
| | 475 | 5S*Ld*-L475 | 766 | |
| ***Larix kaempferi*** | | | | |
| Gènes 5S | 3076 | 5S*Lk*-G3076 | 120 | |
| | 3211 | 5S*Lk*-G3211 | 120 | |
| | 3247 | 5S*Lk*-G3247 | 120 | |
| Espaceurs courts | 3076 | 5S*Lk*-S3076a | 519 | |
| | | 5S*Lk*-S3076b | 531 | |
| | 3211 | 5S*Lk*-S3211a | 516 | |
| | | 5S*Lk*-S3211b | 546 | |
| | 3247 | 5S*Lk*-S3247a | 516 | |
| | | 5S*Lk*-S3247b | 545 | |
| Espaceurs longs | 3076 | 5S*Lk*-L3076 | 763 | |
| | 3211 | 5S*Lk*-L3211 | 741 | |
| | 3247 | 5S*Lk*-L3247 | 755 | |

[a] La provenance de chaque arbre est donnée dans la partie Matériel et Méthodes

[b] La nomenclature utilisée préalablement pour certains clones (**Partie II-A** du **Chapitre IV**) est indiquée entre parenthèses

L'alignement des séquences des 3 clones de mélèze d'Europe (5S*Ld*-G107, 5S*Ld*-G285, 5S*Ld*-G475) et des 3 clones de mélèze du Japon (5S*Lk*-G3076, 5S*Lk*-G3211, 5S*Lk*-G3247) est présenté dans la **Figure 59**. Aucune différence nucléotidique ne permet de distinguer les deux espèces et la séquence consensus majoritaire de 120 pb est comparable à celles des autres gymnospermes connues (consulter la **Partie II** de ce **Chapitre IV**). Les deux clones 5S*Lk*-3076 et 5S*Lk*-3247 (mélèzes du Japon) sont identiques à la séquence consensus et représentent donc probablement des gènes fonctionnels.

Deux autres clones ne se distinguent que par une transversion (5S*Lk*-3211, position 6 de l'alignement) ou deux transitions (5S*Ld*-G107, positions 58 et 77). La transition en position 77 est apparemment sans conséquence pour la structure et la fonction des ARNr 5S transcrits (**Figures 17 et 59**). De même, la transition en position 58 est sans doute neutre elle aussi car, bien que située au sein de la boîte A, elle renforce la cohésion de la région bicaténaire II (appariement $T_{20}$-$A_{58}$ supplémentaire). Enfin, la transversion en position 6 introduit uniquement un mésappariement ($A_6$-$G_{115}$) dans la région bicaténaire I qui comporte par ailleurs 7 bases complémentaires (comparer les positions 1-9 et 112-120 de l'alignement). Il paraît donc raisonnable de considérer que les clones 5S*Ld*-G107 et 5S*Lk*-3211 sont aussi des gènes 5S fonctionnels. En revanche, les clones 5S*Ld*-G285 et 5S*Ld*-G475 (mélèze d'Europe) sont très probablement des pseudogènes. Etrangement, ces deux clones issus d'arbres de provenances éloignées (Sudètes polonaises et Alpes autrichiennes) présentent une très grande similarité de séquence (une seule transition en position 121) mais ils diffèrent significativement en taille (122 pb) et en séquence (7 positions variables) des 4 autres gènes 5S séquencés. Cinq transitions (positions 16, 22, 69, 74 et 109), 2 transversions (positions 52 et 67) et une insertion de 2 pb (GT en position 80-81) sont en effet observées par rapport à la séquence consensus. Seule la transition en position 22 semble être neutre. Les autres mutations ponctuelles introduisent des mésappariements dans les régions bicaténaires II (position 16), III (position 52) et V (positions 67, 69 et 109) et affectent, soit la boîte A (position 52), soit les zones d'interaction des ARNr 5S avec le facteur de transcription TFIIIa (positions 16, 67, 69, 74 et 109). Par ailleurs, l'insertion de 2 pb en position 80-81 n'est sans doute pas dépourvue d'effets dommageables au niveau fonctionnel. Les expériences de mutagénèse dirigée *in vitro* (Pieler *et al.* 1985, 1987) ont en effet montré que les variations de l'espacement entre les boîtes A et C réduisent considérablement l'activité de transcription par la Pol III (voir la **Partie I-1-a-α** du **Chapitre II**).

En raison de la présence de ces deux pseudogènes, la diversité nucléotidique du gène 5S apparaît particulièrement élevée pour le mélèze d'Europe ($\pi = 0{,}066 \pm 0{,}004$) et très faible pour le mélèze du Japon ($\pi = 0{,}006 \pm 0{,}000$). De même, l'estimation globale de la diversité nucléotidique pour l'ensemble des séquences (calcul justifié par l'absence de mutations fixées entre les deux espèces dans cette région) reste substantielle ($\pi = 0{,}050 \pm 0{,}004$). Elle est toutefois plus faible ($\pi = 0{,}012 \pm 0{,}001$) si les pseudogènes 5S*Ld*-G285 et 5S*Ld*-G475 sont exclus du calcul.

β- Variation au sein des espaceurs

*- Alignements et analyses phylogénétiques*

Au contraire des clones de la sous-classe SA, les clones les plus longs, étiquettés "b" (522-546 pb), et le clone 5S*Ld*-S92.1 (514 pb) ne forment pas un ensemble bien défini. Aucun polymorphisme commun n'est en effet détecté entre ces 7 séquences. Trois d'entre elles toutefois (5S*Ld*-S107b, 5S*Lk*-S3076b et 5S*Ld*-S285b) partagent 8 polymorphismes aux positions 135 (G), 191 (C), 255 (G), 353 (G), 436 (C), 441 (A), 359 (T) et 515 (A), ce qui permet de considérer qu'elles sont

apparentées. Il en va de même des séquences de mélèze du Japon 5S*Lk*-S3211b (Mont Fuji) et 5S*Lk*-S3247b (Okunikko) qui partagent 8 mutations ponctuelles aux positions 105 (C), 115 (C), 206 (T), 216 (G), 343 (C), 398 (C), 418 (A), 420 (G), 425 (C) et une insertion de 7 pb (CTTCTTC, positions 402-408).

```
                                                                                  Boîte A
Consensus   GGGTGCGATC ATACCAGCGT TAATGCACCG GATCCCATCA GAACTCCGCA GTTAAGCGCG C
                                                  ****                  *
                I           II          III                   III          II

5SLd-G107   .......... .......... .......... .......... .......... .......A.. .
5SLd-G285   .......... .....G.... .G........ .......... .......... .G........ .
5SLd-G475   .......... .....G.... .G........ .......... .......... .G........ .
5SLk-G3076  .......... .......... .......... .......... .......... .......... .
5SLk-G3211  .....A.... .......... .......... .......... .......... .......... .
5SLk-G3247  .......... .......... .......... .......... .......... .......... .

                  M                  Boîte C
Consensus   TTGGGCTAGA GTAGTACT-- GGGATGGGTG ACCTCCCGGG AAGTCCTAGT GTTGCACCCT C
                                      ****                **
                II   V              IV                IV          V          I

5SLd-G107   .......... .....G..-- .......... .......... .......... .......... .
5SLd-G285   .....A.G.. ..G.....GT .......... .......... .......G.. .........C .
5SLd-G475   .....A.G.. ..G.....GT .......... .......... .......G.. .......... .
5SLk-G3076  .......... .........-- .......... .......... .......... .......... .
5SLk-G3211  .......... .........-- .......... .......... .......... .......... .
5SLk-G3247  .......... .........-- .......... .......... .......... .......... .
```

**Figure 59**. Comparaison des séquences nucléotidiques des clones de gènes 5S obtenus à partir de plusieurs individus de *Larix decidua* (5S*Ld*-G) et *Larix kaempferi* (5S*Lk*-G). Les bases variables par rapport à la séquence consensus majoritaire sont indiquées selon les conventions IUPAC standard. Les séquences fonctionnelles (boîte A, élément intermédiaire M et boîte C) de la région de contrôle interne de la transcription des gènes 5S (ICR) sont encadrées. Les motifs complémentaires impliqués dans la formation des structures secondaires (régions bicaténaires I-V) et tertiaire des ARNr 5S chez les plantes supérieures (Barciszewska *et al.* 1994) sont respectivement soulignées ou indiquées par des astérisques. Les zones d'interaction entre les ARNr 5S et le facteur de transcription TFIIIA apparaissent en gras.

L'alignement des 7 clones d'espaceurs longs (classe L), de tailles comprises entre 738 pb et 771 pb, est présenté dans la **Figure 62**. Comme dans la classe d'espaceurs courts, aucune mutation fixée n'apparaît dans l'une ou l'autre des deux espèces. En revanche, les taux de transitions/transversions y sont comparativement très faibles (1,55 pour *L. decidua* et 1,28 pour *L. kaempferi*). L'analyse phylogénétique (**Figure 63**) conduit à la mise en évidence d'un groupe constitué des trois séquences les plus courtes 5S*Ld*-L107a (738 pb), 5S*Lk*-L3211 (741 pb) et 5S*Lk*-L3247 (755 pb) qui présentent 5 polymorphismes nucléotidiques communs aux positions 488 (C), 523 (T), 591 (T), 593 (A) et 683 (C). Les deux séquences 5S*Ld*-L107a et 5S*Lk*-L3211 partagent en outre 6 polymorphismes supplémentaires. Le premier est une délétion de 15 pb (positions 412-426), sans doute également provoquée par un glissement de brins au cours de la réplication (présence d'un petit élément dupliqué GGC en amont des bordures).

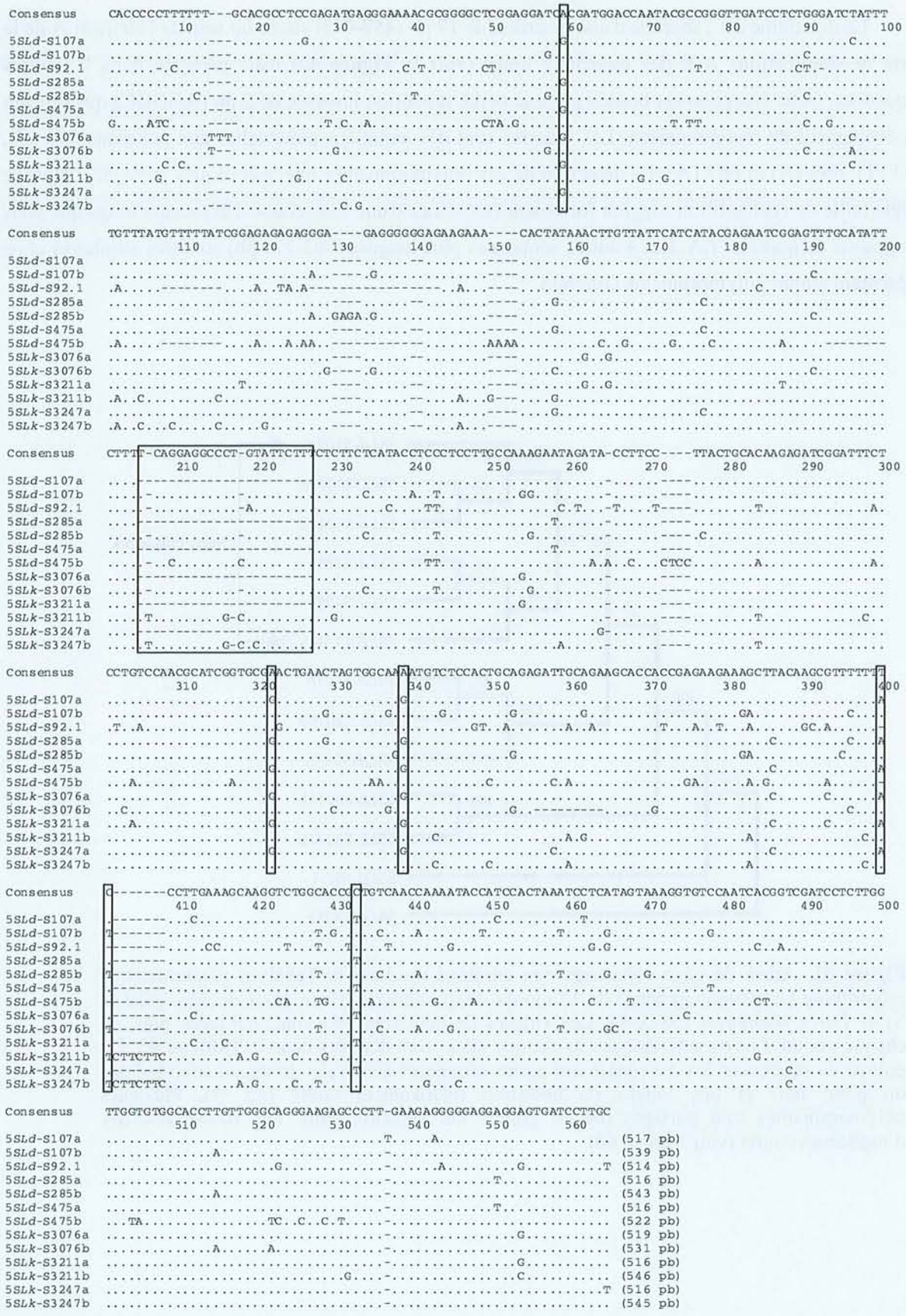



**Figure 60**. Alignement des clones d'espaceurs courts de mélèze d'Europe (5S*Ld*) et de mélèze du Japon (5S*Lk*).

Les bases variables par rapport à la séquence consensus majoritaire sont indiquées selon les conventions IUPAC standard. Les positions encadrées montrent les caractéristiques nucléotidiques communes aux espaceurs courts de la sous-classe SA (voir texte et **Figure 61**).

Le deuxième est l'absence d'une insertion de 17 pb (457-473) située au sein de l'élément A de la série de duplications A-E des espaceurs longs (voir la **Figure 42**) mais présente dans les autres séquences. Cette insertion est bordée par une petite répétition inversée de 5 pb (AGGGC) qui suggère un événement de recombinaison. Les 4 autres sont des mutations ponctuelles aux positions 644 (G), 653 (T), 659 (A) et 667 (A). La détection de ces polymorphismes interspécifiques partagés malgré la faible taille de l'échantillon suggère fortement l'existence d'une sous-classe d'espaceurs longs que nous proposons de nommer LA. Les 4 autres séquences plus longues (763-771 pb) sont très similaires et ne présentent aucun polymorphisme commun.

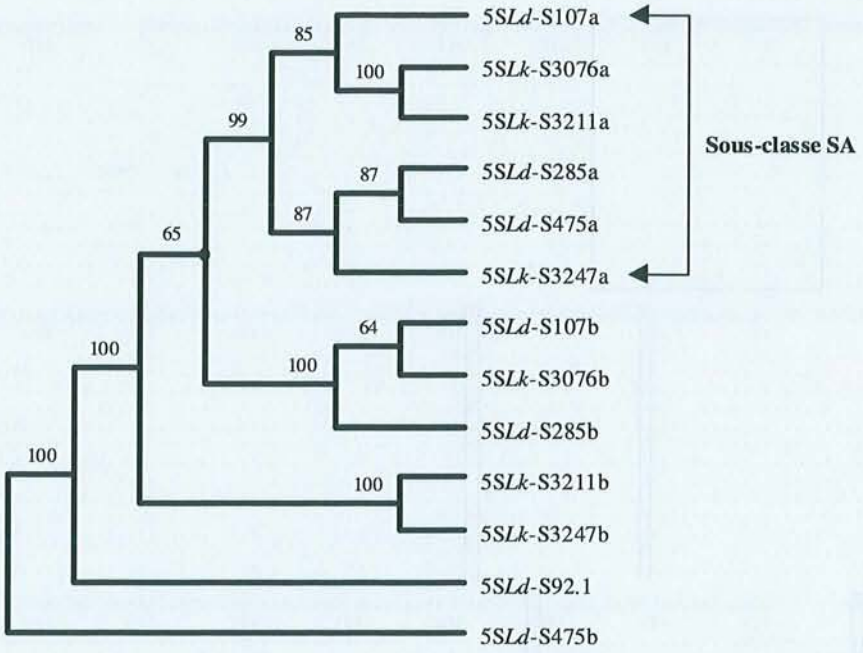



**Figure 61**. Arbre de gène consensus non enraciné résultant de l'analyse parsimonieuse (algorithme heuristique général) de 13 clones d'espaceurs courts de *Larix decidua* (5S*Ld*-S) et *Larix kaempferi* (5S*Lk*-S). Les valeurs de bootstrap (%) sont indiquées près de chaque noeud. Les caractéristiques de chaque clone sont données dans le **Tableau 43**. Le groupe de séquences SA est séparé des autres séquences au niveau du noeud souligné par un point noir et une valeur de bootstrap relativement faible (65 %). Plusieurs polymorphismes sont partagés par ce groupe qui constitue donc une sous-classe SA d'espaceurs courts (voir **Figure 60**).

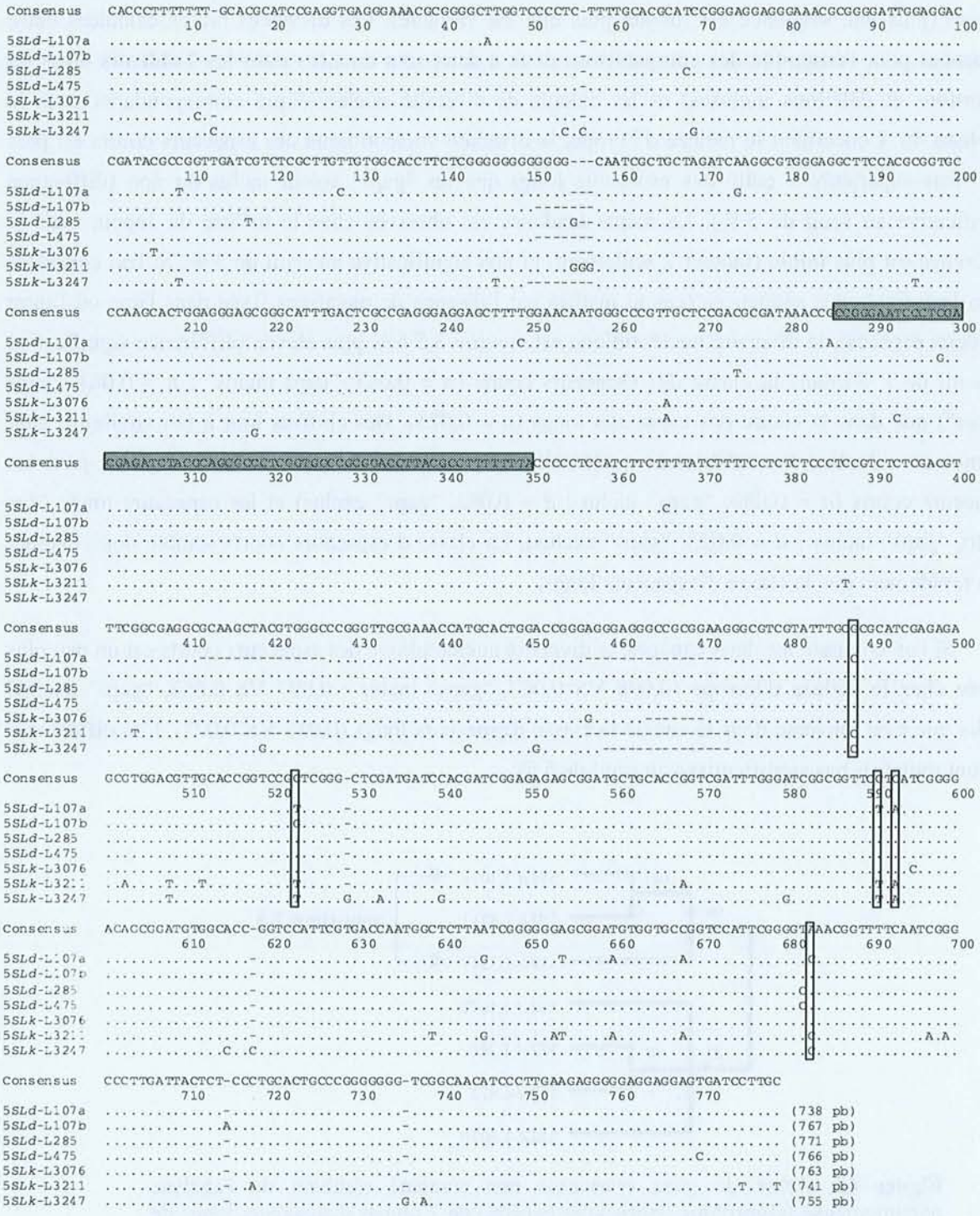



**Figure 62**. Alignement des clones d'espaceurs longs de mélèze d'Europe (5S*Ld*) et de mélèze du Japon (5S*Lk*).

Les bases variables par rapport à la séquence consensus majoritaire sont indiquées selon les conventions IUPAC standard. Les positions encadrées montrent les polymorphismes partagés par les espaceurs de la sous-classe LA (voir également la **Figure 63**). Le motif 5SmsCR1 conservé au sein des espaceurs d'autres gymnospermes est indiqué en gris.

La comparaison des **Figures 60** et **62** montre clairement que la classe d'espaceurs courts est beaucoup plus variable que la classe d'espaceurs longs alors que l'on pouvait s'attendre à la situation inverse (plus une séquence est longue plus elle est variable). Les distances brutes calculées entre séquences pour l'ensemble des comparaisons deux à deux sont données dans les **Tableaux 44** et **45** (insertions et délétions incluses) et les calculs de diversité nucléotidique correspondants dans le **Tableau 46**. Concernant le mélèze d'Europe, la diversité nucléotidique des espaceurs courts est plus de 6 fois supérieure à celle des espaceurs longs que les "gaps" soient inclus ou non (différence significative au seuil de 5 %). La même tendance est observée chez le mélèze du Japon, mais la différence est plus faible (facteur 2 seulement) et non significative au seuil de 5 %. Si l'on considère enfin l'ensemble des séquences (calcul justifié par l'absence de mutations fixée dans l'une ou l'autre des deux espèces), la diversité nucléotidique est environ 3,5 fois plus élevée (différence significative au seuil de 1 %) dans la classe des espaceurs courts ($\pi$ = 0,083, "gaps inclus" ; $\pi$ = 0,081, "gaps exclus") que dans la classe des espaceurs longs ($\pi$ = 0,025). Des chiffres tout à fait similaires sont obtenus en calculant les différences nucléotidiques moyennes (*d*) entre ces deux espèces pour les espaceurs courts (*d* = 0,080, "gaps" inclus ; *d* = 0,082, "gaps" exclus) et les espaceurs longs (*d* = 0,026, "gaps" inclus ; *d* = 0,023, "gaps" exclus). La classe d'espaceurs courts semble donc évoluer plus rapidement que la classe d'espaceurs longs.

Si l'on compare les deux espèces, la diversité nucléotidique des espaceurs courts est un peu plus élevée chez le mélèze d'Europe (0,098 VS 0,067, "gaps" inclus ; 0,097 VS 0,063, "gaps" exclus) tandis que c'est l'inverse dans le cas de la classe d'espaceurs longs (0,015 VS 0,037). Ces différences ne sont toutefois pas significatives au seuil de 5 %.

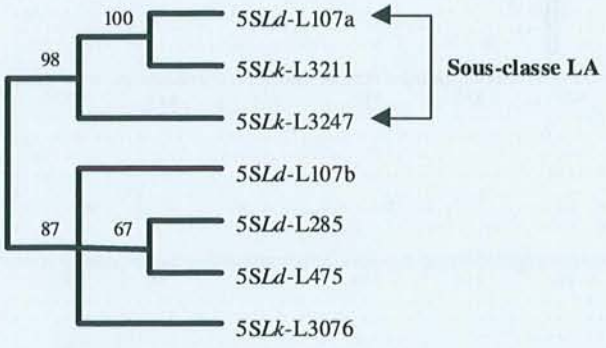



**Figure 63**. Arbre de gène consensus non enraciné résultant de l'analyse parsimonieuse (algorithme heuristique général) de 7 clones d'espaceurs longs de *Larix decidua* (5S*Ld*-L) et *Larix kaempferi* (5S*Lk*-L). Les valeurs de bootstrap (%) sont indiquées près de chaque noeud. Les caractéristiques de chaque clone sont données dans le **Tableau 43**. Les trois séquences de la sous-classe LA présentent plusieurs polymorphismes communs (voir **Figure 62**).

**Tableau 44**. Distances brutes calculées entre les 13 clones d'espaceurs courts, "gaps" inclus (sous la diagonale) ou "gaps" exclus (au-dessus de la diagonale)*.

| Clone | N° | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5S*Ld*-S107a | 1 | - | 0,078 | 0,132 | 0,023 | 0,070 | 0,025 | 0,147 | 0,021 | 0,075 | 0,029 | 0,074 | 0,027 | 0,068 |
| 5S*Ld*-S107b | 2 | 0,074 | - | 0,138 | 0,074 | 0,030 | 0,074 | 0,152 | 0,079 | 0,038 | 0,087 | 0,093 | 0,078 | 0,089 |
| 5S*Ld*-S92.1 | 3 | 0,133 | 0,141 | - | 0,128 | 0,130 | 0,130 | 0,154 | 0,132 | 0,135 | 0,140 | 0,132 | 0,128 | 0,126 |
| 5S*Ld*-S285a | 4 | 0,025 | 0,069 | 0,128 | - | 0,066 | 0,002 | 0,143 | 0,025 | 0,071 | 0,033 | 0,066 | 0,012 | 0,064 |
| 5S*Ld*-S285b | 5 | 0,067 | 0,032 | 0,135 | 0,063 | - | 0,068 | 0,145 | 0,076 | 0,040 | 0,083 | 0,089 | 0,070 | 0,082 |
| 5S*Ld*-S475a | 6 | 0,027 | 0,069 | 0,131 | 0,002 | 0,065 | - | 0,145 | 0,027 | 0,073 | 0,035 | 0,068 | 0,014 | 0,066 |
| 5S*Ld*-S475b | 7 | 0,143 | 0,154 | 0,156 | 0,139 | 0,147 | 0,141 | - | 0,147 | 0,155 | 0,149 | 0,148 | 0,143 | 0,141 |
| 5S*Lk*-S3076a | 8 | 0,023 | 0,076 | 0,133 | 0,027 | 0,074 | 0,029 | 0,143 | - | 0,081 | 0,012 | 0,078 | 0,029 | 0,074 |
| 5S*Lk*-S3076b | 9 | 0,072 | 0,038 | 0,141 | 0,067 | 0,040 | 0,069 | 0,156 | 0,078 | - | 0,089 | 0,095 | 0,073 | 0,091 |
| 5S*Lk*-S3211a | 10 | 0,032 | 0,084 | 0,141 | 0,036 | 0,082 | 0,038 | 0,145 | 0,013 | 0,086 | - | 0,085 | 0,037 | 0,081 |
| 5S*Lk*-S3211b | 11 | 0,067 | 0,091 | 0,133 | 0,059 | 0,084 | 0,061 | 0,143 | 0,072 | 0,093 | 0,080 | - | 0,070 | 0,040 |
| 5S*Lk*-S3247a | 12 | 0,027 | 0,072 | 0,126 | 0,011 | 0,065 | 0,013 | 0,141 | 0,029 | 0,069 | 0,038 | 0,061 | - | 0,064 |
| 5S*Lk*-S3247b | 13 | 0,063 | 0,082 | 0,126 | 0,059 | 0,076 | 0,061 | 0,137 | 0,069 | 0,086 | 0,078 | 0,038 | 0,057 | - |

* Distances calculées en utilisant l'option SHOWDIST du logiciel PAUP (Swofford 1993).

**Tableau 45**. Distances brutes calculées entre les 7 clones d'espaceurs longs, "gaps" inclus (sous la diagonale) ou "gaps" exclus (au-dessus de la diagonale)*.

| Clone | N° | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| 5S*Ld*-L107a | 1 | - | 0,023 | 0,027 | 0,025 | 0,026 | 0,031 | 0,031 |
| 5S*Ld*-L107b | 2 | 0,023 | - | 0,008 | 0,005 | 0,008 | 0,037 | 0,028 |
| 5S*Ld*-L285 | 3 | 0,027 | 0,008 | - | 0,005 | 0,01 | 0,041 | 0,032 |
| 5S*Ld*-L475 | 4 | 0,025 | 0,005 | 0,005 | - | 0,008 | 0,038 | 0,029 |
| 5S*Lk*-L3076 | 5 | 0,026 | 0,007 | 0,01 | 0,007 | - | 0,035 | 0,032 |
| 5S*Lk*-L3211 | 6 | 0,031 | 0,036 | 0,04 | 0,037 | 0,036 | - | 0,045 |
| 5S*Lk*-L3247 | 7 | 0,031 | 0,027 | 0,031 | 0,029 | 0,03 | 0,044 | - |

* Distances calculées en utilisant l'option SHOWDIST du logiciel PAUP (Swofford 1993).

**Tableau 46**. Diversité nucléotidique ($\pi$) au sein des classes d'espaceurs courts et longs pour le mélèze d'Europe (*Ld*), le mélèze du Japon (*Lk*) et l'ensemble des séquences examinées (*Ld* + *Lk*).

| | Classe d'espaceurs courts (S) | | | Classe d'espaceurs longs (L) | | | |
|---|---|---|---|---|---|---|---|
| | | $\pi^b$ | | | $\pi^b$ | | |
| Taxon | N[a] | Avec "gaps" | Sans"gaps" | N[a] | Avec "gaps" | Sans "gaps" | P[c] |
| *Ld* | 7 | 0,098 ± 0,034 | 0,097 ± 0,034 | 4 | 0,015 ± 0,008 | 0,015 ± 0,008 | < 0,05 |
| *Lk* | 6 | 0,067 ± 0,022 | 0,063 ± 0,021 | 3 | 0,037 ± 0,018 | 0,037 ± 0,018 | > 0,30 |
| *Ld* + *Lk* | 13 | 0,083 ± 0,019 | 0,081 ± 0,019 | 7 | 0,025 ± 0,009 | 0,025 ± 0,008 | < 0,01 |

[a] Nombre de séquences comparées

[b] Valeurs ± SE (erreur standard)

[c] Probabilité *P* (test t) que les diversités nucléotidiques observées dans la classe d'espaceurs courts et dans la classe d'espaceurs longs soient identiques (test t, Zar 1996, p 124).

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

## 4- Discussion

### a- Le locus 5SA et la composition hétérogène des séries d'unités d'ADNr 5S

L'étude du croisement entre le mélèze d'Europe 107 et le mélèze du Japon 3076 (famille 179) nous a permis, malgré une légère distorsion de ségrégation des marqueurs étudiés, de conclure à l'existence d'un unique locus d'ADNr 5S (5SA). Ce résultat confirme les conclusions de Lubaretz *et al.* (1996) concernant *Larix decidua* et les élargit indirectement à l'espèce voisine *Larix kaempferi*. Ces auteurs ont en effet détecté *in situ* dans le génome du mélèze d'Europe deux signaux d'hybridation d'une sonde hétérologue d'ADNr 5S obtenue à partir de *Vicia faba*. Ces deux signaux correspondent à un unique locus d'ADNr 5S (probablement 5SA) localisé en position subtélomérique sur les petits bras de la paire de grands chromosomes métacentriques III (10,18 µm) à une distance du centromère équivalente à 82,7 % de leur longueur. Chacun de ces chromosomes III porte par ailleurs sur son bras long l'un des 3 loci d'ADNr identifiés chez *Larix decidua*. Les données que nous avons obtenues devraient permettre d'obtenir rapidement la cartographie génétique du locus 5SA chez les mélèzes d'Europe et du Japon car cette même famille 179 a été utilisée récemment pour construire une carte génétique basée sur la ségrégation de 266 marqueurs RAPD et AFLP pour l'essentiel (Arcade *et al.* 2000). Cette information serait particulièrement utile car les marqueurs RAPD et AFLP sont dominants et donc non transposables d'une carte génétique à l'autre (spécifiques d'un individu). En revanche, l'utilisation des marqueurs codominants et multialléliques fournis par l'ADNr 5S pourrait, comme d'autres tels que les microsatellites (Echt et Nelson 1997, Brondani *et al.* 1998, Barrénèche *et al.* 1998) ou les séquences STS (Sequence-Tagged-Site, Paglia *et al.* 1998, Perry et Bousquet 1998), faciliter la construction d'une carte génétique unifiée entre les parents du croisement 107 x 3076 (marqueurs d'ancrage). L'ADNr 5S présente en outre l'avantage d'être très conservé (famille multigénique impliquée dans la synthèse protéique), ce qui favorise sa détection dans des génomes très différents, facilitant ainsi l'accès à la cartographie génétique comparée entre individus d'une même espèce (carte consensus) ou même entre espèces (synténie). La cartographie génétique de l'ADNr 5S pourrait également fournir des informations intéressantes pour la connaissance du caryotype des mélèzes. Elle permettrait notamment d'établir la correspondance entre la paire de chromosomes III et le groupe de liaison porteur du locus 5SA (probablement l'un des grands groupes de liaison E1, E2, E3 du mélèze d'Europe, ou J1, J2, J3 du mélèze du Japon, voir Arcade *et al.* 2000). Ce type de données serait particulièrement utile car, chez les mélèzes d'Europe et du Japon, la distinction cytologique des 12 paires de chromosomes est difficile. Des auteurs comme Hizume (1988), Hizume *et al.* (1993) et Nkongolo et Klimaszewska (1995) ont par exemple placé les constrictions secondaires associées à l'ADNr sur les bras de différents chromosomes métacentriques (chromosomes I, II, III et IV au sens de Lubaretz *et al.* 1996) en raison de problèmes évidents de reconnaisssance liés à leur taille similaire (9,63-11,68 µm). Si d'autres familles de séquences répétées localisées en partie sur ce même chromosome III (Lubaretz *et al.* 1996), comme l'ADNr, pouvaient être placées sur la carte génétique, il serait alors possible de passer de la distance génétique à la distance physique.

**A.**

```
                -80       -70       -60       -50       -40       -30       -20       -10
Consensus S  ACGGTCGATCCTCTTGGTTGGTGTGGCACCTTGTTGGGCAGGGAAGAGCCCTTGAAGAGGGGGAGGAGGAGTGATCCTTGCGG
Consensus L  CG..C.CT.GA.TACTCTCCC..CACTG..CG.GG...TC..C..C.T.................................

5SLd-G107    ACGgtcgacCCTCTTGGTTG...........A...............T.................................
5SLd-G285    ACGgtcgacCCTCTTGGTTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNccgcgg
5SLd-G475    ACGgtcgacCCTCTTGGTTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNccgcgg
5SLk-G3076   ACGgtcgacCCTCTTGGTTG.............................G....................C.............
5S1K-G3211   ACGgtcgacCCTCTTGGTTG.............................G....................C.............
5SLk-G3247   ACGgtcgacCCTCTTGGTTG.............................................................T.....
```

Région transcrite

*Sac* II

P3 (*Sal* I)

**B.**

```
                                α                                                  β
                 10        20        30        40        50        60        70        80        90       100       110       120       130
Consensus L  CACCC-TTTTTTT-GCACGCATCCGAGGTGAGGGAAA-CGCGGGGCTTGGTCCCCTCTTTTGCACGCATCCGGGAGGAGGGAAA-CGCGGGGATTGGA--GGACCGATACGCCGG-TTGATCGTCTCGCTTGTT
Consensus S  .....CC......-......C......A.........A.........C..A-------------------------------------...CACGAT.....A.........G......C...G.GA.C.A

5SLd-G107    .....-G.C....-...........GATGAGGGAAAccgcggGGC
5SLd-G285    .....-.......T.............A.......G.-G...........C.........G...........AGATGAGGGAAA-ccgcggGGC
5SLd-G475    .....-.......T.............A.......G.-G...........C.........G...........A..T......G.-G......C....CCCCTCTTAGGCACGCATCCGNGATGAGGGAAAccgcggGGCNNNN
5SLk-G3076   .....-.......-...........GATGAGGGAAAccgcggGGC
5SLk-G3211   .....-.......-...........GATGAGGGAAAccgcggGGC
5SLk-G3247   .....CC......-......C.....GATGAGGGAAAccgcggGGC
```

χ

P4 (*Sac* II)

P4 (*Sac* II)

P4 (*Sac II*)

**Figure 64**. Comparaison des zones situées immédiatement en amont (**A**) ou en aval (**B**) de la région transcrite (gène 5S ou pseudogène) de différents clones d'ADNr 5S, issus de l'amplification à l'aide des amorces P3 et P4, avec les zones correspondantes des séquences consensus des espaceurs courts (consensus S) ou longs (consensus L) des mélèzes d'Europe (clones 5S*Ld*) et du Japon (clones 5S*Lk*).
Les positions nucléotidiques sont numérotées par rapport à la région transcrite. Seules les bases qui diffèrent de la séquence consensus des espaceurs courts (**A**) ou longs (**B**) sont indiquées. Les flèches indiquent l'orientation et la position des amorces P3 et P4 dont la séquence figure en gras. Les sites de restriction *Sal* I (gtcgac) ou *Sac* II (ccgcgg) utilisés pour le clonage des fragments P3/P4 sont en minuscules. Noter le site *Sac* II supplémentaire (en gris) observé au sein des clones 5S*Ld*-G285 et 5S*Ld*-G475 qui recouvre l'extrémité 5' de la région transcrite. Les région encadrées α et β correspondent à la vaste duplication en tandem de l'extrémité 5' des espaceurs longs (voir Trontin *et al.* 1999). L'élément χ séquencé pour le clone 5S*Ld*-G475 correspond à une copie supplémentaire de ce motif répété. N : base non séquencée.
Les zones en amont de la région transcrite (**A**) sont clairement apparentées aux espaceurs courts car l'amorce P3 est spécifique de ces derniers. Les zones en aval de la région transcrite (**B**) sont apparentées aux espaceurs longs dans le cas des 5 premières séquences, confirmant ainsi l'organisation principalement alternée des unités d'ADNr 5S chez ces deux espèces. La dernière séquence (5S*Lk*-G3247) présente toutefois les caractéristiques des espaceurs courts (C en positions 6, 7 et 21) et démontre ainsi l'existence d'unités courtes répétées en tandem chez le mélèze du Japon.

Il est tout à fait évident que des unités courtes alternent avec des unités longues au locus 5SA (séries composites). En revanche, la question se pose de savoir si ce locus contient également les séries homogènes d'unités longues révélées précédemment (**Partie II-A** de ce **Chapitre IV**) ou si celles-ci sont situées à un autre locus. L'existence d'un second locus d'ADNr 5S ne peut pas être complètement exclue sur la seule base de l'analyse du croisement 107 X 3076, surtout s'il est de petite taille et/ou si les allèles *cis* et *trans* qui le composent sont les mêmes qu'au locus 5SA. La détection par PCR de marqueurs spécifiques de chaque série d'un locus est en effet difficile. Pour identifier un marqueur spécifique ($S_4$) de l'une des quatre séries parentales (série $S_0S_4L_0$ de l'arbre 3076) au locus principal 5SA, nous avons ainsi dû avoir recours à l'électrophorèse sur gel d'acrylamide avec coloration au nitrate d'argent. Il est donc possible qu'un second locus de petite taille et/ou composé des mêmes allèles *cis* et *trans* qu'au locus 5SA puisse avoir échappé à l'analyse. Dans une telle situation, les unités situées à des loci différents devraient être très divergentes car les événements d'homogénéisation entre loci par les mécanismes de l'évolution concertée sont considérés comme très rares (Scoles *et al.* 1988, Reddy et Appels 1989, Appels *et al.* 1992, Sastri *et al.* 1992, Kellogg et Appels 1995, Cronn *et al.* 1996, voir aussi les **Parties III-4-a** et **III-3-a** du **Chapitre II**). Or, malgré l'identification d'une sous-classe LA, les divergences de séquences observées au sein de la classe d'unités longues paraissent trop faibles pour appuyer cette hypothèse. Dans l'état actuel de nos connaissances, il est donc plus raisonnable de considérer que les séries homogènes d'unités longues se trouvent au locus 5SA. Elles pourraient former des domaines homogènes au sein des séries composites du locus 5SA comme c'est le cas pour de nombreux satellites (voir la **Partie II-2-a-β** du **Chapitre I**). Copenhaver et Pikaard (1996) ont ainsi récemment montré qu'au sein des séries de l'ADNr d'*Arabidopsis thaliana*, les hétéroallèles spécifiques d'une série à un locus donné ne sont pas dispersés, mais forment au contraire des sous-séries homogènes d'unités adjacentes.

Dans les deux espèces étudiées, les séries alléliques au locus 5SA sont clairement hétérogènes. Au-delà de l'alternance des unités courtes et longues qui constitue l'originalité de l'ADNr 5S des mélèzes, nous avons pu en particulier détecter des différences de taille importantes au niveau des espaceurs courts des séries composites $S_0\mathbf{S_1}\mathbf{S_3}L_0$ de l'arbre 107 et $S_0S_4L_0$ de l'arbre 3076. La grande différence de taille des marqueurs $\mathbf{S_3}$ et $S_4$ par rapport aux marqueurs $S_0$ et $\mathbf{S_1}$ pourraient même révéler l'existence d'une nouvelle classe d'unités courtes de 480-500 pb faiblement représentée. Comme nous l'avions constaté dans le cas du chêne pédonculé (voir la **Partie III** du **Chapitre III**), les mécanismes de l'évolution concertée probablement responsables de l'amplification, avant la diversification du genre *Larix* (voir la **Partie IIB** de ce **Chapitre IV**), de l'unité répétée dimérique associant unités courtes et longues, n'ont donc pas encore permis d'homogénéiser complètement les séries de l'ADNr 5S des mélèzes d'Europe et du Japon. La détection de marqueurs spécifiques de chaque série au locus 5SA montre de plus, indirectement, que les échanges interchromosomiques entre séries par crossing-over inégal ou conversion génique sont très limités, ou de moindre fréquence que les échanges intrachromosomiques (intramoléculaire ou entre chromatides soeurs). C'est généralement cette situation qui a été décrite par différents auteurs étudiant l'évolution de l'ADNr 5S

chez les plantes supérieures (Kellogg et Appels 1995, Cronn *et al.* 1996). Le modèle d'évolution concertée intralocus des familles de séquences répétées en tandem proposé par Liao *et al.* (1997) se trouve donc ainsi validé. Il est compatible avec le mode d'hérédité mendélien classique des unités d'ADNr 5S qui en fait d'excellents marqueurs cytogénétiques.

b- La présence de pseudogènes au sein des séries de l'ADNr 5S

Les diversités nucléotidiques du gène 5S (indels incluses) chez le mélèze d'Europe ($\pi = 0{,}066$) et le mélèze du Japon ($\pi = 0{,}006$) sont similaires aux résultats obtenus par Kellogg et Appels (1995) chez 34 espèces diploïdes de *Triticeae* ($0{,}000 < \pi < 0{,}062$ ; en moyenne $\pi = 0{,}023$) et par Cronn *et al.* (1996) chez 16 espèces, diploïdes également, de *Gossypium* ($0{,}013 < \pi < 0{,}083$ ; en moyenne $\pi = 0{,}051$). La plus forte diversité nucléotidique observée chez le mélèze d'Europe est due à la présence probable de gènes 5S non fonctionnels (pseudogènes) au sein des séries de l'ADNr 5S. Deux des trois clones séquencés chez cette espèce (5S*Ld*-G285 et 5S*Ld*-G475) contiennent en effet des gènes 5S dont la séquence est perturbée par plusieurs mutations ponctuelles défavorables au maintien de la structure secondaire des ARNr 5S, ou qui affectent les zones d'interaction avec la Pol III et les ARNr 5S. Ils contiennent en outre une insertion de 2 pb qui modifie la distance entre les boîtes A et C, ce qui pourrait entraîner une diminution importante de l'activité de transcription par la Pol III (Pieler *et al.* 1985, 1987). L'examen des régions non transcrites situées en amont et en aval de ces pseudogènes (régions situées entre les amorces P3 et P4, **Figure 64**) révèle également l'existence d'un site *Sac* II (CCGCGG) supplémentaire recouvrant leur extrémité 5' (positions - 4 à + 2). Ces deux séquences ont donc apparemment subi un clonage symétrique fortuit au niveau du site *Sac* II du plasmide pBluescript KS +/-, tandis que les 4 autres sont sans doute normalement issues du clonage asymétrique attendu en *Sal* I (site de restriction dans P3) et *Sac II* (site de restriction dans P4). Les séquences 5S*Ld*-G285 et 5S*Ld*-G475 sont donc tronquées. Comme l'amorce P3 est spécifique des espaceurs courts (site cible en positions - 81 à - 62 par rapport au gène, **Figure 64**) et que, d'autre part, les séquences en aval des gènes 5S présentent les caractéristiques des espaceurs longs (sauf le clone 5S*Lk*-G3247, **Figure 64**), on peut en déduire que les pseudogènes 5S*Ld*-G285 et 5S*Ld*-G475 proviennent de la série composite où alternent unités courtes et longues. Il faut souligner par ailleurs que ces deux pseudogènes qui proviennent d'arbres éloignés géographiquement (l'arbre 285 provient des Sudètes polonaises tandis que l'arbre 475 provient des Alpes autrichiennes) ont vraisemblablement une origine commune car ils ne diffèrent qu'à une seule position nucléotidique (position 121, **Figure 59**). On peut donc en déduire que les pseudogènes qui émergent au sein des séries composites de l'ADNr 5S des mélèzes sont maintenus durablement dans les populations, qu'ils s'accumulent au sein des séries de l'ADNr 5S et sont transmis au cours des générations successives. Notons que sur la base du critère de conservation des régions d'interaction avec le TFIIIA (boîtes A et C), la séquence 5S*Ld*-G107 correspondrait aussi à un pseudogène puisqu'elle comporte une mutation au sein de la boîte A. La fréquence des gènes 5S partiellement ou totalement inactifs semble donc très importante chez le mélèze d'Europe. Des conclusions similaires ont été rapportées dans le cas de *Linum usitatissimum* (Schneeberger et Cullis 1992), des *Triticeae* (Kellogg et Appels 1995) et du

genre *Gossypium* (Cronn *et al.* 1996). Bien qu'aucune étude fonctionnelle ou structurale n'ait été réalisée directement sur le pool transcriptionnel des ARNr 5S, les analyses de séquences de gènes 5S conduites chez ces différentes espèces suggèrent en effet la présence de nombreux pseudogènes au sein des séries d'ADNr 5S. Kellogg et Appels (1995) ont par exemple déterminé que plus de 17 % des 152 gènes 5S de *Triticeae* examinés (34 espèces) sont apparemment des pseudogènes comportant des mutations dans les boîtes A et/ou C. Ce chiffre est d'ailleurs probablement sous-estimé car lorsque les séquences correspondant à ces pseudogènes présumés sont écartées de l'analyse, la diversité nucléotidique reste remarquablement élevée et les nombreuses mutations observées se trouvent à des positions qui affectent les zones impliquées dans la structure secondaire des ARNr 5S ou dans les interactions gène 5S/ARNr 5S. Dans le cas de *Linum usitatissimum* cité précédemment, Schneeberger et Cullis (1992) ont expliqué la présence de nombreux pseudogènes répartis en plusieurs groupes par la forte dispersion de l'ADNr 5S sur les chromosomes, dispersion caractéristique de cette espèce. L'absence ou la faible homogénéisation interchromosomique par les mécanismes de l'évolution concertée (Nagylaki 1990, Liao *et al.* 1999) aurait conduit à une forte divergence des séries d'ADNr 5S et entrainé l'apparition massive de pseudogènes, certains petits loci étant même totalement constitués de ce type de séquences (vestiges d'anciens loci fonctionnels).

Cette explication n'est cependant pas applicable aux *Triticeae* (seulement deux loci d'unités courtes et longues sur les chromosomes 1 et 5, Scoles *et al.* 1988) et encore moins aux espèces diploïdes ne possédant qu'un seul locus d'ADNr 5S telles que certains *Gossypium* (Hanson *et al.* 1996) ou le mélèze d'Europe (Lubaretz *et al.* 1996). Une autre hypothèse a donc été avancée. Kellogg et Appels (1995) chez les *Triticeae*, puis Cronn *et al.* (1996) chez *Gossypium* ont en effet démontré que malgré le caractère très conservé des gènes 5S entre les espèces, la diversité nucléotidique intragénomique de cette région transcrite est bien souvent aussi élevée que celle de l'espaceur. Ils proposent de résoudre ce paradoxe en introduisant la notion d'un seuil critique d'accumulation des pseudogènes au-delà duquel l'efficacité fonctionnelle d'une série diminue fortement. Les séries d'ADNr 5S accumuleraient donc constamment des pseudogènes par dérive génétique, ces pseudogènes n'étant éliminés d'une population par la sélection naturelle que lorsque le nombre de gènes fonctionnels ne serait plus suffisant pour satisfaire les besoins cellulaires (voir **Figure 20**). La sélection ne s'exercerait donc pas individuellement sur les gènes mais plutôt sur l'ensemble d'une série. Dans le cas du mélèze d'Europe, il se pourrait ainsi que l'organisation alternée des unités courtes et longues (unités dimériques de 1450 pb) soit une conséquence directe de l'accumulation de pseudogènes, l'établissement d'unités dimériques incluant un gène fonctionnel et un pseudogène permettant alors de préserver l'intégrité fonctionnelle des séries. Il est en effet intrigant de constater que les 3 gènes 5S amplifiés à l'aide des amorces P3 (directe) et P4 (indirecte) ne sont probablement pas fonctionnels. L'amorce P3 étant spécifique des espaceurs courts (**Figure 64**), ces pseudogènes se trouvent nécessairement situés entre un espaceur court, côté 5', et un espaceur long côté 3' (gènes 5S "SL"). Il faudrait donc choisir un nouveau couple d'amorces spécifiques des espaceurs longs (directe) et courts (indirecte) afin d'amplifier les gènes 5S "LS" situés entre un espaceur court, côté 5', et un

espaceur long, côté 3'. Une étude comparée de la séquence et de la diversité nucléotidique des gènes 5S "SL" et "LS" permettrait de déterminer plus complètement la distribution des pseudogènes au sein des séries composites et d'en déduire un lien éventuel avec l'organisation principalement alternée des unités.

c- La faible diversité des espaceurs longs par rapport à celle des espaceurs courts

Les diversités nucléotidiques (indels incluses) calculées pour les espaceurs courts (S) ou longs (L) des mélèzes d'Europe ($\pi_S$ = 0,097-0,098 ; $\pi_L$ = 0,015) et du Japon ($\pi_S$ = 0,063-0,067 ; $\pi_L$ = 0,037) se situent dans la gamme observée pour les classes d'espaceurs courts ($0,000 < \pi_S < 0,112$ ; en moyenne $\pi_S = 0,040$) ou longs ($0,010 < \pi_L < 0,079$ ; en moyenne $\pi_L = 0,042$) des 34 espèces de *Triticeae* étudiées par Kellogg et Appels (1995) et pour l'unique classe d'espaceurs ($0,020 < \pi < 0,154$ ; en moyenne $\pi = 0,060$) des 16 espèces diploïdes de *Gossypium* analysées par Cronn *et al.* (1996). Elles sont également comparables aux différences nucléotidiques moyennes obtenues précédemment (voir la **Partie II** du **Chapitre III**) chez le chêne pédonculé ($d = 0,057$) et chez le chêne sessile ($d = 0,047$). La variabilité nucléotidique détectée au sein des espaceurs de l'ADNr 5S chez des espèces ligneuses à temps de génération long semble donc aussi importante que chez les espèces à cycle court.

La diversité nucléotidique au sein de la classe d'espaceurs longs est remarquablement faible par rapport à la classe d'espaceurs courts chez *Larix. decidua* ($\pi_S$ = 0,097-0,098, $\pi_L$ = 0,015) et *Larix kaempferi* ($\pi_S$ = 0,063-0,067, $\pi_L$ = 0,037). La différence est très significative (test t, $p < 0,01$) si l'on considère l'ensemble des espaceurs examinés pour les deux espèces comme un groupe ($\pi_S = 0,025$ ; $\pi_L$ = 0,081-0,083). Si l'on admet, d'une part, que l'importante similarité de séquence détectée entre les classes d'espaceurs courts et longs au niveau des bordures 5' (85 % de similarité sur environ 90 pb, **Figure 64B**) et 3' (81 % de similarité sur environ 50 pb, **Figure 64A**) témoigne de leur homologie et, d'autre part, que ces deux classes évoluent à la même vitesse (taux de mutation similaires), on pourrait être conduit à admettre que les espaceurs longs soient contemporains des espaceurs courts et qu'ils en dérivent.

Cette conclusion ne peut cependant être retenue pour deux raisons majeures. Nous avons en effet tout d'abord montré que les espaceurs longs sont sans doute plus anciens (**Partie II-A** du **Chapitre IV**) avec comme principal argument l'existence d'un motif commun (5SmsCR1) entre les espaceurs longs de mélèze d'Europe et les espaceurs courts ou longs de *Pinus radiata*. Nous avons en outre pu aussi mettre en évidence la présence du motif 5SmsCR1 au sein des unités longues et/ou intermédiaires d'autres espèces de mélèzes eurasiens (dont le mélèze du Japon, **voir Figure 62**) et américains, ainsi que chez *Pseudotsuga menziesii*, *Pinus sylvestris*, *Pinus nigra* et vraisemblablement aussi chez *Pinus pinaster*, *Picea glauca*, *Picea mariana* et *Picea abies* (voir la **Partie IIB** de ce **Chapitre IV**). Ces résultats confortent l'hypothèse du caractère ancestral des espaceurs longs car l'apparition du motif 5SmsCR1 remonte sans doute à une époque antérieure ou contemporaine de la diversification des espèces au sein de la famille des *Pinaceae*. La présence de nombreuses mutations

dans les espaceurs longs, entre les éléments dupliqués en tandem α et β d'une part, et A, B, C, D et E d'autre part (voir la **Figure 42**), va dans le même sens. De plus, si les espaceurs courts étaient les plus anciens, on devrait les trouver non seulement chez les mélèzes mais aussi chez d'autres espèces de *Pinaceae*. Or, seuls des espaceurs de type long (renfermant le motif 5SmsCR1) ont été détectés jusqu'à présent dans les génomes de *Pinus radiata* (Gorman *et al.* 1992, Moran *et al.* 1992) et de *Pseudotsuga menziesii* (Amarasinghe et Carlson 1998). De même, l'examen de la bordure 3' du clone 5S*Pg* d'espaceur court de *Picea glauca* présenté dans la publication de Brown et Carlson (1997) montre une plus grande similarité de séquence avec les espaceurs de type long qu'avec les espaceurs de type court des mélèzes (voir **Figure 47B**). Le clone 5S*Pg* étudié par ces auteurs est en effet caractérisé par 6 mutations communes avec les espaceurs longs de mélèze (positions - 35, - 43, - 44, - 49, - 52 et - 57 en amont du gène 5S). Il est très probable de plus que les espaceurs longs également détectés avec difficulté chez cette espèce (par Southern blot uniquement), renferment aussi le motif 5SmsCR1 (voir la **Partie IIB** de ce **Chapitre IV**). L'absence d'unités de type court chez les *Pinaceae* étudiées autres que les mélèzes suggère donc fortement (mais ne démontre pas) qu'elles ont émergé par délétion ou par perturbation de la zone centrale des espaceurs longs qui contiennent le motif conservé 5SmsCR1. Comme on ne retrouve les duplications observées dans les espaceurs longs des mélèzes (en particulier la duplication du motif α côté 5') ni chez *Pinus radiata* ni chez *Pseudotsuga menziesii*, on peut en conclure que les unités courtes ont probablement évolué à partir des unités longues avant la formation des duplications α, β et A, B, C, D, E. Le scénario d'apparition des séries composites que nous avons retenu (scénario 2, **Partie II-A** du **Chapitre IV**) se confirme donc : les unités courtes auraient divergé en bordure d'une série ancestrale d'unités longues avant que les mécanismes de l'évolution concertée n'amplifient, de façon fortuite ou en réponse à la présence de gènes 5S non fonctionnels, le motif dimérique par succession de crossing-over inégaux et/ou d'autres mécanismes moléculaires susceptibles d'amplifier rapidement une séquence.

La deuxième raison qui peut être évoquée à l'encontre d'une antériorité des espaceurs courts par rapport aux espaceurs longs, est leur association au sein de séries composites au locus 5SA. Cette organisation ne permet raisonnablement pas en effet de considérer que les espaceurs courts et les espaceurs longs évoluent de manière totalement indépendante, au moins depuis que la structure dimérique est apparue, c'est-à-dire avant la diversification du genre *Larix* (voir la **Partie IIB** de ce **Chapitre IV**) entre la fin de l'ère secondaire (il y a 65 à 130 Ma) et le milieu de l'ère tertiaire (il y a 25 à 50 Ma). Si les mécanismes de l'évolution concertée sont bien à l'origine de l'amplification du motif dimérique, seul un petit nombre d'unités courtes et longues originelles ont pu contribuer à la construction initiale des séries composites. A la limite, ces dernières pourraient avoir été construites à partir d'une seule unité courte et d'une seule unité longue (voir par exemple le modèle d'homogénéisation et amplification par crossing-over inégal, **Figure 9A**). Dans ce cas, les diversités nucléotidiques des espaceurs courts et des espaceurs longs au sein des séries composites récemment établies devraient être faibles et similaires. Le déséquilibre de mutation observé actuellement entre les espaceurs courts et les espaceurs longs des mélèzes d'Europe et du Japon s'est donc probablement

développé postérieurement à l'établissement des séries composites dans le genre *Larix*. Cette situation implique que les espaceurs longs aient évolué sous pression de la sélection naturelle, alors que les espaceurs courts auraient plutôt évolué par dérive génétique.

L'interprétation que nous proposons pour expliquer le déséquilibre de mutation au sein des séries composites repose essentiellement sur la présence au sein des unités longues du motif conservé 5SmsCR1 qui présente une similarité de séquence troublante avec une petite région des gènes codant pour la sous-unité α du complexe protéasomique 20S (Parmentier *et al.* 1997) et une analogie de taille et de structure avec un gène d'ARNt accepteur de la cystéine (voir la **Partie IIB** de ce **Chapitre IV**). Si l'on accepte de considérer le motif 5SmsCR1 comme un pseudogène protéasomique ou d'ARNt, il est probable qu'il ait été fonctionnel à une époque remontant à la diversification des *Pinaceae* dont l'origine monophylétique est maintenant bien établie (Chaw *et al.* 1997). Une telle association entre l'ADNr 5S et d'autres familles multigéniques, comme l'ADNr ou les gènes des histones, a été observée chez plusieurs champignons, des protistes, des nématodes, des araignées et des crustacés (revue dans Drouin et Moniz de Sá 1995). Par exemple, Barzotti *et al.* (2000) ont récemment montré l'existence chez le crustacé isopode *Asellus aquaticus* d'une unité répétée en tandem de 6553 pb (200-300 unités/C) constituée d'une copie des 4 gènes d'histones H2A, H2B, H3 et H4 et d'un gène 5S séparés par des espaceurs de taille variable. De même, un pseudogène d'ARNt accepteur de l'histidine a été mis en évidence au sein des espaceurs de l'ADNr 5S chez le riz, suggérant une ancienne "coopération" fonctionnelle entre ces deux familles de gènes (Hariharan *et al.* 1987). Si, chez les mélèzes, le motif 5SmsCR1 a conservé sa fonction pendant un laps de temps suffisant après l'établissement des séries composites, la sélection naturelle a pu alors s'exercer sur les espaceurs longs, selon le modèle proposé par Kellogg et Appels (1995) et Cronn *et al.* (1996) pour les gènes 5S : les mutations s'accumulent au hasard dans une série mais lorsque le nombre de pseudogènes liés au motif 5SmsCR1 devient trop grand la sélection naturelle provoque leur élimination de la population. Les séries qui accumulent par hasard plus de mutations dans les espaceurs longs sont donc régulièrement éliminées tandis que les séries qui accumulent par hasard plus de mutations dans les espaceurs courts, qui ne contiennent pas le motif 5SmsCR1, sont au contraire conservées. Finalement, le motif 5SmsCR1 ayant apparemment perdu sa fonction au cours de l'évolution, les espaceurs longs se sont trouvés soumis aux mêmes contraintes évolutives que les espaceurs courts ("molecular drive", Dover 1982).

Si l'on considère maintenant le cas des espaceurs longs, il est surprenant de constater qu'ils restent très peu variables alors que le motif 5SmsCR1 présente une analogie de séquence assez éloignée avec les gènes fonctionnels (gènes protéasomiques ou d'ARNt) dont il dériverait. On peut cependant remarquer que les séries de l'ADNr 5S ont probablement été soumises pendant une longue période à une pression de sélection double affectant à la fois les gènes 5S et les séquences fonctionnelles ancestrales du motif 5SmsCR1 des espaceurs longs. Comme nous l'avons évoqué dans le cas des chênes pédonculés et sessiles (**Chapitre III**), on peut donc envisager que les espaceurs longs subissent encore actuellement une pression de sélection indirecte liée à leur association avec les gènes 5S. On pourrait également penser que les phénomènes d'évolution concertée sont impliqués

dans l'émergence de cette situation. Il est cependant difficile d'expliquer une éventuelle réduction de l'hétérogénéité de séquence des espaceurs longs par les mécanismes moléculaires classiques d'homogénéisation (crossing-over inégal et conversion génique ectopique) car ils devraient aussi affecter les unités courtes. Les résultats récents de Polanco *et al.* (1998) permettent cependant d'envisager l'existence de mécanismes d'évolution concertée capables d'homogénéiser de manière indépendante deux régions d'une même unité répétée en tandem. Ces auteurs ont en effet démontré en étudiant l'évolution de l'ADNr chez la drosophile, que les ITS subissent uniquement une forte homogénéisation intrachromosomique tandis que les IGS seraient engagés dans de fréquents échanges interchromosomiques du type crossing-over inégal. Des échanges simple brin induits par la présence des sous-répétitions de l'IGS constitueraient un mécanisme moléculaire susceptible de rendre compte de cette situation paradoxale. On peut donc imaginer que les éléments dupliqués $\alpha$, $\beta$ et parfois $\chi$ (**Figure 64B**) ou A, B, C, D, et E (**Figure 42A**) présents dans les espaceurs longs des mélèzes puissent jouer le même rôle que les sous-répétitions de l'IGS, ce qui expliquerait l'homogénéisation plus efficace des espaceurs longs par rapport aux espaceurs courts qui sont dépourvus d'éléments répétés en tandem. Un tel mécanisme d'évolution concertée sélectif pourrait même être considéré comme étant à lui seul suffisant pour rendre compte du déséquilibre de mutations au sein des séries composites des mélèzes d'Europe et du Japon.

d- L'absence de différenciation entre mélèze d'Europe et mélèze du Japon au locus 5SA

Aucune mutation ne permet de distinguer le mélèze d'Europe et le mélèze du Japon au locus 5SA. Cette absence de différenciation ne peut pas être attribuée à la faible variabilité de l'ADNr 5S car la diversité nucléotidique des espaceurs, particulièrement des espaceurs courts, est relativement élevée et du même ordre de grandeur que les estimations obtenues pour les *Triticeae* (Kellogg et Appels 1995) et dans le genre *Gossypium* (Cronn *et al.* 1996). Elle ne peut pas non plus résulter de phénomènes d'hybridation interspécifique répétés comme cela semble être le cas pour les chênes pédonculé et sessile (voir la **Partie II** du **Chapitre III**). En effet, bien que sexuellement compatibles, ces deux espèces occupent des aires de distribution éloignées (espèces allopatriques). La seule hypothèse plausible est donc que le temps qui s'est écoulé depuis la divergence de ces deux mélèzes (34 à 52 Ma selon Le Page et Basinger 1995) est trop faible pour que des mutations aient pu être fixées au locus 5SA. Autrement dit, à cette échelle de temps, la fréquence des mécanismes d'homogénéisation intragénomique (ici évolution concertée intralocus) serait extrêmement faible. Cette situation est certainement liée au long temps de génération de ces espèces ligneuses comparativement aux plantes à cycle court que l'on peut distinguer sur la base des variations des espaceurs de l'ADNr 5S qui se trouvent rapidement fixées dans les populations, même lorsqu'elles sont apparentées (Kolchinsky *et al.* 1991, Cox *et al.* 1992, Kanazin *et al.* 1993, Zanke *et al.* 1995, Ko et Henry 1996). A notre connaissance, aucune mutation spécifique de l'une ou l'autre espèce n'a pu être mise en évidence dans le génome nucléaire jusqu'à présent. Seuls des marqueurs cytoplasmiques à hérédité maternelle (marqueurs mitochondriaux) ou paternelle (marqueurs chloroplastiques) ont pu être identifiés. Les fragments RAPD spécifiques de l'une ou l'autre espèce détectés par Arcade *et al.*

(1996) et Scheepers *et al.* (2000), transformés en SCAR puis séquencés, se sont ainsi révélés être des fragments d'ADN mitochondrial. Un marqueur chloroplastique a par ailleurs été identifié par Acheré (1999). On peut penser que la saturation des cartes génétiques de mélèzes d'Europe et du Japon élaborées récemment par Arcade *et al.* (2000) permettra d'identifier les régions du génome nucléaire impliquées dans la différenciation de ces deux espèces phylogénétiquement très proches (voir Gernandt et Liston 1999) et de dégager des marqueurs nucléaires discriminants.

## Bibliographie


**Acheré V (1999).** Recherche de marqueurs moléculaires pour l'identification des mélèzes hybrides (*Larix x eurolepis*). DEA de l'Université Henri Poincaré, Nancy I. 20 pp.

**Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990).** Basic local alignment search tool. J Mol Biol 215 : 403-410

**Amarasinghe V, Carlson JE (1998).** Physical mapping and characterization of 5S rRNA genes in Douglas-Fir. J Hered 89 : 495-500

**Appels R, Baum BR, Clarke BC (1992).** The 5S DNA units of bread wheat (*Triticum aestivum*). Plant Syst Evol 183 : 183-194

**Arcade A (1999).** Détection de locus contrôlant des caractères quantitatifs dans un plan de croisement factoriel : application au mélèze. Thèse de l'Université d'Orléans, 160 pp.

**Arcade A, Faivre-rampant P, Le Guerroué B, Pâques LE, Prat D (1996).** Heterozygosity and hybrid performance in larch. Theor Appl Genet 93 : 1274-1281

**Arcade A, Anselin F, Faivre Rampant P, Lesage MC, Pâques LE, Prat D (2000).** Application of AFLP, RAPD and ISSR markers to genetic mapping of European and Japanese larch. Theor Appl Genet 100 : 299-307.

**Arnheim N (1983).** Concerted evolution of multigene families. Dans *Evolution of genes and proteins* pp 38-61. Edité par Nei M, Koehn RK. Sinauer associates, Sunderland MA

**Axelrod DI (1990).** Environment of the Middle Eocene (45 Ma) Thunder Mountain flora, central Idaho. National Geographic Research 6 : 355-361

**Barciszewska MZ, Erdmann VA, Barciszewski J (1994).** A new model for the tertiary structure of 5S ribonucleic acids in plants. Plant Mol Biol Rep 12 : 116-131.

**Barrénèche T, Bodénès C, Lexer C, Trontin JF, Fluch S, Streiff R, Plomion C, Roussel G, Steinkellner H, Burg K, Favre JM, Glössl J, Kremer A (1998).** A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers. Theor Appl Genet 97 : 1090-1103

**Barzotti R, Pelliccia F, Bucciarelli E, Rocchi A (2000).** Organization, nucleotide sequence, and chromosomal mapping of a tandemly repeated unit containing the four core histone genes and a 5S rRNA gene in an isopod crustacean species. Genome 43 : 341-345

**Bassam BJ, Caetano-Anollés G, Gresshoff PM (1991).** Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal Biochem 196 : 80-83

**Bastien JC, Keller R (1980).** Intérêts comparés du mélèze hybride (*Larix x eurolepis* Henry) avec les deux espèces parentes. Rev For Fr 32 : 521-530

**Baum BR, Appels R (1992).** Evolutionary change at the 5S DNA loci of species in the *Triticeae*. Pl Syst Evol 183 : 195-208

**Brondani RPV, Brondani C, Tarchini R, Grattapaglia D (1998).** Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. Theor Appl Genet : 97 : 816-827

**Brown GR, Carlson JE (1997).** Molecular cytogenetics of the genes encoding 18S-5.8S-26S rRNA and 5S rRNA in two species of spruce (*Picea*). Theor Appl Genet 95 : 1-9

**Brown GR, Newton CH, Carlson JE (1998).** Organization and distribution of a *Sau*3A tandem repeated DNA sequence in *Picea* (*Pinaceae*) species. Genome 41 : 560-565

**Campell BR, Song Y, Posch TE, Cullis CA, Town CD (1992).** Sequence and organization of 5S ribosomal RNA-encoding genes of *Arabidopsis thaliana*. Gene 112 : 225-228

**Carlson CE, Theroux LE (1993).** Cone and seed morphology of western larch (*Larix occidentalis*), alpine larch (*Larix lyalli*), and their hybrid. Can J For Res 23 : 1264-1269

**Challice JM, Segall J (1989).** Transcription of the 5S rRNA gene of *Saccharomyces cerevisiae* requires a promoter element at +1 and a 14-base pair internal control region. J Biol Chem 264 : 20060-20067

**Chanteranne D (1999).** Recherche d'un motif conservé (5SmsCR1) au sein des espaceurs non transcrits des unités d'ADN ribosomique 5S de quelques gymnospermes. Rapport de stage de Maîtrise de Biologie des Populations et des Ecosystèmes. Université Henri Poincaré, Nancy I. 15 pp

**Chaw SM, Zharkikh A, Sung HM, Lau TC, Li WH (1997).** Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. Mol Biol Evol 14 : 56-68

**Chollet F (1986).** Les vergers à graines. Rev For Fr (n° sp) : 74-80

**Collardet J, Besset J (1988).** Bois commerciaux. Les résineux (conifères). Tome I. Editions H. Vial et Centre Technique du Bois et de l'Ameublement

**Copenhaver GP, Pikaard CS (1996).** Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in *Arabidopsis thaliana* suggesting local spreading of variants as the mode for gene homogenization during concerted evolution. Plant J 9 : 273-282

**Cox AV, Bennett MD, Dyer TA (1992).** Use of the polymerase chain reaction to detect spacer size heterogeneity in plant 5S-rRNA gene clusters and to locate such clusters in wheat (*Triticum aestivum* L.). Theor Appl Genet 83 : 684-690

**Cronn RC, Zhao X, Paterson AH, Wendel JF (1996).** Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. J Mol Evol 42 : 685-705

**Cullis CA, Creissen GP, Gorman SW, Teasdale RD (1988).** The 25S, 18S, and 5S ribosomal RNA genes from *Pinus radiata* D. Don. Dans *Molecular genetics of forest trees*, pp 34-40. Edité par Cheliak WM, Yapa AC. Petawawa Natl For Inst Rep PI-X-80, Chalk River, Ontario, Canada

**Debazac EF (1977).** Manuel des conifères. ENGREF, Nancy.

**De Verno LL, Charest PJ, Bonen L (1993).** Inheritance of mitochondrial DNA in the conifer *Larix*. Theor Appl Genet 86 : 383-388

**Dhillon SS (1987).** DNA in tree species. Dans *Cell and tissue culture in forestry*, pp 298-313. Edité par Bonga JM, Durzan DJ. Martinus Nijhoff Publisher

**Doudrick RL, Heslop-Harrison JS, Nelson CD, Schmidt T, Nance WL, Schwarzacher T (1995).** Karyotype of Slash Pine (*Pinus elliottii* var. *elliottii*) using patterns of fluorescence *in situ* hybridization and fluorochrome banding. J of Heredity 86 : 289-296

**Dover GA (1982).** Molecular drive: a cohesive mode of species evolution. Nature 299 : 111-117

**Doyle JJ, Doyle JL (1987).** A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19 : 11-15

**Drouin G, Moniz de Sá M (1995).** The concerted evolution of 5S ribosomal genes linked to the repeat units of other multigene families. Mol Biol Evol 12 : 481-493

**Dvorák J, Zhang HB, Kota RS, Lassner M (1989).** Organization and evolution of the 5S ribosomal RNA gene family in wheat and related species. Genome 32 : 1003-1016

**Echt CS, Nelson CD (1997).** Linkage mapping and genome length in eastern white pine (*Pinus strobus* L.). Theor Appl Genet 94 : 1031-1037.

**Ellis THN, Lee D, Thomas CM, Simpson PR, Cleary WG, Newman MA, Burcham KWG (1988).** 5S rRNA genes in *Pisum* : sequence, long range and chromosomal organization. Mol Gen Genet 214 : 333-342

**Ennos RA, Qian T (1994).** Monitoring the output of a hybrid larch seed orchard using isozyme markers. Forestry 67 : 63-73

**Farjon A (1990).** *Pinaceae*. Koeltz Scientific Books, Konigstein, Germany.

**Felsenstein J (1995).** PHYLIP (Phylogeny Inference Package) version 3.57c. Department of Genetics, University of Washington, Seattle.

**Ferrand JC (1986).** Mélèzes. Rev For Fr 38 (n° sp) : 142-145

**Gernandt DS, Liston A (1999).** Internal transcribed spacer region evolution in *Larix* and *Pseudotsuga* (*Pinaceae*). Amer J Bot 86 : 711-723

**Goldsbrough PB, Ellis THN, Cullis CA (1981).** Organisation of the 5S RNA genes in flax. Nucleic Acids Res 9 : 5895-5904

**Gorman SW, Teasdale RD, Cullis CA (1992).** Structure and organization of the 5S rRNA genes (5S DNA) in *Pinus radiata* (*Pinaceae*). Pl Syst Evol 183 : 223-234

**Gottlob-McHugh SG, Lévesque M, MacKenzie K, Olson M, Yarosh O, Johnson DA (1990).** Organization of the 5S rRNA genes in the soybean *Glycine max* (L.) Merrill and conservation of the 5S rDNA repeat structure in higher plants. Genome 33: 486-494

**Grandemange C (1994).** L'ADN ribosomiqque 5S chez le mélèze : clonage, séquençage et essai d'utilisation pour le typage de différentes espèces. DEA en Biologie Forestière, 18 pp. Université Henri Poincaré, Nancy I

**Greilhuber J (1986).** Severely distorted Feulgen-DNA amounts in *Pinus* (*Coniferophytina*) after non-additive fixation as a result of meristematic self-tanning with vacuole contents. Can J Gen Cyt 28 : 409-415

**Häcker M, Bergmann F (1991).** The proportion of hybrids in seed from a seed orchard composed of two larch species (*L. europaea* and *L. leptolepis*). Ann Sci For 48 : 631-640

**Hanson RE, Islam-Faridi NM, Percival EA, Crane CF, Ji Y, McKnight TD, Stelly DM, Price HJ (1996).** Distribution of 5S and 18S-28S rDNA loci in a tetraploid cotton (*Gossypium hirsutum* L.) and its putative diploid ancestors. Chromosoma 105 : 55-61

**Hariharan N, Reddy PS, Padayatty JD (1987).** 5S rRNA genes in rice embryos. Plant Mol Biol 9 : 443-451

**Hemleben V, Werts D (1988).** Sequence organization and putative regulatory elements in the 5S rRNA genes of two higher plants (*Vigna radiata* and *Matthiola incana*). Gene 62 : 165-169

**Higgins DG, Bleasby AJ, Fuchs R (1992).** CLUSTAL V: improved software for multiple sequence alignment. Comput Appl Biosci 8 : 189-191

**Hizume M (1988).** Karyomorphological studies in the family *Pinaceae*. Mem Fac Educ Ehime Univ Nat Sci 8 : 1-108

**Hizume M, Tominaga HH, Kondo K, Gu Z, Yue Z (1993).** Fluorescent chromosome banding in six taxa of Eurasian *Larix*, *Pinaceae*. Kromosomo 69 : 2342-2354

**Hori H, Lim BL, Osawa S (1985).** Evolution of green plants as deduced from 5S rRNA sequences. Proc Natl Acad Sci USA 82 : 820-823

**Kanazin V, Ananiev E, Blake T (1993).** The genetics of 5S rRNA encoding multigene families in barley. Genome 36 : 1023-1028

**Kellogg EA, Appels R (1995).** Intraspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. Genetics 140 : 325-343

**Kimura M (1980).** A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16 : 111-120

**Kisanuki H, Kurahashi A, Kato H, Terauchi R, Kawano S, Ide Y, Watanabe S (1995).** Interspecific relationship of genus *Larix* inferred from the RFLPs of chloroplast DNA. J Jap For Soc 77 : 83-85

**Ko HL, Henry RJ (1996).** Specific 5S ribosomal RNA primers for plant species identification in admixtures. Plant Mol Biol Rep 14 : 33-43

**Kolchinsky A, Kolesnikova M, Ananiev E (1991).** "Portraying" of plant genomes using polymerase chain reaction amplification of ribosomal 5S genes. Genome 34 : 1028-1031

**Korn LJ (1982).** Transcription of *Xenopus* 5S ribosomal RNA genes. Nature 295 : 101-105

**Khoshoo TN (1961).** Chromosome numbers in gymnosperms. Silvae genet 10 : 1-9

**Lapitan NLV (1992).** Organization and evolution of higher plant nuclear genomes. Genome 35 : 171-181

**Le Cam J (1981).** Contribution à la connaissance des mélèzes hybrides. AFOCEL, Annales de Recherches Sylvicoles, pp. 107-137

**LePage BA, Basinger JF (1991).** A new species of *Larix* (*Pinaceae*) from the early Tertiary of Axel Heiberg Island, Arctic, Canada. Review of Palaeobotany and Palynology 70 : 89-111

**LePage BA, Basinger JF (1995).** The evolutionary history of the genus *Larix* (*Pinaceae*). Symposium on ecology and management of *Larix* forests: a look ahead. USDA, Forest Service, Intermountain Research Station, General Technical Report GTR-INT-319. pp 19-29

**Lewandowski A (1997).** Genetic relationships between European and Siberian larch, *Larix* spp. (*Pinaceae*), studied by allozymes: is the Polish larch a hybrid between these two species ? Plant Syst Evol 204 : 65-73

**Liao D (1999).** Concerted evolution : molecular mechanism and biological implications. Am J Hum Genet 64 : 24-30

**Liao D, Pavelitz T, Kidd JR, Kidd KK, Weiner AM (1997).** Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the *RNU2* locus) involves rapid intrachromosomal homogenization and rare interchromosomal gene conversion. EMBO J 16 : 588-598

**Linares AR, Bowen T, Dover GA (1994).** Aspects of nonrandom turnover involved in the concerted evolution of intergenic spacers within the ribosomal DNA of *Drosophila melanogaster*. J Mol Evol 39 : 151-159

**Loconte H (1996).** Comparison of alternative hypotheses for the origin of the angiosperms. Dans *Flowering plant: origin, evolution and phylogeny* pp 267-285. Edité par Taylor DW, Hickey LJ. Chapman & Hall, New-York

**Lubaretz O, Fuchs J, Ahne R, Meister A, Schubert I (1996).** Karyotyping of three *Pinaceae* species *via* fluorescent *in situ* hybridization and computer-aided chromosome analysis. Theor Appl Genet 92 : 411-416

**Maier J (1992).** Genetic variation in European larch (*Larix decidua* Mill.). Ann Sci For 49 : 39-47

**Mashkova TD, Barciszewska MZ, Joachimiak A, Nalaskowska M, and Barciszewski J (1990).** Molecular evolution of plants as deduced from changes in free energy of 5S ribosomal RNAs. Int J Biol Macromol 12 : 247-250

**McMahon ME, Stamenkovitch D, Petes TD (1984).** Tandemly arranged variant 5S ribosomal RNA genes in the yeast *Saccharomyces cerevisiae*. Nucleic Acids Res 12 : 8001-8016

**Melekhovets YF, Troitsky AV, Valiejo-Roman KM, Bobrova VK, Antonov AS (1988).** Nucleotide sequences of cytosolic 5S ribosomal RNAs from two gymnosperms, *Gnetum gnemon* and *Ephedra kokanica*. Nucleic Acids Res 16 : 4155

**Mill RR (1999).** A new species of *Larix* (*Pinaceae*) from Southeast Tibet and other nomenclatural notes on Chinese *Larix*. NOVON 9 : 79-82.

**Moran GF, Smith D, Bell JC, Appels R (1992).** The 5S RNA genes in *Pinus radiata* and the spacer region as a probe for relationships between *Pinus* species. Pl Syst Evol 183 : 209-221

**Nagylaki T (1984).** The evolution of multigene families under intra-chromosomal gene conversion. Genetics 106 : 529-548

**Nagylaki T (1990).** Gene conversion, linkage, and the evolution of repeated genes dispersed among multiple chromosomes. Genetics 126 : 261-276

**Nalaskowska M, Mashkova TD, Barciszewski J (1990).** The primary structure of 5S ribosomal RNAs from *Magnolia cobus* and *Magnolia stellata*. Nucleic Acids Res 18 : 666

**Nei M (1987).** Molecular Evolutionary Genetics. Columbia University Press, New York

**Nkongolo KK, Klimaszewska K (1995).** Cytological and molecular relationships between *Larix decidua*, *L. leptolepis* and *Larix x eurolepis*: identification of species-specific chromosomes and synchronization of mitotic cells. Theor Appl Genet 90 : 827-834

**Ohta T, Dover GA (1983)**. Population genetics of multigene families that are dispersed into two or more chromosomes. Proc Natl Acad Sci USA 80 : 4079-4083

**Page CN (1990)**. Coniferophytina. Dans *The families and genera of vascular plants* (édité par Kubitski K), *Vol I : Pteridophytes et gymnospermes*, pp. 282-361. Edité par Kramer U, Green PS. Springer-Verlag, Berlin.

**Paglia GP, Olivieri AM, Morgante M (1998)**. Towards second-generation STS (sequence-tagged sites) linkage maps in conifers: a genetic map of Norway spruce (*Picea abies* K.). Mol Gen Genet 258 : 466-478

**Pâques LE (1989)**. A critical review of larch hybridization and its incidence on breeding strategies. Ann Sci For 46 : 141-153

**Pâques LE (1992)**. Les mélèzes. Dans *Amélioration des espèces végétales cultivées*, pp. 720-731. Edité par Gallais A, Bannerot H. INRA Editions, Paris.

**Pâques LE (1994)**. Mélèzes d'Europe et du Japon. Forêt Entreprise 96 : 73-75

**Pâques LE (1996a)**. Variabilité naturelle du mélèze. I. Mélèze d'europe : bilan de 34 ans de test comparatif de provenances. Ann Sci For 53 : 51-67

**Pâques LE (1996b)**. Variabilité naturelle du mélèze. II. Mélèze du Japon : bilan de 36 ans de test comparatif de provenances. Ann Sci For 53 : 69-78

**Parmentier Y, Bouchez D, Fleck J, Genschik P (1997)**. The 20S proteasome gene family in Arabidopsis thaliana. FEBS Letters 416 : 281-285

**Perry DJ, Bousquet J (1998)**. Sequence-tagged-site (STS) markers of arbitrary genes: the utility of black spruce-derived STS primers in other conifers. Theor Appl Genet 97 : 735-743.

**Pieler T, Hamm J, Roeder RG (1987)**. The 5S gene internal control region is composed of three distinct sequence elements, organized as two functional domains with variable spacing. Cell 48 : 91-100

**Pieler T, Oei SL, Hamm J, Engelke U, Erdmann VA (1985)**. Functional domains of the *Xenopus laevis* 5S gene promoter. EMBO J 4 : 3751-3756

**Polanco C, González AI, de la Fuente Á, Dover GA (1998)**. Multigene family of ribosomal DNA in *Drosophila melanogaster* reveals contrasting patterns of homogenization for IGS and ITS spacer regions: a possible mechanism to resolve this paradox. Genetics 149 : 243-256

**Qian T, Ennos RA, Helgason T (1995)**. Genetic relationships among larch species based on analysis of restriction fragment variation for chloroplast DNA. Can J For Res 25 : 1197-1202

**Reddy P, Appels R (1989)**. A second locus for the 5S multigene family in *Secale* L.: sequence divergence in two lineages of the family. Genome 32 : 456-467

**Riesberg LH, Brunsfeld SJ (1992)**. Molecular evidence and plant introgression. Dans *Molecular Systematics of Plants*, pp. 151-176. Edité par Soltis PS, Soltis DE, Doyle JJ. Chapman & Hall, New York

**Saitou N, Nei M (1987)**. The Neighbor-Joining Method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4 : 406-425

**Sambrook J, Fritsch EF, Maniatis T (1989)**. Molecular cloning: a laboratory manual. Second edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

**Sastri DC, Hilu K, Appels R, Lagudah ES, Playford J, Baum BR (1992)**. An overview of evolution in plant 5S DNA. Pl Syst Evol 183 : 169-181

**Scheepers D, Eloy MC, Briquet M (2000)**. Identification of larch species (*Larix decidua*, *Larix kaempferi*, and *Larix x eurolepis*) and estimation of hybrid fraction in seed lots by RAPD fingerprints. Theor Appl Genet 100 : 71-74

**Schneeberger RG, Creissen GP, Cullis CA (1989)**. Chromosomal and molecular analysis of 5S RNA gene organization in the flax, *Linum usitatissimum*. Gene 83 : 75-84

**Schneeberger RG, Cullis CA (1992)**. Intraspecific 5S rRNA gene variation in flax, *Linum usitatissimum* (*Linaceae*). Pl Syst Evol 183 : 265-280

**Schondelmaier J, Schmidt T, Heslop-Harrison JS, Jung C (1997)**. Genetic and chromosomal localization of the 5S rDNA locus in sugar beet (*Beta vulgaris* L.). Genome 40 : 171-175

**Schorn HE (1994)**. A preliminary discussion of fossil larches (*Larix*, *Pinaceae*) from the Arctic. Quaternary International 22/23 : 173-183

**Scoles GJ, Gill BS, Xin ZY, Clarke BC, McIntyre CL, Chapman C, Appels R (1988)**. Frequent duplication and deletion events in the 5S RNA genes and the associated spacer regions of the *Triticeae*. Pl Syst Evol 160 : 105-122

**Selker EU, Morzycka-Wróblewska E, Stevens JN, Metzenberg RL (1986)**. An upstream signal is required for *in vitro* transcription of *Neurospora* 5S RNA genes. Mol Gen Genet 205 : 189-192

**Semerikov V, Lascoux M (1999)**. Genetic relationship among Eurasian and American *Larix* species based on allozymes. Heredity 83 : 62-70

**Semerikov V, Semerikov LF, Lascoux M (1999)**. Intra- and interspecific allozyme variability in Eurasian *Larix* Mill. species. Heredity 82 : 193-204

**Shin DI, Podila GK, Huang Y, Karnosky DF (1994)**. Transgenic larch expressing genes for herbicide and insect resistance. Can J For Res 24 : 2059-2067

**Singer M, Berg P (1992)**. Gènes et génomes. Editions Vigot, Paris

**Smith GP (1976)**. Evolution of repeated DNA sequences by unequal crossover. Science 191 : 528-535

**Steele KP, Holsinger KE, Jansen RK, Taylor DW (1991)**. Assessing the reliability of 5S rRNA sequence data for phylogenetic analysis in green plants. Mol Biol Evol 8 : 240-248

**Szmidt AE, Alden T, Hällgren JE (1987)**. Paternal inheritance of chloroplast DNA in *Larix*. Plant Mol Biol 9 : 59-64

**Szymanski M, Specht T, Barciszewska MZ, Barciszewski J, and Erdmann VA (1998).** 5S rRNA data bank. Nucleic Acids Res 26 : 156-159

**Swofford DL (1993).** PAUP: phylogenetic analysis using parsimony, version 3.1.1. Illinois Natural History Survey, Champaign, IL

**Trontin (1993).** L'ADN ribosomique 5S chez les mélèzes (*Larix decidua* Mill., *Larix x eurolepis* Henry, *Larix x leptoeuropaea* Dengler) : éléments d'organisation, stabilité au sein de cultures embryogènes. DEA en Biologie Forestière, 20 pp. Université Henri Poincaré, Nancy I

**Trontin JF, Grandemange C, Favre JM (1999).** Two highly divergent 5S rDNA unit size classes occur in composite tandem array in european larch (*Larix decidua* Mill.) and japanese larch (*Larix kaempferi* (Lamb.) Carr.). Genome 42 : 837-848

**Tyler-Smith C, Brown WRA (1987).** Structure of the major block of alphoid satellite DNA on the human Y chromosome. J Mol Biol 195 : 457-470

**Van de Peer Y, De Baere R, Cauwenbergh J, De Watcher R (1990).** Evolution of green plants and their relationship with other photosynthetic eucaryotes as deduced from 5S ribosomal RNA sequences. Pl Syst Evol 170 : 85-96

**Vendramin GG, Lelli L, Rossi P, Morgante M (1996).** A set of primers for amplification of 20 chloroplast microsatellites in *Pinaceae*. Mol Ecol 5 : 595-598

**Venkateswarlu K, Lee SW, Nazar RN (1991).** Conserved upstream sequence elements in plant 5S ribosomal RNA-encoding genes. Gene 105 : 249-253

**Wendel JF, Schnabel A, Seelanan T (1995).** Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc Natl Acad Sci USA 92 : 280-284

**Zanke C, Borisjuk N, Ruoss B, Schilde-Rentschler L, Ninnemann H, Hemleben V (1995).** A specific oligonucleotide of the 5S rDNA spacer and species-specific elements indentify symmetric somatic hybrids between *Solanum tuberosum* and *S. pinnatisectum*. Theor Appl Genet 90 : 720-726

**Zar JH (1996).** Biostatistical analysis. Troisième édition. Prentice Hall. Upper Saddle River, New Jersey.

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

# BILAN GENERAL ET PERSPECTIVES

Nous avons donné à la fin des différentes parties de chacun des deux chapitres expérimentaux concernant les chênes pédonculé et sessile (**Chapitre III**) et les mélèzes d'Europe et du Japon (**Chapitre IV**) un certain nombre de conclusions et discuté les principaux résultats obtenus. Dans cette dernière partie, nous nous efforcerons donc plutôt de dresser un bilan comparatif des connaissances acquises sur la structure, l'organisation, le polymorphisme et l'évolution des unités d'ADNr 5S chez ces deux modèles différents d'espèces forestières à cycle long. Nous essaierons en même temps de développer une réflexion prospective visant à évaluer l'intérêt de cette famille multigénique et à dégager quelques pistes pour de futures recherches.

## I- Structure des unités d'ADNr 5S

### 1- Taille et séquence nucléotidique des gènes 5S et des espaceurs

Les amorces universelles et spécifiques des gènes 5S nucléaires que nous avons utilisées (couples F/B et P1/P2) ont permis de révéler une unique classe d'espaceurs relativement courts chez les chênes pédonculé et sessile (environ 230 pb) et deux classes nettement plus longues chez les mélèzes d'Europe et du Japon (environ 530 pb et 750 pb). Il est cependant possible que d'autres classes d'espaceurs aient échappé à l'analyse. Il est en effet envisageable que des mutations affectent les bordures des gènes 5S où se fixent les amorces universelles, en particulier au niveau de leur extrémité 3' à partir de laquelle la *Taq* polymérase initie la synthèse au cours de l'amplification PCR. Cela pourrait expliquer la faiblesse des amplifications que nous avons obtenues à l'aide de ces amorces chez certains conifères comme les pins ou les épicéas. Il est possible aussi que les gènes 5S puissent adopter de manière transitoire une structure secondaire proche de celle des ARNr 5S au cours des cycles d'amplification par PCR qui limiterait alors l'accessibilité de l'ADN cible aux amorces (Buckler *et al.* 1997). Cette éventualité est appuyée expérimentalement par le fait que la technique, très sensible, de coloration des produits de PCR au nitrate d'argent après électrophorèse sur gel d'acrylamide révèle l'existence d'espaceurs relativement courts (310-330 pb) chez certains mélèzes d'Europe et du Japon dont il faudrait vérifier la nature (nouvelle classe d'espaceurs ou variants de longueur d'une des 2 classes S et L). Cependant, si de telles classes d'espaceurs supplémentaires existent réellement, elle sont très divergentes et/ou leur importance est nettement en deçà du seuil de détection par les méthodes de Southern blot à l'aide de sondes marquées à la digoxigénine que nous avons employées.

Au-delà de ces différences de taille, les espaceurs de chênes et de mélèzes sont extrêmement divergents comme l'on pouvait logiquement s'y attendre pour une telle région de l'ADN, intergénique et largement non transcrite. La comparaison des séquences consensus majoritaires établies soit pour l'unique classe d'espaceurs de chênes, soit pour les classes courtes et longues de mélèzes ne montre en effet aucune homologie significative, même au niveau des extrémités des espaceurs impliquées dans

la régulation de la transcription (comparer les **Figures 29, 60** et **62**). Si l'on considère que les spermaphytes actuelles ont une origine monophylétique (Chaw *et al.* 1997), on peut donc admettre que la structure (= taille + séquence) des espaceurs de l'ADNr 5S a dû considérablement évoluer depuis la séparation des angiospermes et des gymnospermes. En revanche, elle est encore sensiblement la même pour les chênes pédonculé et sessile d'une part, et pour les mélèzes d'Europe et du Japon, d'autre part, car aucune différence qualitative de longueur ou de séquence (mutation fixée) n'a été relevée entre les deux espèces de chacun des deux genres.

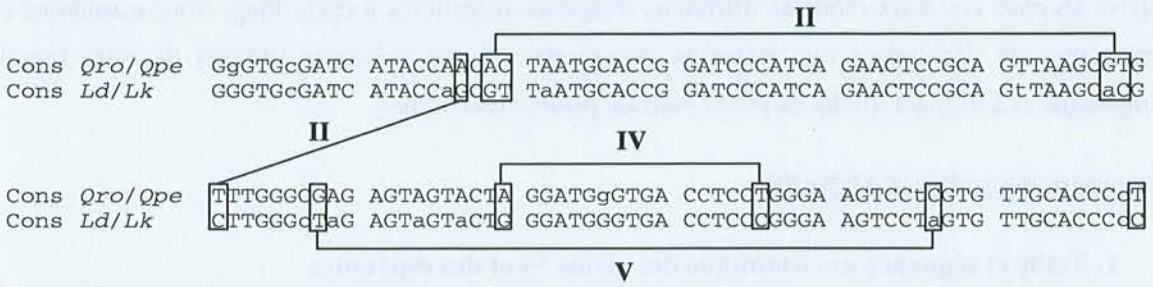



**Figure 65**. Comparaison des séquences nucléotidiques majoritaires de gènes 5S obtenues pour les chênes pédonculé et sessile (Cons *Qro/Qpe*) et les mélèzes d'Europe et du Japon (Cons *Ld/Lk*).
Les bases variables au sein de chaque séquence consensus sont indiquées en minuscules. Les positions qui diffèrent entre les deux séquences consensus sont encadrées. Les mutations non indépendantes qui apparaissent au sein des régions bicaténaires II, IV et V (voir **Figures 17, 40** ou **59**) de la structure secondaire des ARNr 5S (mutations compensatrices) sont reliées par des traits fins.

A la différence des espaceurs, la structure des gènes 5S apparaît quant à elle extrêmement conservée entre les 4 espèces étudiées. La taille des gènes 5S séquencés chez les chênes pédonculé et sessile (119-120 pb) et chez les mélèzes d'Europe et du Japon (120-122 pb) est très similaire à celle des autres plantes supérieures (120 pb). De même, les séquences consensus majoritaires des chênes et des mélèzes ont une taille de 120 pb et ne diffèrent l'une de l'autre qu'à 11 positions nucléotidiques (**Figure 65**). La plupart sont des transitions (9/11), particulièrement de type C→T (5/11) qui, selon certains auteurs (Goldsbrough *et al.* 1982, Fulnecek *et al.* 1998), pourraient être expliquées par une désamination des cytosines méthylées. On note en outre, comme Steele *et al.* (1991), que beaucoup de ces mutations (10/11) ne sont pas indépendantes, mais compensatrices (**Figure 65**), probablement en raison de la pression de sélection qui s'exerce sur les gènes 5S. L'intérêt du gène 5S pour les études phylogénétiques semble donc très limité, même à des niveaux taxinomiques élevés. Seules 6 mutations sont ainsi réellement informatives et indépendantes entre chênes pédonculé et sessile et mélèzes d'Europe et du Japon.

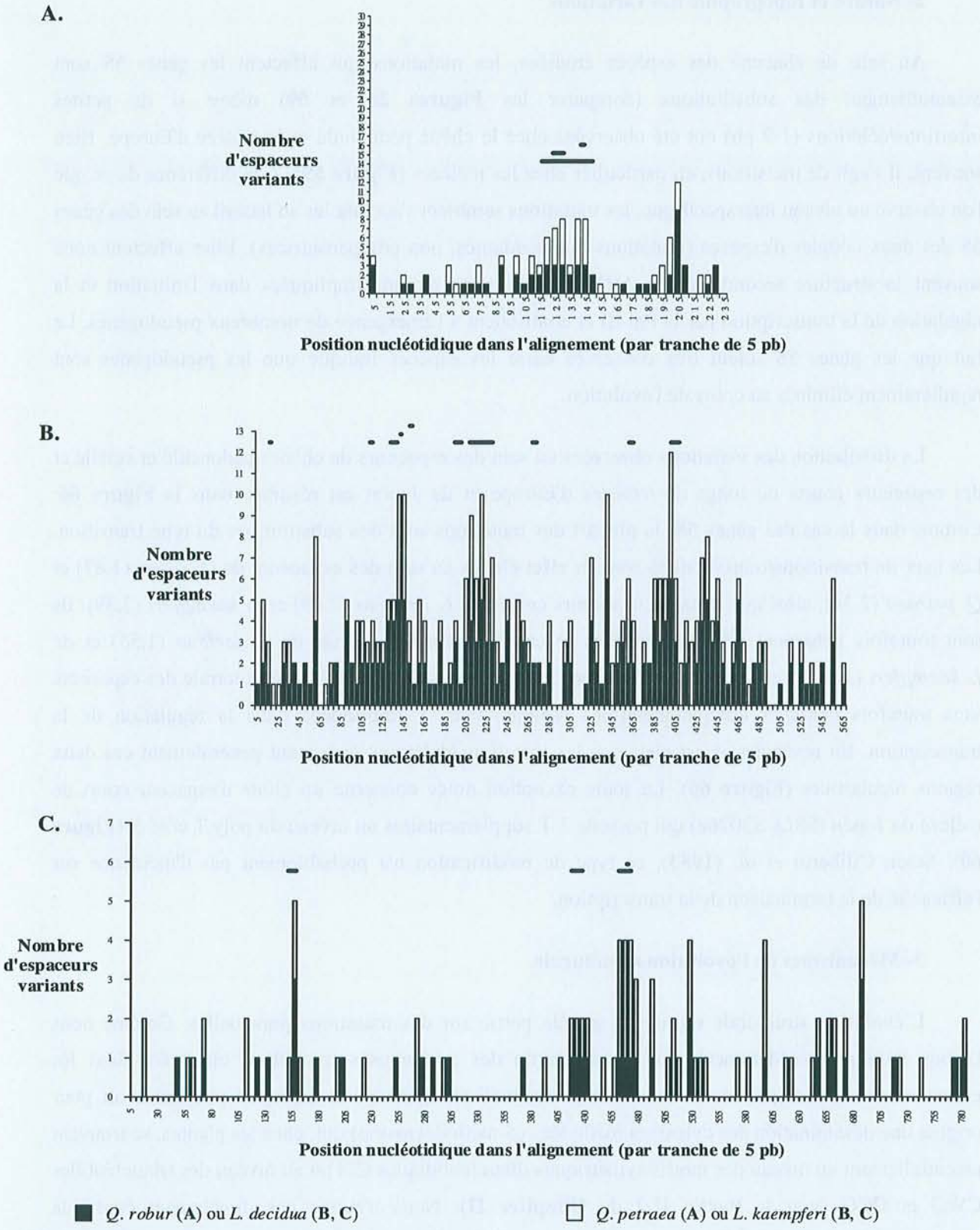



**Figure 66.** Distribution des variations de séquence au sein de l'unique classe d'espaceurs des chênes pédonculé et sessile (**A,** 30 séquences) et des 2 classes d'espaceurs courts (**B,** 13 séquences) ou longs (**C,** 7 séquences) des mélèzes d'Europe et du Japon.
Le nombre d'espaceurs porteurs d'une ou plusieurs différences nucléotidiques par rapport à la séquence consensus majoritaire est indiqué en ordonnée pour chaque intervalle de 5 pb indiqués en abscisse (voir **Figures 29**, **60** et **62**). Les traits noirs indiquent les positions et tailles approximatives des délétions et/ou insertions (> 1 pb).

**2- Nature et topographie des variations**

Au sein de chacune des espèces étudiées, les mutations qui affectent les gènes 5S sont essentiellement des substitutions (comparer les **Figures 26** et **59**) même si de petites insertions/délétions (1-2 pb) ont été observées chez le chêne pédonculé et le mélèze d'Europe. Bien souvent, il s'agit de transitions, en particulier chez les mélèzes (**Figure 59**). A la différence de ce que l'on observe au niveau interspécifique, les mutations semblent s'accumuler au hasard au sein des gènes 5S des deux couples d'espèces (mutations indépendantes, non compensatrices). Elles affectent donc souvent la structure secondaire des ARNr 5S et/ou les régions impliquées dans l'initiation et la régulation de la transcription par la Pol III et contribuent à l'émergence de nombreux pseudogènes. Le fait que les gènes 5S soient très conservés entre les espèces indique que les pseudogènes sont régulièrement éliminés au cours de l'évolution.

La distribution des variations observées au sein des espaceurs de chênes pédonculé et sessile et des espaceurs courts ou longs de mélèzes d'Europe et du Japon est résumée dans la **Figure 66**. Comme dans le cas des gènes 5S, la plupart des mutations sont des substitutions du type transition. Les taux de transitions/transversions sont en effet élevés au sein des espaceurs de *Q. robur* (1,87) et *Q. petraea* (2,34), ainsi que dans les espaceurs courts de *L. decidua* (2,38) et *L. kaempferi* (3,39). Ils sont toutefois nettement plus faibles dans le cas des espaceurs longs de *L. decidua* (1,55) et de *L. kaempferi* (1,28). Les substitutions ont tendance à s'accumuler dans la zone centrale des espaceurs sans toutefois épargner complètement les régions 5' et 3' impliquées dans la régulation de la transcription. En revanche, il est clair que les insertions/délétions épargnent généralement ces deux régions régulatrices (**Figure 66**). La seule exception notée concerne un clone d'espaceur court de mélèze du Japon (5S*Lk*-S3076a) qui possède 3 T supplémentaires au niveau du poly T côté 5' (**Figure 60**). Selon Ciliberto *et al.* (1983), ce type de modification n'a probablement pas d'incidence sur l'efficacité de la terminaison de la transcription.

**3- Mécanismes de l'évolution structurale**

L'évolution structurale repose en grande partie sur des mutations ponctuelles. Comme nous l'avons indiqué précédemment, une bonne partie des nombreuses transitions observées dans les espaceurs des chênes pédonculé et sessile et des mélèzes d'Europe et du Japon pourrait avoir pour origine une désamination des cytosines méthylées (5-méthylcytosine) qui, chez les plantes, se trouvent essentiellement au niveau des motifs symétriques dinucléotidiques CG ou au niveau des trinucléotides CWG et CCG (voir la **Partie II-2** du **Chapitre II**). Nous n'avons pas directement étudié la méthylation de l'ADNr 5S. En revanche, nous avons été confronté à plusieurs reprises à des difficultés de digestion des unités avec des enzymes sensibles à la méthylation des cytosines, comme *Pst* I, *Sma* I et *Ksp* I. Il est donc possible, comme chez la plupart des plantes, que l'ADNr 5S des chênes et des mélèzes soit très méthylé au niveau des cytosines, avec pour conséquence l'obtention de taux de transitions/transversions élevés. Des travaux de séquençage génomique permettant d'étudier

l'importance et la distribution des cytosines méthylées dans les unités d'ADNr 5S (voir Fulnecek *et al.* 1998) pourraient confirmer cette hypothèse et expliquer pourquoi la classe d'unités longues des mélèzes présente des taux de transitions/transversions aussi faibles.

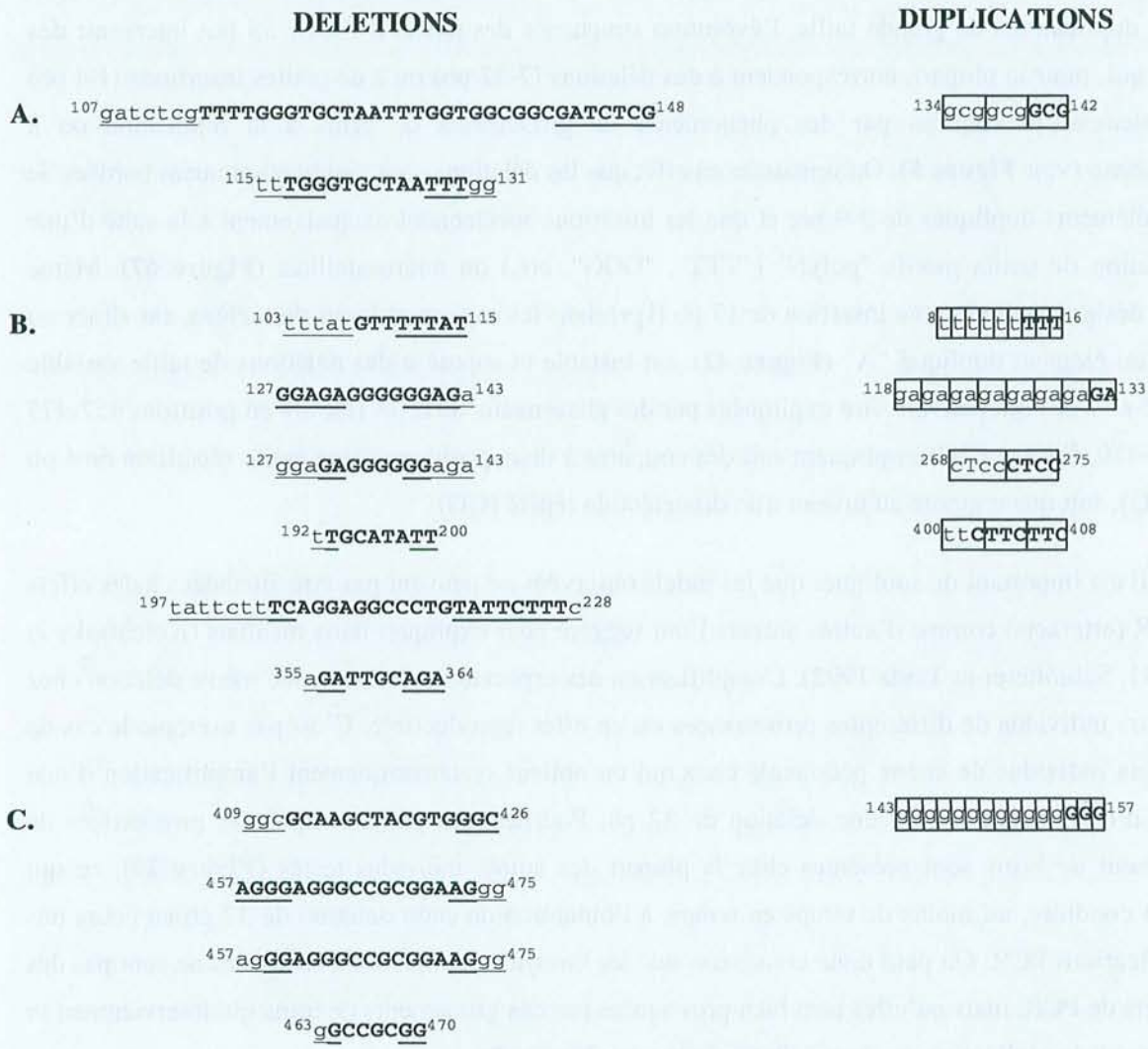



**Figure 67**. Liste des délétions et petites duplications pouvant être attribuées à des phénomènes de glissement de brins au sein des espaceurs de chênes pédonculé et sessile (**A**) et des espaceurs courts (**B**) ou longs (**C**) de mélèzes d'Europe et du Japon. Les motifs délétés ou insérés sont indiqués en majuscules et en gras. Les petits éléments répétés promoteurs de glissements de brins sont respectivement soulignés (délétions) ou encadrés (duplication). La position des séquences dans les alignements (**Figures 29**, **60** et **62**) est indiquée (en pb).

Concernant les mutations de plus grande ampleur, les observations réalisées permettent de conclure que les unités d'ADNr 5S évoluent en longueur surtout par duplication en tandem et délétion. D'importantes duplications en tandem (environ 50 pb) ont ainsi été mises en évidence au sein des espaceurs longs de mélèzes d'Europe et du Japon, côté 5' d'une part (éléments $\alpha$ et $\beta$, peut-être $\chi$ dans certains cas) et dans la partie centrale (éléments A-E), d'autre part (voir les **Figures 42** et **64**). De même, le séquençage de 3 espaceurs de *Quercus suber* et de *Quercus libani* a révélé la présence d'une

grande duplication en tandem de 83 pb qui occupe, là encore, l'extrémité 5' des espaceurs (**Figure 31**). Le mécanisme généralement proposé pour des duplications en tandem d'une telle importance est le crossing-over inégal dont la réalisation ne dépendrait que de la présence de petites régions similaires relativement proches (**Figure 6**, consulter également les **Parties II-2c** et **II-3** du **Chapitre I**). En plus de ces duplications de grande taille, l'évolution structurale des unités d'ADNr 5S fait intervenir des indels qui, pour la plupart, correspondent à des délétions (7-32 pb) ou à de petites insertions (1-4 pb) probablement provoquées par des phénomènes de glissements de brins à la réplication ou à l'interphase (voir **Figure 5**). On remarque en effet que les délétions sont systématiquement bordées de petits éléments dupliqués de 2-9 pb, et que les insertions surviennent exclusivement à la suite d'une duplication de petits motifs "polyN" ("TTT", "GGG", etc.) ou microsatellites (**Figure 67**). Même l'indel désignée comme une insertion de 17 pb ($i_{17}$) dans les espaceurs longs de mélèze, car située au sein d'un élément dupliqué "A" (**Figure 42**), est instable et sujette à des délétions de taille variable (17, 15 et 7 pb) qui peuvent être expliquées par des glissements de brins (motifs en positions 457-475 et 463-470, **Figure 67C**) impliquant soit des coupures à deux positions d'une petite répétition de 4 pb (AGGG), soit une coupure au niveau d'un dinucléotide répété (GG).

Il est important de souligner que les indels observées ne peuvent pas être attribuées à des effets de PCR (artefacts) comme d'autres auteurs l'ont suggéré pour expliquer leurs résultats (Kolchinsky *et al.* 1991, Schlötterer et Tautz 1992). L'amplification des espaceurs présentant une même délétion chez plusieurs individus de différentes provenances est en effet reproductible. C'est par exemple le cas de plusieurs individus de chêne pédonculé chez qui on obtient systématiquement l'amplification d'une classe d'espaceurs sièges d'une délétion de 32 pb. Pourtant, les petites séquences promotrices de glissement de brins sont présentes chez la plupart des autres individus testés (**Figure 29**), ce qui devrait conduire, au moins de temps en temps, à l'obtention de cette délétion de 32 pb au cours des amplifications PCR. On peut donc considérer que les insertions et délétions observées ne sont pas des artefacts de PCR, mais qu'elles sont bien provoquées par des glissements de brins qui interviennent *in vivo* lors de la réplication ou durant l'interphase (voir **Figure 5**).

**4- Aspects fonctionnels**

Les séquences consensus des gènes 5S des chênes pédonculé et sessile et des mélèzes d'Europe et du Japon (**Figure 65**) sont très similaires aux gènes 5S des autres plantes supérieures (Szymanski *et al.* 1998) et parfaitement compatibles avec le modèle de structure secondaire et tertiaire des ARNr 5S proposé par Barciszewka *et al.* (1994). La comparaison des **Figures 17** et **40** ainsi que la consultation du **Tableau 31** montrent clairement que les différences de structure observées entre angiospermes et gymnospermes sont très faibles (quelques différences au sein des régions bicaténaires).

Les motifs de l'espaceur potentiellement impliqués dans la régulation de l'initiation et de la terminaison de la transcription des gènes 5S (voir **Figures 14** et **15**) apparaissent à des positions semblables mais leur composition est assez différente entre chênes et mélèzes. Le signal de

terminaison de la transcription (poly T, **Figure 68**) est ainsi assez long (14-18 T) et segmenté en trois parties de taille inégale chez *Q. robur*, *Q. petraea*, *Q. suber* et *Q. libani* tandis qu'il est court (6-7 T) et ininterrompu chez *L. decidua* et *L. kaempferi*. On peut en déduire que si le poly T est nécessaire pour la terminaison de la transcription, sa taille n'est probablement pas le seul facteur qui détermine son efficacité. Selon Bogenhagen et Brown (1981), les bases adjacentes C et G qui flanquent le poly T dans tous les cas (**Figure 68**) créeraient un environnement favorable à son expression.

```
                                +1                                  +35
Quercus robur                   CTCTTTTTCTTCGGATTTTTTTT---GTCGCCGGG
Quercus petraea

Quercus suber                   CTCTTTTTCTTCGCATTTTTTT------CGTTGGG

Quercus libani (S)              CTCTTTTTCTTCGCATTTTTTTTTTTTCTCGTTGGG
Quercus libani (L)              CTCTTTTTCTTCGCATTTTTTTTTTTT--CGTTGGG

Larix decidua     (S)           CACCCCCTTTTTTGCACGCCTCCGAGATGAGGGAA
Larix kaempferi

Larix decidua     (L)           CACCC-TTTTTTTGCACGCATCCGAGGTGAGGGAA
Larix kaempferi
```

**Figure 68**. Signaux présumés de terminaison de la transcription observés en aval des gènes 5S au sein des espaceurs de *Quercus robur*, *Quercus petraea*, *Quercus suber*, *Quercus libani*, *Larix decidua* et *Larix kaempferi*. Les régions correspondantes des espaceurs sont alignées. S et L désignent respectivement les classes d'espaceurs courts ou longs. Le motif poly T potentiellement fonctionnel est indiqué en gras.

```
                                        -24          -11          -1
Quercus robur                   CGGATCCTTT-ATAAGGATTTAAACGCGCTTATGCTGAT
Quercus petraea

Quercus suber                   CGGATCCTTTTATAAGAATTAAACGCGCTTATGCTGAT
Quercus libani (S+L)            CGGATCCTTTTATAAGAATTAAACGCGCTTATGCTGAT
                                        -26          -11          -1
Larix decidua     (S)           AAGAGCCCTTGAAGAGGGGGAGGAGGAGTGATCCTTGC
Larix kaempferi

Larix decidua     (L)           AACATCCCTTGAAGAGGGGGAGGAGGAGTGATCCTTGC
Larix kaempferi
```

**Figure 69**. Signaux présumés de régulation de l'initiation de la transcription observés en amont des gènes 5S au sein des espaceurs de *Quercus robur*, *Quercus petraea*, *Quercus suber*, *Quercus libani*, *Larix decidua* et *Larix kaempferi*.
S et L désignent respectivement les classes d'espaceurs courts ou longs. La région riche en AT ainsi que la base en position -1 sont encadrées. La région riche en GC (chênes) ou en GA (mélèzes) apparaît en gris.

De la même façon, il n'est pas non plus possible d'identifier des motifs conservés en position et en séquence au sein des espaceurs, en amont des gènes 5S (**Figure 69**). Comme chez d'autres angiospermes (**Figure 14**), on reconnaît facilement chez les chênes les motifs riches en AT (position -24) et en GC (position -11) désignés par Venkateswarlu *et al.* (1991) comme éléments régulateurs de l'initiation de la transcription. Chez les mélèzes en revanche, la région riche en AT (position -26) semble atypique (motif TTGAA) et la région riche en GC (position -11) est remplacée par une région plus longue, riche en GA qui est également observée chez d'autres gymnospermes comme *Pinus radiata*, *Picea glauca* et *Pseudotsuga menziesii* (**Figure 14**). Il semble donc qu'il y ait une différence d'organisation fonctionnelle notable entre angiospermes et gymnospermes. On remarque par ailleurs que l'on trouve bien une cytosine en position -1 par rapport au gène 5S chez les mélèzes (**Figure 69**) comme chez de nombreuses autres plantes supérieures (**Figure 14**). Chez les 4 chênes étudiés, on trouve en revanche à cette position une thymine (**Figure 69**). La présence d'une base C en position -1 n'a donc probablement pas l'importance qu'on lui prête habituellement (base importante pour l'initiation de la transcription selon Venkateswarlu *et al.* 1991).

Seules des expériences de mutagénèse dirigée et de transcription *in vitro*, ou *in vivo* après transformation génétique, pourraient permettre d'apprécier l'importance par rapport à l'ICR des différents éléments de l'espaceur supposés intervenir dans la régulation de la transcription. A notre connaissance, les données expérimentales de ce type concernent surtout les espèces animales et sont extrêmement limitées chez les plantes. Une seule étude, portant sur le riz (Hariharan *et al.* 1987), a montré qu'il est possible de transcrire *in vitro* un gène 5S dissocié de son espaceur et tronqué à son extrémité 3' de 30 pb. Les éléments régulateurs de l'espaceur semblent donc jouer un rôle relativement secondaire dans l'initiation de la transcription.

## II- Organisation des unités d'ADNr 5S

Comme chez la plupart des plantes supérieures, nous avons montré que l'ADNr 5S des chênes pédonculé et sessile et des mélèzes d'Europe et du Japon est organisé en longues séries d'unités répétées en tandem. Cette organisation est illustrée notamment par l'apparition, soit d'échelles de multimères après analyse RFLP de l'ADN génomique (voir **Figures 25** et **45**), soit de fragments multimériques après amplification par PCR.

### 1- Nombre d'unités

Les estimations du nombre d'unités d'ADNr 5S que nous avons obtenues pour un chêne pédonculé (1000-2000 copies par génome 2C) et un mélèze d'Europe (1500-3000 copies par génome 2C de chacune des classes d'unités courtes et longues) sont du même ordre de grandeur que celles données pour d'autres plantes supérieures (**Tableau 13**). L'ADNr 5S représenterait donc approximativement 0,02 à 0,04 % du génome du chêne pédonculé (1,84 pg/2C, **Tableau 19**) et 0,01 à 0,02 % du génome du mélèze d'Europe (19,70 pg/2C, **Tableau 29**). Ce type de données doit

cependant être considéré avec prudence. En effet, il est bien connu tout d'abord que le nombre de copies varie beaucoup au sein des espèces, aussi bien entre les populations qu'au sein des populations, en raison du brassage chromosomique et des mécanismes de l'évolution concertée (voir la **Partie III-1** du **Chapitre II**). De plus, la détermination du nombre de copies par dot blot ou Southern blot est une entreprise délicate qui se traduit par une grande incertitude sur les valeurs obtenues. Ces méthodes nécessitent ainsi de connaître avec précision la taille du génome des espèces étudiées. Or, les chiffres disponibles pour les chênes et les mélèzes diffèrent significativement selon les auteurs (voir les **Tableaux 19** et **29**). Il est nécessaire ensuite de pouvoir quantifier avec précision les extraits d'ADN génomique à tester par des méthodes spectrophotométriques ou plus souvent fluorimétriques. L'utilisation de ces dernières exige de disposer d'un ADN étalon hautement purifié possédant une composition en base GC aussi proche que possible de l'ADN à tester. Nous avons utilisé un ADN génomique commercial hautement purifié (extrait du thymus de veau) dont la composition en base GC (42 %) est similaire à celle des chênes pédonculé et sessile bien que, là aussi, les estimations varient selon les auteurs (38,8-42 %, **Tableau 19**). A notre connaissance, la composition en bases GC du génome des mélèzes d'Europe et du Japon est en revanche encore inconnue. L'utilisation par défaut d'ADN de thymus de veau comme référence est donc susceptible d'avoir entraîné une erreur d'estimation non négligeable pour ces 2 espèces. Pour envisager une étude plus large du nombre de copies d'ADNr 5S chez les chênes et les mélèzes, il conviendrait donc d'utiliser comme standard des ADN génomiques ultrapurs préparés directement à partir de ces espèces. Une solution alternative, peut-être plus simple à utiliser que des méthodes basées sur une hybridation quantitative, consisterait à mettre au point la coamplification des unités d'ADNr 5S avec une quantité connue d'un segment d'ADN de référence de taille voisine. La comparaison par densitométrie de l'intensité de coloration après électrophorèse sur gel d'agarose ou gel d'acrylamide permettrait d'obtenir une évaluation précise du nombre d'unités (Deng *et al.* 1993).

Malgré ces difficultés méthodologiques, l'analyse des variations du nombre de copies d'ADNr 5S, au sein, et entre les populations de chênes ou de mélèzes, demeure intéressante à envisager. Plusieurs auteurs (voir Govindaraju et Cullis 1992) ont en effet détecté des variations du nombre de copies d'ADNr clairement associées à des différences morphologiques et/ou liées à différents stress environnementaux. Chez le chêne par exemple, une telle étude réalisée selon différents transects permettrait peut-être d'apporter des éléments d'information intéressants concernant une éventuelle structuration géographique des variations des séries d'ADNr 5S. Le mélange en quantité équimoléculaire des ADN génomiques d'une dizaine d'individus par population permettrait de s'affranchir en partie des variations entre les individus.

**2- Structure et composition des séries**

Les différents éléments d'information obtenus au cours des investigations concernant la structure et la composition des séries d'unités d'ADNr 5S montrent que les séries sont hétérogènes,

tant au niveau des gènes 5S qu'au niveau des espaceurs. Malgré le faible nombre de séquences de gènes 5S analysées (3 par espèce, de 3 individus différents), nous avons pu détecter à la fois chez le chêne pédonculé et chez le mélèze d'Europe (**Figures 26** et **59**) la présence de pseudogènes qui portent des mutations, très probablement défavorables, dans la région de contrôle de l'initiation de la transcription (ICR), dans les zones d'interaction entre les ARNr 5S et le facteur de transcription TFIIIA (formation de la particule RNP 7S, **Figures 16**), ou bien encore au sein des régions impliquées dans les structures secondaire et tertiaire des ARNr 5S (**Figure 17**). Si l'on admet, conformément à la théorie du "molecular drive" (Dover 1982) que la proportion de pseudogènes dans une population d'individus est approximativement égale à la proportion de pseudogènes présents dans l'ADNr 5S d'un génome individuel (voir la **Partie II-2-d** du **Chapitre I**), ce résultat suggère que les pseudogènes sont très nombreux au sein des génomes individuels. L'analyse d'un plus grand échantillon de séquences (au moins une centaine) permettrait d'estimer la fréquence des pseudogènes dans un génome et donc d'avoir une idée du seuil critique d'accumulation à partir duquel ils réduisent l'efficacité d'une série et provoquent l'élimination de l'individu porteur par la sélection naturelle (voir **Figure 20**). Une telle donnée serait particulièrement intéressante à prendre en compte dans les études théoriques de l'évolution des familles de séquences répétées (Kellogg et Appels 1995, Cronn *et al.* 1996).

Au niveau des espaceurs, les séries sont particulièrement hétérogènes chez les mélèzes d'Europe et du Japon. Les unités appartenant aux deux classes courtes (environ 650 pb) et longues (environ 850 pb) alternent en effet pour la plupart au sein de séries composites (**Figure 45B**) dont l'établissement remonte approximativement à la période de diversification du genre *Larix* (**Partie IIB** du **Chapitre IV**). Une telle alternance n'a été que très rarement observée chez les plantes. Sa présence à grande échelle chez les mélèzes laisse donc supposer qu'elle est nécessaire à la préservation de l'intégrité fonctionnelle des séries. Cette organisation alternée peut néanmoins, et paradoxalement, être considérée comme homogène si l'on admet que le motif dimérique de 1450 pb formé des unités courtes et longues constitue l'unité répétée. Une situation tout à fait similaire a été décrite pour l'ADNr 5S des ovocytes de *Xenopus borealis* dont les unités de taille importante (1000-2000 pb) apparaissent dimériques (Peterson *et al.* 1980). Elles se composent en effet d'un gène 5S, d'un espaceur court (80 pb), d'une seconde copie de gène 5S et d'un espaceur plus long, de taille variable. L'analyse fonctionnelle a finalement montré que l'un des deux gènes 5S de chaque unité est un pseudogène.

Dans le cas du mélèze, nous n'avons pas pu formellement conclure à la présence au sein des unités d'une combinaison gène/pseudogène car les couples d'amorces utilisés n'étaient pas adéquats (voir **Discussion** du **Chapitre IV**). Il serait important de reprendre l'étude structurale, et surtout fontionnelle, du motif dimérique de 1450 pb pour mieux comprendre son origine et les raisons de son maintien au cours de l'évolution.

Au-delà des particularités de l'organisation générale des unités, l'étude de la structure et de la composition des séries nous a aussi permis de révéler la présence de nombreuses unités variantes aussi bien au sein des séries des chênes pédonculé et sessile que chez les mélèzes d'Europe et du Japon. L'analyse des croisements en particulier (voir les **Partie III** des **Chapitres III** et **IV**), a montré que les séries d'ADNr 5S à chaque locus sont composites (ou mosaïques), c'est-à-dire constituées de multiples variants alléliques. Des différences de taille importantes se superposant à une microhétérogénéité de séquence ont été mises en évidence au niveau des espaceurs chez les mélèzes. Chez les chênes en revanche, seule une microhétérogénéité de séquence a été détectée (**Figure 28**). Les profils d'amplification obtenus sur gel d'acrylamide pour les mélèzes sont difficilement analysables en raison de leur complexité (données non présentées). Cette difficulté suggère également une très forte microhétérogénéité de séquence superposée à la variation de taille.

Si l'on considère l'existence d'un unique locus d'ADNr 5S chez les chênes pédonculés et sessiles ainsi que chez les mélèzes d'Europe et du Japon, l'analyse RFLP à l'aide de sondes marquées couplée à l'électrophorèse en champ pulsé pourrait s'avérer intéressante pour tenter de quantifier la microhétérogénéité des séries de l'ADNr 5S des chênes pédonculés et sessiles (taille estimée : 175-350 kpb) et des mélèzes d'Europe/du Japon (taille estimée : 1125-2250 kpb). L'analyse des profils de digestion en terme de nombre, taille et intensité des fragments obtenus permettrait ainsi d'évaluer la fréquence de différents sites de restriction au sein des séries de l'ADNr 5S (Röder *et al.* 1995).

**3- Héritabilité des variants alléliques, nombre de loci et cartographie génétique**

L'étude de deux familles issues d'un croisement intraspécifique de chênes pédonculés et d'un croisement interspécifique de mélèzes d'Europe et du Japon a permis de montrer que la plupart des marqueurs de l'ADNr 5S amplifiés par PCR à l'aide d'amorces universelles présentent un mode d'hérédité mendélien (**Figures 33** et **58**) gouverné par un unique locus. Dans les deux croisements étudiés, des distorsions de ségrégation plus ou moins importantes de quelques marqueurs ont toutefois été notées. Elles sont probablement explicables par les mécanismes de l'évolution concertée des unités d'ADNr 5S (Dover 1982) même si, pour les chênes, l'éventuelle présence de chromosomes B porteurs d'ADNr 5S pourrait être une piste envisageable (Ohri et Ahuja 1991). A ces difficultés s'ajoute l'impossibilité de discriminer la variation allélique *cis* de la variation allélique *trans* sur la simple base des profils d'amplification. Malgré l'importance des variations qu'elle recèle, l'utilisation de cette famille multigénique pour des études de génétique des populations s'avère donc délicate. Pour en exploiter au moins en partie les possibilités, on pourrait envisager de cloner et de séquencer les régions de l'ADN génomique qui flanquent les séries de l'ADNr 5S (supposées conservées entre les individus d'une même espèce) afin de synthétiser de nouvelles amorces. Leur utilisation en combinaison avec les amorces situées dans les unités permettrait d'amplifier les bordures internes 5' et 3' des séries. Ces régions sont en effet considérées comme des zones d'accumulation privilégiées des unités variantes car, par rapport aux régions plus centrales des séries, elles sont moins

fréquemment soumises aux mécanismes d'homogénéisation par crossing-over inégal ou conversion génique ectopique (voir la **Partie IIA-5-c** du **Chapitre IV**). Si l'on prend soin de se placer dans des conditions d'amplification qui limitent l'apparition des multimères (saturation en amorces), on pourrait alors retrouver une situation analogue à celle d'un locus monogénique avec l'avantage de pouvoir distinguer le polymorphisme *trans* (allélique au sens génétique) du polymorphisme *cis* au niveau de chacune des deux "régions" 5' et 3' du locus. Cette approche pourrait être progressivement étendue aux dimères et trimères obtenus à partir de ces bordures 5' et 3' dans des conditions de PCR qui les favorisent (déficit en amorces se fixant dans les unités d'ADNr 5S).

L'analyse des deux croisements étudiés (chênes pédonculés et mélèzes d'Europe et du Japon) n'a révélé qu'un seul locus "5SA". Ces résultats ont récemment été confirmés chez le chêne pédonculé par Zoldos *et al.* (1999) qui n'ont détecté qu'un seul locus d'ADNr 5S par hybridation *in situ*. Ces auteurs ont montré que ce locus est placé en position paracentromérique sur la seconde paire de chromosomes (par la taille) comme chez les 11 autres espèces de chêne testées. Cette paire de chromosomes N°2 peut donc être clairement associée aux groupes de liaison G5F (arbre 3P) et G5M (arbre A4) de la carte génétique de chêne pédonculé sur lesquels a été cartographié le locus 5SA (Barrénèche *et al.* 1998). A notre connaissance, c'est la première association établie entre groupes de liaison et chromosomes chez le chêne. L'intérêt d'une telle information est évidente dans la perspective des futurs programmes d'amélioration génétique s'appuyant sur la localisation physique d'un gène d'intérêt ou de QTL cartographiés sur les groupes de liaison.

Dans le cas du mélèze, les données obtenues confirment les travaux de Lubaretz *et al.* (1996) qui ont identifié par hybridation *in situ* un unique locus d'ADNr 5S sur la paire de chromosomes III du mélèze d'Europe. L'association de cette paire de chromosomes III à l'un des groupes de liaison établis par Arcade *et al.* (2000) pour un mélèze d'Europe et un mélèze du Japon (famille 179) devrait être facilement réalisée sur la base de l'analyse de ségrégation des marqueurs de l'ADNr 5S que nous avons réalisée dans cette même famille.

## III- Polymorphisme et évolution des unités d'ADNr 5S

### 1- Diversité moléculaire intraspécifique et différenciation interspécifique

La diversité moléculaire intraspécifique des unités d'ADNr 5S (gènes 5S + espaceur) des chênes pédonculé et sessile et des mélèzes d'Europe et du Japon a été étudiée au niveau intragénomique (au sein d'un individu) et entre les individus. Bien que la présence d'un unique locus d'ADNr 5S chez ces 4 espèces constitue une situation simple, l'importance de la variation allélique *cis* observée (voir les **Parties III** des **Chapitres III** et **IV**) ne permet pas d'accéder à la variation allélique *trans* et donc de calculer des fréquences alléliques. Le recours à une approche basée sur la quantification de la diversité phénotypique n'est pas non plus apparu satisfaisant. Par exemple, les estimations de diversité phénotypique de Shannon obtenues pour les espaceurs de chêne pédonculé

(0,779) et de chêne sessile (0,899) sont apparemment très élevées. Il n'est cependant pas possible de conclure car ce paramètre est extrêmement sensible à la taille de l'échantillon et à la sensibilité de la méthode mise en œuvre pour détecter les variations. Les chiffres obtenus sont donc difficilement comparables aux résultats obtenus pour les familles de séquences répétées d'autres espèces. On peut seulement en déduire que les espaceurs de chêne sessile sont plus variables que les espaceurs de chêne pédonculé au locus 5SA.

Pour quantifier la variation au locus d'ADNr 5S, l'analyse de séquences nucléotidiques nous a donc semblé être la méthode la plus appropriée. Nous avons choisi d'effectuer un séquençage systématique des différents variants électrophorétiques repérés sur gel d'agarose et/ou d'acrylamide à partir d'un échantillon réduit d'individus de différentes provenances, 3 par espèce pour la région du gène 5S et 10 (chênes pédonculés ou sessile) ou 3 (mélèzes d'Europe ou du Japon) par espèce pour la région de l'espaceur. Pour chaque classe de variants électrophorétiques, les clones analysés ont été sélectionnés au hasard. Ceci nous a permis d'accéder à la diversité nucléotidique et de nous soustraire au moins partiellement aux effets de la taille de l'échantillon (Nei 1987).

Les résultats ont tout d'abord révélé une importante diversité moléculaire intragénomique des espaceurs. La différence nucléotidique moyenne entre espaceurs variants d'un même individu est en effet de 0,058 ("gaps" inclus) ou 0,022 ("gaps" exclus) pour le chêne pédonculé et 0,053 ("gaps" inclus") ou 0,017 ("gaps" exclus) pour le chêne sessile. Dans le cas des espaceurs courts des mélèzes, elle atteint 0,093 ("gaps" inclus) ou 0,096 ("gaps exclus") chez le mélèze d'Europe et 0,072 ("gaps" inclus") ou 0,077 ("gaps exclus") chez le mélèze du Japon. La diversité moléculaire intraspécifique (variation intragénomique + variation entre les individus) apparaît très similaire et suggère que l'essentiel de la variation des espaceurs réside donc au niveau intragénomique. En moyenne, elle est en effet de 0,057 ("gaps inclus) ou 0,023 ("gaps exclus") pour le chêne pédonculé, et 0,047 ("gaps inclus") ou 0,023 ("gaps exclus") pour le chêne sessile. De même, la diversité moléculaire intraspécifique des espaceurs courts de mélèze a été estimée à 0,098 ("gaps" inclus) ou 0,097 ("gaps" exclus) pour le mélèze d'Europe et 0,067 ("gaps" inclus) ou 0,063 ("gaps" exclus) pour le mélèze du Japon. Dans le cas des espaceurs longs au sein desquels un élément conservé entre gymnospermes a été détecté (5SmsCR1, **Partie II-B** du **Chapitre IV**), les estimations obtenues sont nettement plus faibles, de l'ordre de 0,015 pour le mélèze d'Europe et de 0,037 pour le mélèze du Japon. En considérant l'alternance des unités courtes et longues, il pourrait être intéressant de compléter ces données en séquençant au hasard quelques unités dimériques d'environ 1520 pb (unités courtes et longues).

Aucune variation significative n'a pu être en revanche mise en évidence au niveau du gène 5S. La diversité nucléotidique ("gaps" inclus") calculée à partir de 3 séquences de gènes 5S issues de 3 individus par espèce est de 0,017 pour le chêne pédonculé, 0,005 pour le chêne sessile (0,014 pour l'ensemble des deux espèces), 0,066 pour le mélèze d'Europe et 0,006 pour le mélèze du Japon (0,050

pour l'ensemble des deux espèces). Les chiffres relativement élevés obtenus pour le chêne pédonculé et surtout le mélèze d'Europe peuvent être expliqués par la présence de pseudogènes.

Les valeurs obtenues aussi bien pour les gènes 5S que pour les espaceurs chez les mélèzes et les chênes sont du même ordre de grandeur que celles calculées pour d'autres espèces à temps de génération beaucoup plus court, comme les *Triticeae* (Kellogg et Appels 1995) ou différentes espèces de *Gossypium* (Cronn *et al.* 1996). Nos résultats indiquent donc que les séries d'ADNr 5S des espèces à cycle court et à cycle long présentent des niveaux de variation intragénomique tout à fait comparables. Chez les plantes à cycle court, l'espaceur constitue généralement un excellent marqueur de discrimination des espèces (Sastri *et al.* 1992) car le niveau de divergence interspécifique y est corrélé au polymorphisme intraspécifique (Kellogg et Appels 1995). Il est donc remarquable de constater que malgré cette similitude des niveaux de variation avec les espèces à court cycle, aucune différence qualitative n'a pu être enregistrée entre chêne pédonculé et chêne sessile, d'une part, et entre mélèze d'Europe et mélèze du Japon, d'autre part. Les différences quantitatives sont elles-aussi très faibles : un seul marqueur présentant une différence de fréquence significative entre chêne pédonculé et chêne sessile ("*e*") a pu être détecté.

Dans le cas des chênes pédonculé et sessile, l'hybridation interspécifique pourrait être évoquée pour expliquer cette situation (espèces sympatriques) mais cette hypothèse ne peut évidemment pas être retenue pour les mélèzes d'Europe et du Japon (espèces allopatriques). Une alternative pourrait être d'admettre que variation intraspécifique et divergence interspécifique au locus d'ADNr 5S sont découplées chez les chênes pédonculé et sessile et les mélèzes d'Europe et du Japon. L'hypothèse est sans doute plausible pour la région du gène 5S qui apparaît aussi variable que la région de l'espaceur au sein des espèces, mais en revanche très conservée entre les espèces car soumise à la pression de la sélection naturelle. Elle pourrait également être retenue pour les espaceurs longs de mélèzes qui renferment un élément conservé (5SmsCR1, probablement un pseudogène) chez d'autres gymnospermes. En revanche, il semble difficile de l'appliquer aux espaceurs de chênes et aux espaceurs courts de mélèzes qui ne contiennent aucun éléments particulier en dehors des éléments de régulation de l'initiation et de la terminaison de la transcription.

Il est cependant évident que la dynamique de propagation d'un variant au sein d'une famille de gènes (homogénéisation intragénomique ou évolution concertée) et simultanément dans une population au cours des générations (fixation) dépend essentiellement (Dover 1982) de la fréquence des échanges interchromosomiques (entre chromosomes homologues ou non homologues à la méiose, donc à chaque génération) et de l'efficacité du brassage chromosomique (fonction du temps de génération). L'hétérogénéité des séries d'ADNr 5S des espèces à cycle long reflèterait ainsi surtout la faible fréquence des événements d'homogénéisation intragénomique tandis que chez les espèces à cycle court elle traduirait plutôt l'efficacité du brassage chromosomique. Pour les chênes pédonculé et sessile et pour les mélèzes d'Europe et du Japon qui sont des espèces à cycle long, il est donc

probable que le nombre de générations écoulées depuis leur divergence est insuffisant pour permettre l'émergence de différences interspécifiques au locus d'ADNr 5S. L'importance de la variation intraspécifique semble en tout cas avoir une assez mauvaise valeur prédictive de la divergence interspécifique chez ces espèces.

**2- Intensité de l'évolution concertée et conséquences**

Bien que la relative conservation de la taille des unités chez les chênes pédonculé et sessile (317-353 pb) et les mélèzes d'Europe et du Japon (634-668 pb pour les unités courtes, 858-893 pb pour les unités longues) atteste de son action, l'ensemble des données de séquences dont nous venons de considérer les conséquences en terme de diversité démontrent clairement que l'intensité de l'évolution concertée n'est pas suffisante ou qu'elle n'a pas encore suffisamment joué pour conduire à l'homogénéisation parfaite des séries d'ADNr 5S. L'homogénéisation incomplète des séries d'ADNr 5S est une situation courante chez les plantes à cycle court (revue dans Cronn *et al.* 1996, voir le **Tableau 15**) ce qui suggère que la fréquence et/ou l'amplitude des mécanismes de l'évolution concertée (nombre d'unités impliquées à chaque fois) est relativement faible. Contrairement à la situation rencontrée chez les plantes à cycle court fortement apparentées qui ont été les plus souvent étudiées, les différents gènes ou espaceurs séquencés chez les chênes pédonculé et sessile et les mélèzes d'Europe et du Japon ont un ancêtre commun apparemment plus ancien que l'événement de spéciation qui leur a donné naissance. En effet, nous avons montré que les unités d'une espèce (chêne ou mélèze) ne sont pas plus apparentées entre elles qu'elles ne le sont de celles des autres espèces. L'homogénéisation intragénomique des séries d'ADNr 5S se réalise donc à une vitesse probablement plus faible que les événements de spéciation, sauf à considérer que ces espèces sont encore en cours de spéciation, l'argument principal en ce sens étant qu'elles ont conservé leur compatibilité sexuelle. Dans ce contexte, l'équivalence des niveaux de variation intragénomique et intraspécifique peut s'expliquer par le brassage chromosomique qui intervient à chaque génération, c'est-à-dire beaucoup plus rapidement que les mécanismes de l'évolution concertée. Les effets lents de l'évolution concertée au niveau intragénomique se trouvent ainsi constamment diffusés à l'échelle des populations ("molecular drive", Dover 1982). Sur un plan pratique, ceci signifie qu'une étude moléculaire de diversité moléculaire peut en théorie être réalisée de manière équivalente soit sur un échantillon d'unités choisies au hasard à partir d'un unique individu, soit sur un échantillon d'individus représentatifs de l'espèce (une unité par individu).

Si l'on considère maintenant comment l'évolution concertée joue au sein d'un locus, l'étude des croisements de chênes pédonculés et de mélèzes d'Europe et du Japon a montré qu'il est relativement aisé de détecter des marqueurs spécifiques de chacune des séries au locus d'ADNr 5S. Cela indique clairement que la fréquence des échanges entre chromatides soeurs est supérieure à la fréquence des échanges entre chromosomes homologues. Le modèle d'évolution concerté intralocus (**Figure 10**) proposé par Liao *et al.* (1997) semble donc applicable pour au moins 3 des 4 espèces étudiées (nous n'avons pas de données concernant le chêne sessile).

**3- Phylogenèse des variants alléliques - Paralogie et orthologie**

La phylogénèse des variants alléliques (*cis* et *trans*, indistinctement) séquencés pour les chênes pédonculé et sessile n'a pas révélé de classes d'espaceurs paralogues suffisamment divergentes pour être traitées séparément dans les études phylogénétiques et taxinomiques (**Figure 32**). Seules quelques positions nucléotidiques sont informatives et permettent de conclure que la classe de variant électrophorétique "*f*", présente uniquement à faible fréquence chez le chêne sessile, comporte des caractéristiques nucléotidiques compatibles avec une origine ancienne (polymorphismes partagés avec des séquences de *Quercus libani* et *Quercus suber* appartenant à la section voisine *Cerris* et utilisés comme racine). Les relations phylogénétiques entre les variants n'ont pas pu être mises en relation avec une éventuelle structuration géographique des variations en raison de la taille limitée de l'échantillon. Elles suggèrent toutefois que le variant de type "*e*", présent dans notre échantillon à une fréquence significativement plus élevée chez le chêne sessile (33,5 %) que chez le chêne pédonculé (3,5 %), est d'origine relativement récente. Il est intéressant de remarquer que ce marqueur a essentiellement été détecté chez des individus du sud de l'Europe (9 individus sur 13) à proximité des 3 refuges glaciaires présumés de ces espèces (Péninsule Ibérique, Sud de l'Italie, Balkans, Dumolin-Lapègue *et al.* 1997), ce qui tend à indiquer qu'il pourrait être en phase de diffusion selon un gradient sud-nord. Etant donné qu'il est peu probable qu'il ait pu émerger simultanément dans les 3 refuges à la fois, on pourrait en déduire que son apparition remonte à une période antérieure à la dernière glaciation. Il serait apparu chez le chêne sessile et se trouverait progressivement transmis au chêne pédonculé par hybridation introgressive. Pour conclure de façon plus assurée, il faudrait évidemment envisager une étude plus large incluant la dimension "population".

Concernant les mélèzes d'Europe et du Japon, il est évident que les deux classes très divergentes d'unités courtes (S) et longues (L) doivent être considérées comme des séquences paralogues et donc traitées séparément dans les analyses taxinomiques et phylogénétiques. La phylogenèse des espaceur séquencés pour chacun de ces deux groupes paralogues révèle même des sous-classes relativement divergentes, SA (**Figure 61**) et LA (**Figure 63**), susceptibles d'être traitées, elles-aussi, comme des séquences paralogues en raison de leurs caractéristiques nucléotidiques communes. La faible taille de l'échantillon de séquences ne permet toutefois pas de conclure définitivement sur les relations de paralogie au sein des deux classes d'unités courtes et longues. Dans ces conditions, l'interprétation des relations phylogénétiques entre espèces peut s'avérer très discutable. Les travaux menés depuis 1994 par Baum et Johnson pour tenter de décrire les relations de paralogie des unités d'ADNr 5S au sein du genre *Hordeum* (synthèse dans Baum et Johnson 1998), montrent en effet que la méconnaissance des relations de paralogie/orthologie conduit à des interprétations erronées.

L'alternance des unités courtes et longues, non seulement chez les mélèzes d'Europe et du Japon, mais aussi chez *Larix laricina*; *Larix occidentalis*, *Larix russica*, *Larix gmelinii* (**Partie II-B**

du **Chapitre IV**) peut toutefois offrir une solution avantageuse pour résoudre ces difficultés, au moins en partie. Plutôt que de faire porter les études taxinomiques et phylogénétiques sur l'une ou l'autre des unités courtes ou longues, on pourrait utiliser directement l'unité dimérique entière. Son clonage est tout a fait réalisable. Nous avons ainsi déjà obtenu fortuitement quelques clones composés d'un espaceur court, d'un gène 5S et d'un espaceur long dont la taille (environ 1500 pb) permet, par ailleurs, un séquençage relativement aisé. L'intérêt d'une telle approche est évident puisqu'elle élimine le problème de paralogie entre unités courtes et longues. On trouverait alors une situation plus simple, proche de celle rencontrée chez les chênes pédonculé et sessile. Par rapport aux séquences ITS (1759-1770 pb) analysées par Gernandt et Liston (1999) pour étudier l'évolution du genre *Larix*, l'utilisation des unités dimériques d'ADNr 5S offrent un intérêt comparable au niveau taille, mais elles présentent en plus l'avantage d'être issues d'un seul locus contre 3 pour les unités d'ADNr (Lubaretz *et al.* 1996) ce qui réduit les risques de paralogie entre séquences appartenant à plusieurs loci.

**4- Evaluation du niveau d'information taxinomique fourni par l'ADNr 5S**

Au sein des 4 espèces étudiées, la diversité moléculaire des unités d'ADNr 5S, particulièrement dans la région de l'espaceur, est importante car la variation allélique interchromosomique (*trans*) est conjuguée à la variation allélique intrachromosomique (*cis*). Il s'agit toutefois le plus souvent d'une microhétérogénéité difficile à détecter. Il n'est pas toujours possible en effet de distinguer facilement deux génotypes de chênes (**Figure 27**) ou de mélèzes (**Figure 41**) après amplification et électrophorèse classique sur gel d'agarose. Pour atteindre cet objectif, il peut être nécessaire de digérer les produits de PCR par les enzymes de restriction et/ou de recourir à des méthodes d'électrophorèse sur gel d'acrylamide comme dans le cas des chênes (**Figures 28** et **33**). Certains génotypes de chênes pédonculés ou de chênes sessiles demeurent même indiscernables par ces méthodes résolutives (voir le **Tableau 21**) et les séquences issues d'un individu donné sont souvent très proches des séquences consensus majoritaires (**Figure 29**).

Contrairement à la situation généralement observée chez les espèces à cycle court (**Partie III-3-c** du **Chapitre II**), le niveau d'information taxinomique fourni par les unités d'ADNr 5S n'est pas non plus suffisant pour permettre de distinguer (ou d'identifier) des espèces proches à cycle long encore interfertiles, qu'elles soient sympatriques (chênes pédonculé et sessile) ou allopatriques (mélèzes d'Europe et du Japon). L'analyse des séquences nucléotidiques n'a en effet révélé aucune mutation fixée entre ces espèces, que ce soit au niveau des gènes 5S (**Figures 26** et **59**) ou, de manière plus inattendue, au niveau des espaceurs qui évoluent beaucoup plus rapidement (**Figures 29**, **60** et **62**).

En revanche, les différences nucléotidiques observées entre les séquences d'espaceurs de différentes espèces de *Pinaceae* comme *Larix decidua* ou *Larix kaempferi*, *Pinus radiata*, *Picea glauca* et *Pseudotsuga menziesii* sont déjà si importantes, que les alignements ne peuvent être réalisés et exploités en phylogénie qu'au niveau des bordures 5' et 3' (**Figures 43** et **47**) ou de certaines régions

centrales conservées comme le motif 5SmsCR1 (**Figures 44** et **48**). De même, les amplifications des espaceurs (ou de certaines de leurs régions) de l'ADNr 5S que nous avons réalisées chez les *Pinaceae* révèlent des profils très différents entre les genres *Larix* (**Figures 52** et **54**), *Pinus* (**Figure 56**) et *Picea* (**Figure 55**). Les profils obtenus pour plusieurs mélèzes eurasiens (**Figure 52**) ou américains (**Figure 54**) ainsi que pour différentes espèces d'épicéas (**Figure 55**) sont relativement proches et suggèrent une bonne conservation des séquences et donc des possibilités d'alignement importantes au sein des genres. Ils apparaissent néanmoins nettement différents chez les pins comme le montre par exemple la comparaison des profils obtenus chez *Pinus radiata*, *Pinus pinaster* et *Pinus nigra* ou *Pinus sylvestris* (**Figure 56**). Chez les chênes, on observe de manière similaire que les possibilités d'alignement des espaceurs des 2 espèces étudiées de la section *Quercus* (*Quercus robur* et *Quercus petraea*) et des 2 espèces étudiées de la section *Cerris* (*Quercus suber* et *Quercus libani*) sont excellentes. On peut de plus identifier de nombreuses positions nucléotidiques informatives entre ces deux groupes (environ 10 % des positions de l'alignement, **Figure 31**).

Sur la base de ces données, il apparaît donc que la pertinencee taxinomique des espaceurs de l'ADNr 5S chez les espèces forestières à cycle long que nous avons étudiées se situe quelque part entre l'échelle du genre (étude des variations entre espèces) et l'échelle de la famille (étude des variations entre les genres). Elle équivaut donc à celle des ITS de l'ADNr (voir **Tableau 1**). Gernandt et Liston (1999) ont en effet montré que les séquences des ITS sont clairement informatives pour établir les relations phylogénétiques entre les grands groupes géographiques de mélèzes (mélèzes américains et eurasiens) ainsi qu'entre les genres *Larix* et *Pseudotsuga* qui sont considérés comme très proches. De la même façon, Samuel *et al.* (1998) puis Manos *et al.* (1999) ont récemment démontré l'intérêt des ITS pour l'étude de la différenciation moléculaire des grands groupes taxinomiques (sections) au sein du genre *Quercus*. Etant donné que chez les chênes et les mélèzes la taille des espaceurs de l'ADNr 5S est comparable à celle des ITS, l'utilisation de cette région dans des études phylogénétiques similaires, particulièrement à l'échelle du genre, pourrait être envisagée. La confrontation d'arbres phylogénétiques indépendants issus de l'analyse de différentes régions du génome nucléaire (arbres de gènes) est en effet importante pour mieux comprendre la phylogénie des espèces (arbres d'espèces).

**Tableau 47**. Caractéristiques des clones d'espaceurs séquencés pour différentes espèces de chênes*.

| Espèce | Abréviation utilisée | Source | Origine géographique | Libellé du clone | Taille (pb) |
|---|---|---|---|---|---|
| *Q. myrsinifolia* | *Qmy* | Université Paris XI | Asie (Est) | S*Qmy* | 327 |
| *Q. rubra* | *Qru* | Université Paris XI | Amérique du Nord (Est) | S*Qru* | 216 |
| *Q. palustris* | *Qpa* | Université Paris XI | Amérique du Nord (Est) | S*Qpa* | 218 |
| *Q. pubescens* | *Qpu* | Université Paris XI | Europe | S*Qpu* | 230 |
| *Q. frainetto* | *Qfr* | Université Paris XI | Europe | S*Qfr* | 230 |

* Voir également le **Tableau 22**

A titre de test prospectif, nous avons cloné et séquencé un espaceur de l'ADNr 5S de 5 nouvelles espèces de chêne afin de disposer d'un échantillon couvrant les grandes subdivisions du genre *Quercus* au sens de Camus et Nixon (**Tableau 17**). Il s'agit, pour le sous-genre *Cyclobalanopsis* de *Quercus myrsinifolia* Blume (= *Cyclobalanopsis myrsinaefolia* Blume) et, pour le sous-genre *Quercus*, de *Quercus palustris* Muench. et *Quercus rubra* L. (section *Lobatae*), *Quercus pubescens* Willd. (section *Quercus* comme *Quercus robur* et *Quercus petraea*) et *Quercus frainetto* Ten. (section *Mesobalanus*). Si l'on rajoute *Quercus suber* L. et *Quercus libani* Oliv. (sous-genre *Quercus*, section *Cerris*) déjà étudiées, seule la section *Protobalanus* du sous-genre *Quercus* n'est pas représentée. Les informations relatives aux clones séquencés pour ces nouvelles espèces sont indiquées dans le **Tableau 47**. Les informations concernant *Quercus robur*, *Quercus petraea*, *Quercus suber* et *Quercus libani* ont été données précédemment dans le **Tableau 22**. L'alignement de l'ensemble des séquences avec la séquence consensus majoritaire des chênes pédonculé et sessile (Cons S*Qro*/S*Qpe*) est présenté dans la **Figure 70**. Les séquences de *Q. pubescens* est *Q. frainetto* ont des tailles identiques à celle de la séquence Cons S*Qro*/S*Qpe* (230 pb) dont elles ne diffèrent respectivement qu'à 2 et 3 positions nucléotidiques. Les espaceurs des 5 autres espèces (S*Qmy*, S*Qli*, S*Qsu*, S*Qpa*, S*Qru*) présentent des tailles assez différentes, comprises entre 216 pb (S*Qru*) et 327 pb (S*Qmy*). De manière remarquable, la plupart de ces séquences sont caractérisées par une large duplication en tandem de l'extrémité 5' (éléments répétés 1 et 2) qui prend naissance 5 pb en amont, dans le gène 5S. Elle est seulement absente du clone d'espaceur le plus court de *Quercus libani* (*Qli*S). Le fait que cette duplication soit partagée par les espaceurs de différentes espèces et que les deux éléments répétés diffèrent considérablement en taille (71-99 pb) et en séquence (35 positions variables sans compter les indels), indique que son origine est ancienne et remonte probablement à la période de diversification du genre *Quercus*. Les deux séquences de la section *Lobatae*, S*Qru* et S*Qpa*, comportent en outre une grande et même délétion centrale d'environ 85 pb située immédiatement en aval (positions 208-303 de l'alignement) de la duplication en tandem. Ces deux événements paraissent donc corrélés. Plusieurs indels de moindre amplitude (< 24 pb) sont également notées dans cette région centrale des espaceurs (positions 234-241, 261-269, 283-293) ainsi qu'au sein des éléments répétés de la duplication (positions 88-91), particulièrement au niveau de la région du poly T (positions 23-50 et 128-156). La grande duplication en tandem (S*Qmy*, S*Qli*L, S*Qsu*, S*Qpa*, S*Qru*), la délétion centrale de 85 pb

(S*Qru*, S*Qpa*) ainsi que l'indel en position 283-293 (S*Qmy*, S*QliS*, S*QliL*, S*Qsu*) constituent 3 événements phylogénétiquement informatifs. Par ailleurs, 93 positions sur un total de 353 (26 %) présentent des substitutions dont 48 se révèlent informatives. Au total, environ 14 % des positions de l'alignement sont informatives. Ce chiffre est comparable à ce qui a été obtenu par Manos *et al.* (1999) dans le cas des ITS de chêne (19 % de positions informatives).

Les distances brutes (= divergences nucléotidiques) calculées entre ces 9 séquences d'espaceurs de l'ADNr 5S sont consignées dans le **Tableau 48**. Elles varient de 0,009 à 0,192 (indels incluses, 353 caractères) ou de 0,007 à 0,209 (indels exclues, 134 caractères). La divergence nucléotidique moyenne de S*Qmy* est de 0,171 (indels incluses) ou 0,143 (indels exclues), tandis que chez les autres clones elle se situe dans la gamme 0,115-0,131 (indels incluses) ou 0,118-0,138 (indels exclues). Le clone S*Qmy* est donc la séquence la plus divergente de notre échantillon. Comme par ailleurs la position phylogénétique basale du sous-genre *cyclobalanopsis* est maintenant bien établie (références dans Manos *et al.* 1999), nous avons utilisé cette séquence pour enraciner les arbres de gènes déduits de l'alignement présenté dans la **Figure 70**.

**Tableau 48**. Distances brutes, indels inclues (sous la diagonale) ou exclues (au-dessus de la diagonale), calculées entre 9 séquences d'espaceurs de l'ADNr 5S de divers chênes*.

| Clone | N° | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Cons S*Qro*/S*Qpe* | 1 | - | 0,007 | 0,022 | 0,134 | 0,157 | 0,179 | 0,187 | 0,134 | 0,127 |
| S*Qpu* | 2 | 0,009 | - | 0,015 | 0,142 | 0,164 | 0,187 | 0,194 | 0,142 | 0,134 |
| S*Qfr* | 3 | 0,017 | 0,017 | - | 0,149 | 0,179 | 0,201 | 0,209 | 0,149 | 0,149 |
| S*Qmy* | 4 | 0,160 | 0,164 | 0,174 | - | 0,149 | 0,157 | 0,164 | 0,127 | 0,119 |
| S*QLiS* | 5 | 0,153 | 0,162 | 0,166 | 0,169 | - | 0,067 | 0,015 | 0,127 | 0,134 |
| S*QliL* | 6 | 0,163 | 0,172 | 0,176 | 0,181 | 0,054 | - | 0,067 | 0,119 | 0,127 |
| S*Qsu* | 7 | 0,156 | 0,165 | 0,170 | 0,192 | 0,053 | 0,023 | - | 0,127 | 0,134 |
| S*Qpa* | 8 | 0,134 | 0,148 | 0,148 | 0,168 | 0,134 | 0,133 | 0,144 | - | 0,022 |
| S*Qru* | 9 | 0,127 | 0,141 | 0,148 | 0,159 | 0,141 | 0,139 | 0,149 | 0,023 | - |

* Distances calculées en utilisant l'option SHOWDIST du logiciel PAUP (Swofford 1993). Les abréviations utilisées sont indiquées dans les **Tableaux 22** et **47**.

Ces arbres de gènes ont été construits selon la méthode de parcimonie maximum en utilisant l'algorithme heuristique général sans pondération des états de caractères fourni par le logiciel PAUP, version 3.1.1 (Swofford 1993). Ils présentent la même topologie générale selon que l'on écarte ou non de l'analyse les positions comportants des indels (seule la longueur des branches change). La **Figure 71** présente uniquement l'arbre consensus déduit de l'alignement en conservant les positions correspondant à des indels (les "gaps" sont alors simplement traités comme des données manquantes). On y observe 3 groupes de séquences bien distincts et supportés par des valeurs de bootstrap élevées (> 97 %). Le premier, en position basale, comporte les 3 séquences des sections *Quercus* et *Mesobalanus*, le second, les 2 séquences de la section *Lobatae* et enfin le troisième, les 3 séquences de la section *Cerris*.

```
                    1         10        20        30        40        50        60        70        80        90       100
Cons SQro/SQpe cccctCTCTTTTTCTTCGGATTTTTTTT-----------------------GTCGCCGGGCTTACCGTTTACGGCGGTCCAGCACGCTCGC----TTCGTAGGCT
SQpu           ..........................-----------------------.................G.......................----..........
SQfr           ..........................-----------------------.............T.........................----..........
SQmy 1         .............C...C.A.C.....TTTTTTTTGTTGTTGTTGTTGTTG.G.C.ATC......G.-------................CTTG..T...C..G
SQliS          ................C.........TTT---------------------C...TT.......T........T................----..T.....T.
SQliL 1        ................C.........TTT-----------------------..TT...G..GT.........................----.CT.C.T..A
SQsu 1         ................C.........--------------------------..TT...G..GT.........................----.CT.C.T..A
SQpa 1         .....T......G.......A.......----------------------------.TT.......T..C.......................----..T...T..A
SQru 1         ...-.T......G.......A.......----------------------------..T.......T..C...........G..........----..T...T..A

                        110       120       130       140       150       160       170       180       190       200
SQmy 2         .....T..C..C......C.C..C..----------------------------C..TT....ACCT.....G..T....GA...........----.CT.C.....
SQliL 2        .....T..A.....CCA.C.........TTTTT-------------------C...T........T........T...................----..T.......
SQsu 2         .....T..A.....CCT.C.........-----------------------C...TT.......T....C...C...................----..T.......
SQpa 2         .....T.G..........C.........T---------------------------..T......GT........T..T....A.........----..T...T..A
SQru 2         .....T............C.........----------------------------..T......GT........T..T....A.........----..T.A.T..A

                  210       220       230       240       250       260       270       280       290       300
Cons SQro/SQpe TAGTCTGCCCCAGGCCAAAAATACGAGCCGCTGATCTCGTTTTGGGTGCTAATTTGGTGGCGGCGATCTCGAAAT----------GCCCTCGGG-C
SQpu           ..........................................................................----------.........-.
SQfr           ..........................................................................----------.........-A
SQmy           C....G.....G..G.C.G..-....--------......C........CGGAGC-..C................CCGCGGGAGGC........G.
SQliS          ...C.C...............C.........A....................AGC-..C.......CG.......CGGCGGAAG-C........-T
SQliL          ...CGC...............C.........G....................AGC-..C.......CG.......CCGCGGAAG-C........-T
SQsu           ...C.C...............C.........A....................A---------....CG.......CCGCGGAAG-C........-T
SQpa           ------------------------------------------------------------------------------------------------
SQru           ------------------------------------------------------------------------------------------------

                     310       320       330       340       350       360       370
Cons SQro/SQpe TGCAGCTCCCTGGTGTTAGAGTGCATCGCTGCATTCGGATCCTTT-ATAAGGATTTAACGCGCTTATGCTGAT
SQpu           ..........................................C.-...........................
SQfr           ........................................T...-...........................
SQmy           ....C...T.........A.......G..................-.....A................G....
SQliS          ...GC...T.C..C....AG.....C..T...G............T.....A...A.................
SQliL          ....C...T.C..C....AG........T...G............T.....A...A.................
SQsu           ....C...T.C..C....AG.A......T...G............T.....A...A.................
SQpa           ....C...T........GA..........C...........A...-...........................
SQru           ....C...T........GA..........C...............-...........................
```

**Figure 70**. Alignement des séquences de clones d'espaceurs de l'ADNr 5S de *Q. pubescens* (S*Qpu*), *Q. frainetto* (S*Qfr*), *Q. myrsinifolia* (S*Qmy*), *Q. libani* (S*Qli*S et S*Qli*L), *Q. suber* (S*Qsu*), *Q. palustris* (S*Qpa*) et *Q. rubra* (S*Qru*) avec la séquence consensus majoritaire déduite de la comparaison des espaceurs de *Q. robur* et *Q. petraea* (Cons S*Qro*/S*Qpe,* voir la **Figure 29**). Les bases variables par rapport à la séquence Cons S*Qro*/S*Qpe* sont indiquées selon les conventions IUPAC standard. Les deux éléments (1 et 2) dupliqués en tandem dans S*Qmy* (99 et 75 pb), S*Qli*L (75 et 83 pb), S*Qsu* (71 et 78 pb), S*Qpa* (75 et 76 pb) et S*Qru* (74 et 74 pb) sont alignés afin de faciliter la comparaison de leurs séquences. Les duplications observées dans S*Qmy*, S*Qpa*, S*Qru*, S*Qli*L et S*Qsu* prennent leur origine 5 pb en amont de l'espaceur dans le gène 5S (indiqué en minuscules et en gras dans la séquence consensus). Noter la grande délétion centrale (85 pb) au sein des séquences S*Qpa* et S*Qru* immédiatement en aval des duplications.

Les variations des espaceurs de l'ADNr 5S sont donc globalement concordantes avec la définition des grands groupes taxinomiques généralement admise pour le genre *Quercus*. L'arbre obtenu ne semble pas en revanche confirmer les phylogénies moléculaires déduites de l'analyse des régions ITS de l'ADNr et du polymorphisme de l'ADN chloroplastique de Manos *et al.* (1999), qui placent la section *Cerris* en position basale au sein du sous-genre *Quercus*. Samuel *et al.* (1998) avaient également obtenu des résultats partiellement en conflit avec ceux de Manos *et al.* (1999) en analysant pourtant également ces mêmes régions ITS de l'ADNr.

Dans ce contexte, une étude phylogénétique plus complète basée sur l'analyse des espaceurs de l'ADNr 5S permettrait probablement d'amener de nouvelles informations pour discuter de ces questions de différenciation moléculaire des grands groupes taxinomiques au sein du genre *Quercus*.

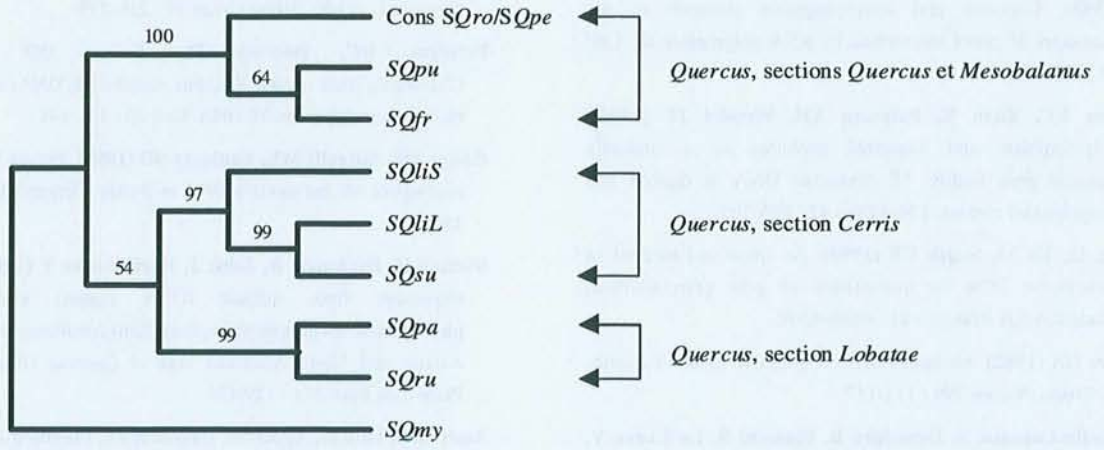



**Figure 71**. Arbre de gènes consensus résultant de l'analyse parsimonieuse (algorithme heuristique général) de 38 clones d'espaceurs de l'ADNr 5S de divers chênes du sous-genre *Quercus*, sections *Quercus* (*Qro*, *Qpe*, *Qpu*), *Mesobalanus* (*Qfr*), *Cerris* (*Qli*, *Qsu*) et *Lobatae* (*Qru*, *Qpa*). L'arbre est enraciné par une séquence issue d'un chêne du sous-genre *Cyclobalanopsis* (*Qmy*). L'alignement des séquences est donné dans la **Figure 70**. Les abréviations et caractéristiques de chaque clone sont consignées dans les **Tableaux 22** et **47**. Cons S*Qro*/S*Qpe* désigne la séquence consensus déduite de l'analyse de 30 clones d'espaceurs de *Quercus robur* et *Quercus petraea* (**Figure 29**). Les valeurs de bootstrap (pour 100 réplications) sont indiquées près de chaque noeud.

## Bibliographie

**Arcade A, Anselin F, Faivre Rampant P, Lesage MC, Pâques LE, Prat D (2000)**. Application of AFLP, RAPD and ISSR markers to genetic mapping of European and Japanese larch. Theor Appl Genet 100 : 299-307.

**Barciszewska MZ, Erdmann VA, Barciszewski J (1994)**. A new model for the tertiary structure of 5S ribonucleic acids in plants. Plant Mol Biol Rep 12 : 116-131.

**Barrénèche T, Bodénès C, Lexer C, Trontin JF, Fluch S, Streiff R, Plomion C, Roussel G, Steinkellner H, Burg K, Favre JM, Glössl J, Kremer A (1998)**. A genetic linkage

map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers. Theor Appl Genet 97 : 1090-1103

**Baum BR, Johnson DA (1998).**The 5S rRNA gene in sea barley (*Hordeum marinum* Hudson *sensu lato*): sequence variation among repeat units and relationship to the X haplome in barley (*Hordeum*). Genome 41 : 652-661

**Bogenhagen DF, Brown DD (1981).** Nucleotide sequences in *Xenopus* 5S DNA required for transcription termination. Cell 24 : 261-270

**Buckler ES, Ippolito A, Holtsford TP (1997).** The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. Genetics 145 : 821-832

**Chaw SM, Zharkikh A, Sung HM, Lau TC, Li WH (1997).** Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. Mol Biol Evol 14 : 56-68

**Ciliberto G, Rangei G, Constanzo F, Dente L, Cortese R (1983).** Common and interchangeable elements in the promoters of genes transcribed by RNA polymerase III. Cell 32 : 725-733

**Cronn RC, Zhao X, Paterson AH, Wendel JF (1996).** Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. J Mol Evol 42 : 685-705

**Deng G, Yu M, Smith HS (1993).** An improved method of competitive PCR for quantitation of gene copy number. Nucleic Acids Research 21 : 4848-4849.

**Dover GA (1982).** Molecular drive: a cohesive mode of species evolution. Nature 299 : 111-117

**Dumolin-Lapègue S, Demesure B, Fineschi S, Le Corre V, Petit RJ (1997).** Phylogeographic structure of white oaks throughout the European continent. Genetics 146 : 1475-1487

**Fulnecek J, Matyàsek R, Kovarik A, Bezdek M (1998).** Mapping of 5-methylcytosine residues in *Nicotiana tabacum* 5S rRNA genes by genomic sequencing. Mol Gen Genet 259 : 133-141

**Gernandt DS, Liston A (1999).** Internal transcribed spacer region evolution in *Larix* and *Pseudotsuga* (*Pinaceae*). Amer J Bot 86 : 711-723

**Goldsbrough PB, Ellis THN, Lomonossoff GP (1982).** Sequence variation and methylation of the flax 5S RNA genes. Nucleic Acids Res 10 : 4501-4514

**Govindaraju DR, Cullis CA (1992).** Ribosomal DNA variation among populations of a Pinus rigida Mill. (pitch pine) ecosystem: I. Distribution of copy numbers. Heredity 69 : 133-140

**Hariharan N, Reddy PS, Padayatty JD (1987).** 5S rRNA genes in rice embryos. Plant Mol Biol 9 : 443-451

**Kellogg EA, Appels R (1995).** Intraspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. Genetics 140 : 325-343

**Kolchinsky A, Kolesnikova M, Ananiev E (1991).** "Portraying" of plant genomes using polymerase chain reaction amplification of ribosomal 5S genes. Genome 34 : 1028-1031

**Lubaretz O, Fuchs J, Ahne R, Meister A, Schubert I (1996).** Karyotyping of three *Pinaceae* species *via* fluorescent *in situ* hybridization and computer-aided chromosome analysis. Theor Appl Genet 92 : 411-416

**Manos PS, Doyle JJ, Nixon KC (1999).** Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (*Fagaceae*). Molecular Phylogenetics and Evolution 12 : 333-349

**Nei M (1987).** Molecular Evolutionary Genetics. Columbia University Press, New York

**Ohri D, Ahuja MR (1990).** Giemsa C-banded karyotype in *Quercus* L. (Oak). Silvae Genet 39 : 216-219

**Peterson RC, Doering JL, Brown DD (1980).** Characterization of two *Xenopus* somatic 5S DNAs and one minor oocyte-specific 5S DNA. Cell 20 : 131-141

**Röder MS, Sorrells ME, Tanksley SD (1995).** Pulsed-field gel analysis of 5S and satellite DNA in Barley. Genome 38 : 153-157

**Samuel R, Bachmair A, Jobst J, Ehrendorfer F (1998).** ITS sequences from nuclear rDNA suggest unexpected phylogenetic relationships between Euro-Mediterranean, East Asiatic and North American taxa of *Quercus* (*Fagaceae*). Plant Syst Evol 211 : 129-139

**Sastri DC, Hilu K, Appels R, Lagudah ES, Playford J, Baum BR (1992).** An overview of evolution in plant 5S DNA. Plant Syst Evol 183 : 169-181

**Schlötterer C, Tautz D (1992).** Slippage synthesis of simple sequence DNA. Nucleic Acids Res 20 : 211-215

**Steele KP, Holsinger KE, Jansen RK, Taylor DW (1991).** Assessing the reliability of 5S rRNA sequence data for phylogenetic analysis in green plants. Mol Biol Evol 8 : 240-248

**Szymanski M, Specht T, Barciszewska MZ, Barciszewski J, and Erdmann VA (1998).** 5S rRNA data bank. Nucleic Acids Res 26 : 156-159

**Zoldos V, Papes D, Cerbah M, Panaud O, Besendorfer V, Siljak-Yakovlev (1999).** Molecular-cytogenetic studies of ribosomal genes and heterochromatin reveal conserved genome organization among 11 *Quercus* species. Theor Appl Genet 99 : 969-977

# *ANNEXE*

Theor Appl Genet (1998) 97: 1090–1103 © Springer-Verlag 1998

T. Barreneche · C. Bodenes · C. Lexer · J.-F. Trontin
S. Fluch · R. Streiff · C. Plomion · G. Roussel · H. Steinkellner
K. Burg · J.-M. Favre · J. Glössl · A. Kremer

# A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers

Received: 30 January 1998 / Accepted: 12 May 1998

**Abstract** A genetic map of Pedunculate oak (*Quercus robur*) was constructed based on one 5S rDNA, 271 RAPD, ten SCAR, 18 microsatellite, one minisatellite, and six isozyme markers. A total of 94 individuals from a full-sib family was genotyped. Two maps, including 307 markers, were constructed according to the "two-way pseudo-testcross" mapping strategy. Testcross markers segregating in the 1:1 ratio were first used to establish separate maternal (893.2 cM, 12 linkage groups) and paternal (921.7 cM, 12 linkage groups) maps. Both maps provided 85–90% genome coverage. Homologies between the male and female linkage groups were then identified based on 74 intercross markers segregating in the 3:1, 1:2:1 and 1:1:1:1 ratios (RAPDs, SCARs, SSRs, 5S rDNA and isozymes) in the hybrid progeny. In each map, approximately 18% of the studied markers showed segregation distortion. More than 60% of the skewed markers were due to an excess of heterozygote genotypes. This map will be used for: (1) studying the molecular organisation of genomic regions involved in inter- and intraspecific differentiation in oaks and (2) identification of QTLs for adaptive traits.

Communicated by M. A. Saghai Maroof

T. Barreneche · C. Bodenes · R. Streiff · C. Plomion · G. Roussel · A. Kremer (✉)
Laboratoire de Génétique et d'Amélioration des Arbres Forestiers, INRA, B.P. 45, Pierroton, F-33610 Cestas Cedex, France
Fax: (33) 5 57 97 90 88
E-mail: kremer@pierroton.inra.fr

C. Lexer · J. Glossl · H. Steinkellner
Zentrum für Angewandte Genetik, Universität für Bodenkultur Wien, Muthgasse 18, A-1190 Vienna, Austria

J.-F. Trontin · J.-M. Favre
Laboratoire de Biologie Forestière, associé INRA, Université Henri Poincaré, Nancy, B.P. 239, 54506 Vandoeuvre-lès-Nancy, France

S. Fluch · K. Burg
Forschungszentrum Seibersdorf, A-2444 Seibersdorf, Austria

**Key words** *Quercus robur* L · Linkage map · RAPD · SCAR · Microsatellite · Minisatellite · 5S rDNA · Isozymes

## Introduction

Depending on the authors the number of existing oak species varies between 300 and 600. They are spread over the northern hemisphere (Europe, Africa, North America and Asia) from boreal zones to semi-arid areas. Pedunculate oak (*Quercus robur* L.) is one of the most important and widespread oak species in Europe. It belongs to the white oak section (*Lepidobalanus*) and is distributed throughout Europe from central Spain to the Urals. According to Camus (1954) the *Lepidobalanus* section is the richest with respect to numbers and diversity and comprises more than 150 species spread over all four continents.

*Quercus robur* L. has a broad ecological range, from very acid to chalky and from moist to mesophilic soils, but it prefers to grow on neutral and moist soils (Becker and Levy 1990). Genetic studies in pedunculate oak have mostly concentrated on population genetics, provenance research and the inheritance of economically important traits (Kremer et al. 1993, for a review).

Cytogenetic investigations have been restricted to karyotypic analysis. *Quercus robur* L. is a diploid species ($2n = 2x = 24$), although in some cases supernumerary chromosomes (B type) have been reported (Ohri and Ahuja 1990; Besendorfer et al. 1996). Natural triploid individuals have been described (Johnsson 1946; Butorina 1993) that usually are larger than diploid trees. Compared to data available on other woody angiosperms (e.g. *Populus*, *Eucalyptus*, *Acacia*, *Pyrus*, *Fraxinus*) the physical genome size of *Q. robur* is larger (1.88 pg/2C) (Zoldoš et al. 1998), although its size is among the lower values found in plants (Favre

and Arnould 1996; Favre and Brown 1996). Segregation studies with gene markers have been conducted in *Quercus robur* with isozymes (Zanetto et al. 1996), SCARs (sequence characterised amplified regions, Bodénès et al. 1996) and microsatellites (Steinkellner et al. 1997a). However, due to the small number of loci analysed in each of these studies, no linkage groups could be constructed except with isozymes when two linkage groups were identified (Zanetto et al. 1996).

Our objective was to construct a genetic map with all earlier developed markers (isozymes, SCARs, microsatellites, minisatellites, 5S rDNA). In order to locate these markers in the genome, we added randomly distributed markers using amplification with arbitrarily designed primers [random amplified polymorphic DNA (RAPD) Williams et al. 1990]. The pedigree used for the map construction is a full-sib $F_1$ progeny of *Quercus robur* L. Linkage groups were assembled according to the "two-way pseudo-testcross" method described by Grattapaglia and Sederoff (1994).

To our knowledge, this is the first reported linkage map in the genus *Quercus*, and in the *Fagaceae* family as well. Some of the markers used in the construction of the map have been reported to be applicable to other oak species: isozymes in *Q. petraea* and *Q pubescens* (Müller-Starck et al. 1996), microsatellites in other sections of the genus *Quercus* (Steinkellner et al. 1997b). Because oaks of a given botanical section usually exhibit high genetic similarities (Kremer and Petit 1993), we expect that the present map will provide a useful tool for various applications in population and quantitative genetics in different white oak species.

## Materials and methods

### Plant material

Segregation data from a two-generation full-sib pedigree including 94 individuals were used to construct two genetic maps, one for each parental tree. The male parent originated from near Arcachon (southwest of France), and the female parent was located on the Forestry Research Station in Pierroton (southwest of France). The controlled cross was made in the spring of 1992 according to the techniques described in Zanetto et al. (1996). About 400 acorns were collected in pollination bags, of which 94 were used for map construction. Preliminary segregation analysis of the 94 offspring with a few isozymic loci did not detect any acorn resulting from fertilisation with foreign pollen (Zanetto et al. 1996).

### Marker analysis

Genomic DNA was extracted from leaves according to Saghai-Maroof et al. (1984) with some minor modifications described by Moreau et al. (1994).

Procedures used to study different markers are summarized in Table 1. Specific primer sequences and annealing temperatures for the SCAR, microsatellite, minisatellite and 5S rDNA markers are indicated in Table 2.

### Data analysis

After segregation was scored for each marker, Mendelian segregation was tested in the offspring using Chi-square goodness-of-fit tests. Linkage analysis of the markers was performed using a combination of MAPMAKER (Lander et al. 1987) and JOINMAP (Stam 1993) software (see Results section). The Kosambi (Kosambi 1944) mapping function was used to convert recombination frequencies to map distances in centiMorgans.

Genome length, E(G), of both parental maps was estimated under the assumption of random marker distribution according to method "3" of Chakravarti et al. (1991): $G = 2MX/K$, where $M = n(n-1)/2$ is the number of informative meioses (n is the number of loci analysed), X is the maximum observed map distance among the locus pairs above a threshold LOD, Z, and K is the number of locus pairs having LOD values at or above Z. The values tested for Z were 3, 4 and 5. The values of X and K were obtained using the "LODs" function of MAPMAKER with the maximum recombination fraction set to 0.5 and minimum LOD threshold set to Z. Values for X were obtained using the Kosambi map function. These estimations were calculated by taking into consideration only pairwise comparisons between framework markers.

## Results

### Segregation analysis

Markers were subdivided in four different groups on the basis of the segregation patterns in the progeny:

A) Loci that were in the testcross configuration (Grattapaglia and Sederoff 1994) between the parents (heterozygous in one parent and homozygous null in the other, or vice-versa). A 1 : 1 segregation ratio was observed in the $F_1$ family for most RAPDs, seven SCARs (six dominant and one codominant), two isozymes, 1 microsatellite (ssrQpZAG104) and one minisatellite (Omin3).
B) Loci that were heterozygous in both parents and that segregated in a 3 : 1 ratio in the offsprings. Fifty RAPDs followed this pattern.
C) Loci that were heterozygous in both parents and that segregated in a 1 : 2 : 1 ratio. This was the case for 2 microsatellites and three isozymes.
D) Loci that were heterozygous in both parents and that segregated in a 1 : 1 : 1 : 1 ratio. Three SCARs, 15 microsatellites, one isozyme and one 5S rDNA locus followed this pattern. Markers segregating in the 1 : 1 : 1 : 1 ratio were partitioned according to the alleles contributed by the relevant parent (Jacobs et al. 1995), resulting in a marker segregating 1 : 1 in both the female and the male parent, as for category A markers.

Approximately 80% of the markers analysed in the progeny fit the expected Mendelian segregation ratios (1 : 1, 3 : 1, 1 : 2 : 1 or 1 : 1 : 1 : 1). A total of 61 markers showed segregation distortion at $\alpha = 0.05$.

**Table 1** Protocols used for the molecular analysis of the different markers

| Marker type | Number of markers | PCR protocol | | | | Electrophoresis | Staining conditions | References |
|---|---|---|---|---|---|---|---|---|
| | | Primer type | Amplification conditions | | | | | |
| RAPD | 271 | Ten-base primers Operon Technologies kits A-Z | 1 step | 94°C | 4 min | 1.8% agarose gel | Ethidium bromide | Williams et al. 1990 |
| | | | 35 cycles | 92°C<br>40°C<br>72°C | 45 s<br>45 s<br>1 min 45 s | | | |
| SCAR | 10 | * | 1 step | 94°C | 3 min | 1% agarose gel or 8% non-denaturating polyacrylamide gel | Ethidium bromide Silver staining | Bodénès et al. 1997 |
| | | | 30 cycles | 94°C<br>T°C<br>72°C | 45 s<br>45 s<br>1 min 30 s | | | |
| Microsatellite | 18 | * | 1 step | 95°C | 6 min | 6% denaturating polyacrylamide gel | Fluorescence (ALF express) | Steinkellner et al. 1997b |
| | | | 30 cycles | T°C<br>72°C<br>92°C | 1 min<br>30 s<br>1 min | | Silver staining | Streiff et al. 1998 |
| | | | 1 step | 72°C | 8 min | | | |
| Minisatellite | 1 | * | 1 step | 94°C | 4 min | 2% agarose gel | Ethidium bromide | Fluch and Burg unpublished data |
| | | | 30 cycles | 94°C<br>50°C<br>72°C | 45 s<br>45 s<br>30 s | | | |
| 5S rDNA | 1 | * | 1 step | 96°C | 3 min | 8% non-denaturing polyacrylamide gel | Silver staining | Specht et al. 1997 |
| | | | 30 cycles | 94°C<br>55°C<br>72°C | 40 s<br>1 min<br>40 s | | | |
| Isozymes | 6 | – | | – | | 11%–12% starch gel or IEF | [b] | Zanetto et al. 1996 |

[a] See Table 2 for specific primer sequences and annealing temperatures
[b] See Zanetto et al. (1996) for staining conditions of various enzyme systems

**Table 2** Primer pairs and annealing temperatures of SCAR, microsatellite, minisatellite and 3Sr DNA markers

| Locus name | Primer sequence | Annealing temperature (°C) | References |
|---|---|---|---|
| A17-700 | F: GAC CGC TTG TTG GTA<br>B: GAC CGC TTG TGA TTA | 55 | Bodénès et al. 1997 |
| B11-1500 | F: AGA CCC GTA GAG GAG ACA TT<br>B: CCC GTG TAG TAT TCC ACA AA | 52 | Bodénès et al. 1997 |
| B12-500 | F: GTC TTC GAC TGG GGT GAA<br>B: TCA GTC CAT TCC GAA AGA | 52 | Bodénès et al. 1997 |
| B12-800 | F: CCT TGA ACG CAT TAT GAC AT<br>B: CCT TGA ACG CAG CAC AAT TC | 48 | Bodénès et al. 1997 |
| F14-750 | F: CAG AAG AAG CAA TGG TAA CA<br>B: CAA CAT TTG GTG TGT CTT AG | 50 | Bodénès et al. 1997 |
| I13-300 | F: GTG TGG TGC AGA AAA<br>B: ACT CCA GGT CTA TCC | 52 | Bodénès et al. 1997 |
| P14-450 | F: TGC AAA ACA TAA ACA TT<br>B: AAT CCA GTG GCA AGT TTT AA | 48 | Bodénès et al. 1997 |
| P17-1400 | F: CGC ATG GAT ATA CTA TTC<br>B: CTA CCT AGT GTC CTA TGG | 54 | Bodénès et al. 1997 |
| R12-500 | F: CGA GCG TTG ATA GCC AAT AG<br>B: TCG GAA GCA AAA GGG TAA TT | 51 | Bodénès et al. 1997 |
| U7-790 | F: GCC CTA ACA AAT CAT CTC<br>B: TAA TAT AGA AAG GGA AAG | 48 | Steinkellner et al. 1997a |
| ssrQpZAG1/5 | F: GCT TGA GAG TTG AGA TTT GT<br>B: GCA ACA CCC TTT AAC TAC CA | 55 | Steinkellner et al. 1997a |
| ssrQpZAG16 | F: CTT CAC TGG CTT TTC CTC CT<br>B: TGA AGC CCT TGT CAA CAT GC | 59 | Steinkellner et al. 1997a |
| ssrQpZAG3/64 | F: TAG AAA GCC CAA AAC CAA AAC C<br>B: CTT TTT GGA AGC CGC TTC CGT A | 50 | Steinkellner et al. 1997a |
| ssrQpZAG3/62 | F: CTT GAG CAT GGA ATC CTA TG<br>B: TCT AGA GGA GCT TTC CTT TAC AC | 58 | Steinkellner et al. 1997a |
| ssrQpZAG9 | F: GCA ATT ACA GGC TAG GCT GG<br>B: GTC TGG ACC TAG CCC TCA TG | 50 | Steinkellner et al. 1997a |
| ssrQpZAG15 | F: CGA TTT GAT AAT GAC ACT ATG G<br>B: CAT CGA CTC ATT GTT AAG CAC | 50 | Steinkellner et al. 1997a |
| ssrQpZAG36 | F: GAT CAA ATT TGG AAT ATT AAG AGA G<br>B: ACT GTG GTG GTA GTC TAA CAT GTA G | 50 | Steinkellner et al. 1997a |
| ssrQpZAG46 | F: CCC CTA TTG AAG TCC TAG CCG<br>B: TCT CCC ATG TAA GTA GCT CTG | 48 | Steinkellner et al. 1997a |
| ssrQpZAG58 | F: CTG CAA GAT TCG GAC AAG CAA<br>B: TCT TTT TTC CTA ATC TCA CCT G | 50 | Steinkellner et al. 1997a |
| ssrQpZAG102 | F: CAA CAA TGA AAG GGA GAA AGC<br>B: CGC TTT TGA CTT TCC TAC CTT | 50 | Steinkellner et al. 1997a |
| ssrQpZAG104 | F: ATA GGG AGT GAG GAC TGA ATG<br>B: GAT GGT ACA GTC GCA ACA TTC | 50 | Steinkellner et al. 1997a |
| ssrQpZAG108 | F: CTA GCC ACA ATT CAG GAA CAG<br>B: CCT CTT TTG TGA ATG ACC AAG | 50 | Steinkellner et al. 1997a |
| ssrQpZAG110 | F: GGA GGC TTC CTT CAA CCT ACT<br>B: GAT CTC TTG TGT GCT GTA TTT | 48 | Steinkellner et al. 1997a |
| ssrQpZAG119 | F: GAT CAG TGA TAG TGC CTC TC<br>B: GAT CAA CAA GCC CAA GGC AC | 50 | Steinkellner et al. 1997a |
| MSQ4 | F: TCT CCT CTC CCC ATA AAC AGG<br>B: GTT CCT CTA TCC AAT CAG TAG TGA G | 50 | Dow et al. 1995 |

**Table 2** Continued

| Locus name | Primer sequence | Annealing temperature (°C) | References |
|---|---|---|---|
| MSQ13 | F: TGG CTG CAC CTA TGG CTC TTA G<br>B: ACA CTC AGA CCC ACC ATT TTT CC | 50 | Dow et al 1995 |
| MSQ16 | F: GGA ACA ACT AGA GAG AAC CAA GTC AGG<br>B: TTG CCT ATC CTG CC CCG TAT CAC | 50 | Dow and Ashley 1996 |
| MicJ-AG22 | F: CGA TTT TGA ACA CCT TCT TC<br>B: CGA TGT GGG ATT TTG ATT TT | 46 | T. Kawahara (personal communication) |
| Omin3 | F: AGA ATC GAC TTC AAT GCA AG<br>B: CCA TTC ACA GAG GTT GTC GG | 50 | Fluch et al. (personal communication) |
| 5S rDNA | F: TGG GAA GTC CTC GTG TTG CA<br>B: ATT AGT GCT GGT ATG ATC GC | 55 | Specht et al. 1997 |

## Map construction

The map construction was done in two steps. First, two separate parental maps (paternal and maternal) were constructed using markers segregating in the 1 : 1 ratio (category A and D markers) with a confidence level of $P \leq 0.01$. These "testcross" markers were grouped with the help of MAPMAKER (Lander et al. 1987) using the following criteria:

- a minimum LOD score of 3.0 for statistical acceptance of linkage.
- a maximum recombination fraction $\theta$ of 0.40.

In the initial step of map construction, the order of non-distorted markers was approximated using the "FIRST ORDER" command. From this order, a subset of markers that could be locally ordered with a likelihood ratio greater than LOD = 2 was used to establish a framework map using the "RIPPLE" command. Markers that could not meet this ordering criteria were dropped from the framework and placed as accessory markers. Accessory markers were then located beside their closest framework marker. Distorted markers that were excluded from the framework map were also placed in the established framework map as accessory markers using the "NEAR" command. These markers are indicated with a boldface asterisk in Fig. 1.

In a second step, intercross markers segregating in both parents (category B, C) in addition to category D markers already mapped in the first step were used to identify homologous counterparts between parental maps. Markers of category D were mapped using MAPMAKER. However, because MAPMAKER does not accept mixed segregation data, intercross dominant and codominant markers of category B and C were placed to an approximate location on both maps with the JOINMAP programme (Stam 1993).

Finally, a total of 307 Loci (271 RAPDs, ten SCARs, 18 microsatellites, one minisatellite, six isozymes and one 5S rDNA marker) were located on both oak linkage maps (Fig. 1). Among the 50 intercross RAPD loci (category B) only the 24 markers showing strong linkage (LOD $\geq 4.5$) with framework markers in both maps were retained.

### *Female map* (*Fig.* 1)

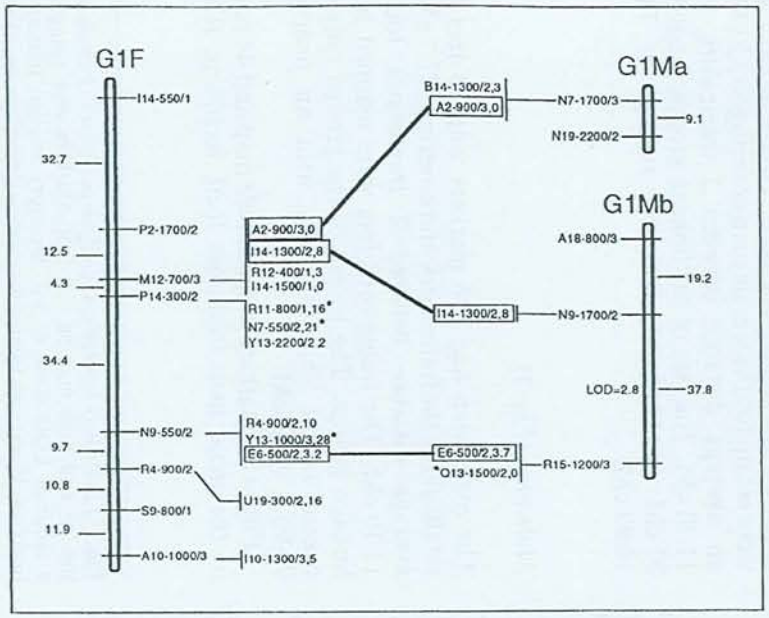
A total of 136 markers were used to establish the female map. Approximately 58% of them (79 loci) were located on the framework map. Twelve linkage groups were set up for the female map covering 893.2 cM, with an average distance between 2 framework loci of 11.30 cM. The size of the linkage groups ranged from 33 cM to 124.8 cM, with an average of 74.40 ± 16.90 cM.

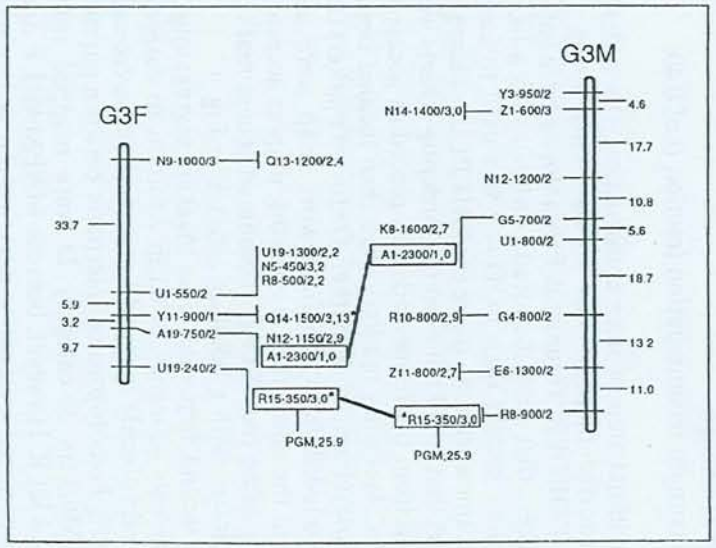
### *Male map* (*Fig.* 1)

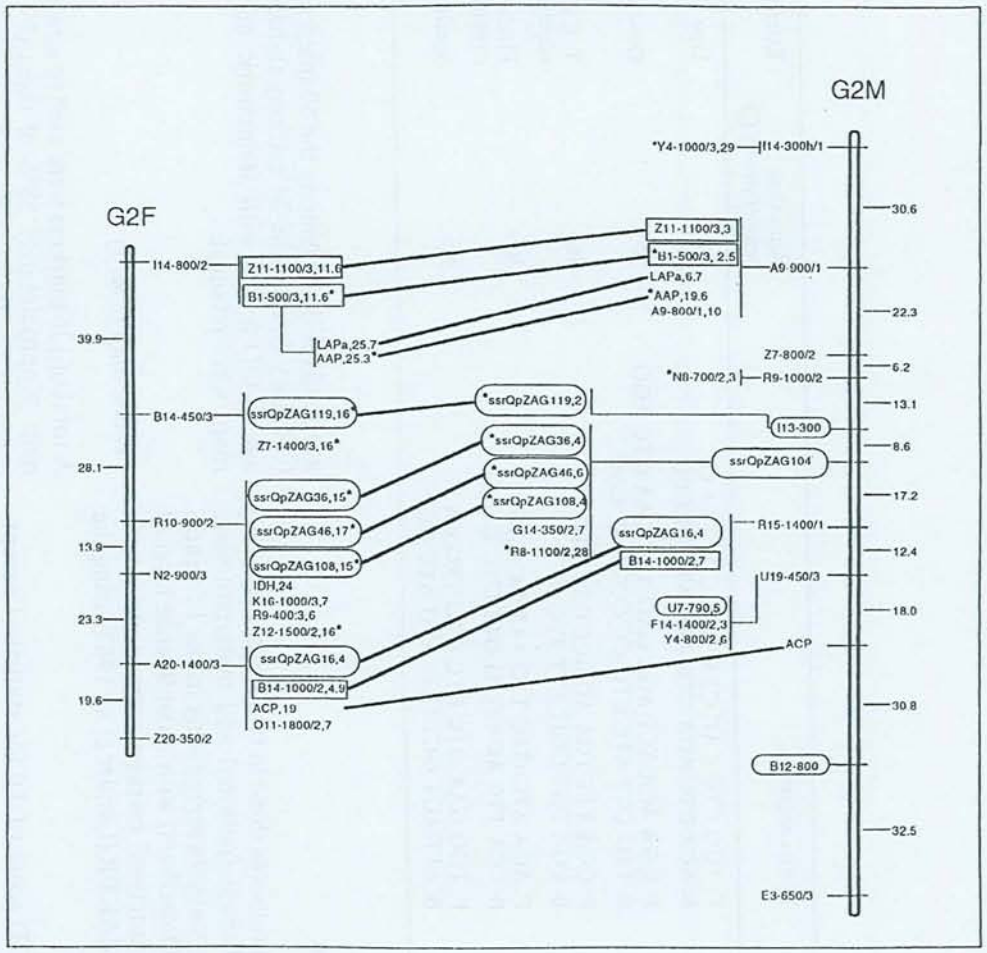
The male map had 136 markers with 83 loci (60%) establishing the framework and covering 921.7 cM. The average distance between 2 framework loci was 11.10 cM. The framework loci were mapped into 12 linkage groups. The length of the groups ranged between 9.1 cM and 191.8 cM, with an average of 70.90 ± 27.80 cM.

Fifty-two markers in the female map and 44 markers in the male map (testcross loci), including RAPDs,

**Fig. 1** Framework linkage maps of *Quercus robur* L. Female map on the *left* and male map on the *right*. Markers were mapped with a minimal LOD score of 3.0. Accessory markers (testcross and intercross types) were located beside their closest framework marker. Intercross markers (category B and C, see Results section) and some recoded testcross markers (category D) were used to identify homologous linkage groups. These markers are interconnected with *bold lines*. Intercross RAPD markers are *boxed*, and microsatellite, minisatellite, SCAR and 5S rDNA markers are *circled*. Markers displaying distorted segregation ($\alpha = 0.05$) are denoted by *boldface asterisk*


G1F
G1Ma
G1Mb
I14-550/1
32.7
P2-1700/2
12.5
M12-700/3
4.3
P14-300/2
34.4
N9-550/2
9.7
R4-900/2
10.8
S9-800/1
11.9
A10-1000/3
I10-1300/3,5
U19-300/2,16
R4-900/2,10
Y13-1000/3,28*
E6-500/2,3.2
A2-900/3,0
I14-1300/2,8
R12-400/1,3
I14-1500/1,0
R11-800/1,16*
N7-550/2,21*
Y13-2200/2.2
B14-1300/2,3
A2-900/3,0
N7-1700/3
N19-2200/2
9.1
I14-1300/2,8
N9-1700/2
A18-800/3
19.2
LOD=2.8
37.8
E6-500/2,3.7
*O13-1500/2,0
R15-1200/3


G3F
G3M
N9-1000/3
Q13-1200/2,4
33.7
U1-550/2
5.9
Y11-900/1
3.2
A19-750/2
9.7
U19-240/2
U19-1300/2,2
N5-450/3,2
R8-500/2,2
Q14-1500/3,13*
N12-1150/2,9
A1-2300/1,0
K8-1600/2,7
A1-2300/1,0
R15-350/3,0*
PGM,25.9
*R15-350/3,0
PGM,25.9
N14-1400/3,0
Y3-950/2
Z1-600/3
4.6
17.7
N12-1200/2
10.8
G5-700/2
5.6
U1-800/2
18.7
R10-800/2,9
G4-800/2
13.2
Z11-800/2,7
E6-1300/2
11.0
R8-900/2


G2F
G2M
I14-800/2
39.9
B14-450/3
28.1
R10-900/2
13.9
N2-900/3
23.3
A20-1400/3
19.6
Z20-350/2
Z11-1100/3,11.6
B1-500/3,11.6*
LAPa,25.7
AAP,25.3*
ssrQpZAG119,16*
Z7-1400/3,16*
ssrQpZAG36,15*
ssrQpZAG46,17*
ssrQpZAG108,15*
IDH,24
K16-1000/3,7
R9-400:3,6
Z12-1500/2,16*
ssrQpZAG16,4
B14-1000/2,4.9
ACP,19
O11-1800/2,7
*ssrQpZAG119,2
*ssrQpZAG36,4
*ssrQpZAG46,6
*ssrQpZAG108,4
G14-350/2,7
*R8-1100/2,28
Z11-1100/3,3
*B1-500/3, 2.5
LAPa,6.7
*AAP,19.6
A9-800/1,10
ssrQpZAG16,4
B14-1000/2,7
U7-790,5
F14-1400/2,3
Y4-800/2,6
*Y4-1000/3,29
I14-300h/1
30.6
A9-900/1
22.3
Z7-800/2
6.2
*N8-700/2,3
R9-1000/2
13.1
I13-300
8.6
ssrQpZAG104
17.2
R15-1400/1
12.4
U19-450/3
18.0
ACP
30.8
B12-800
32.5
E3-650/3

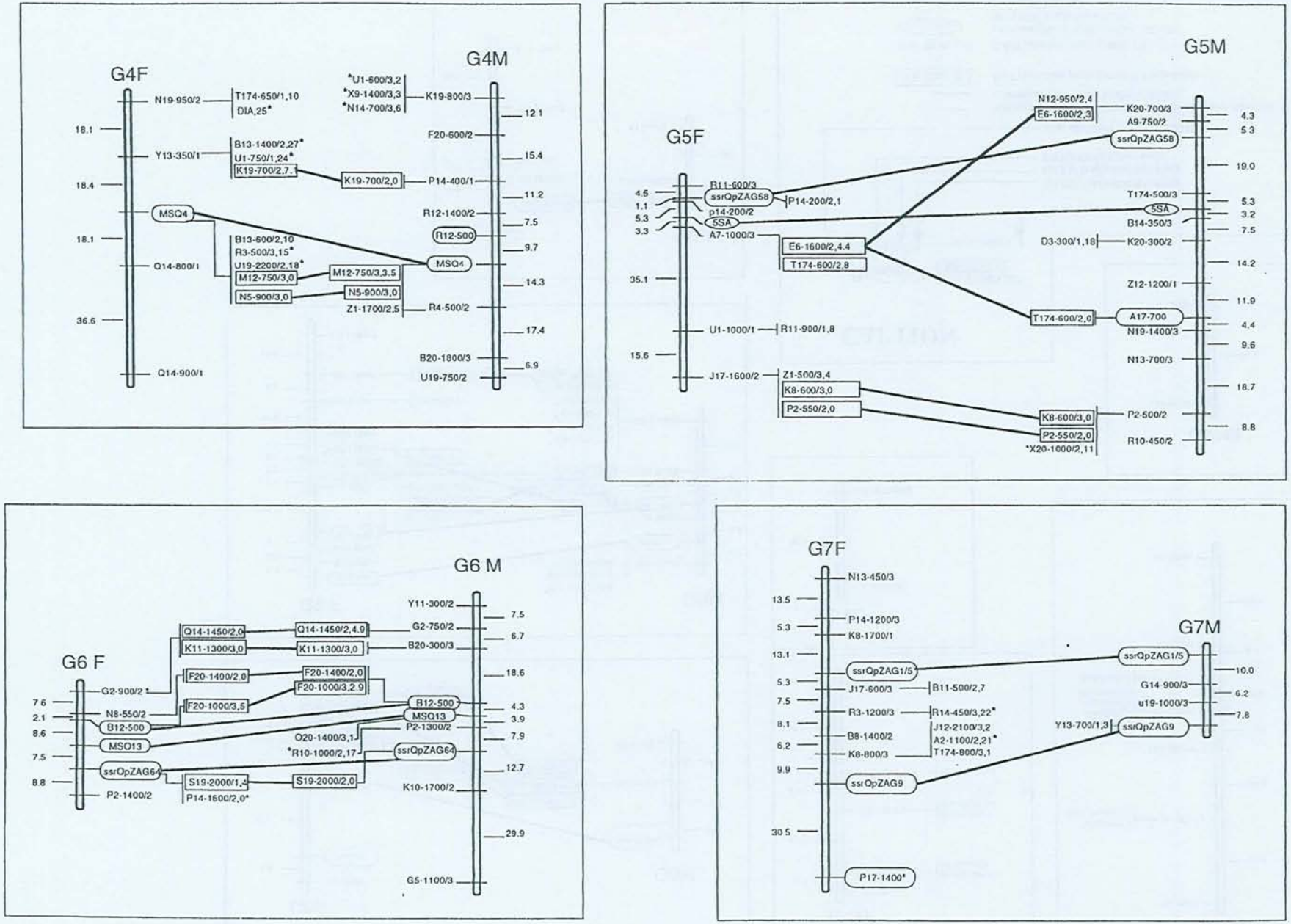



**Fig. 1** Continued. (See page 1094 for legend)

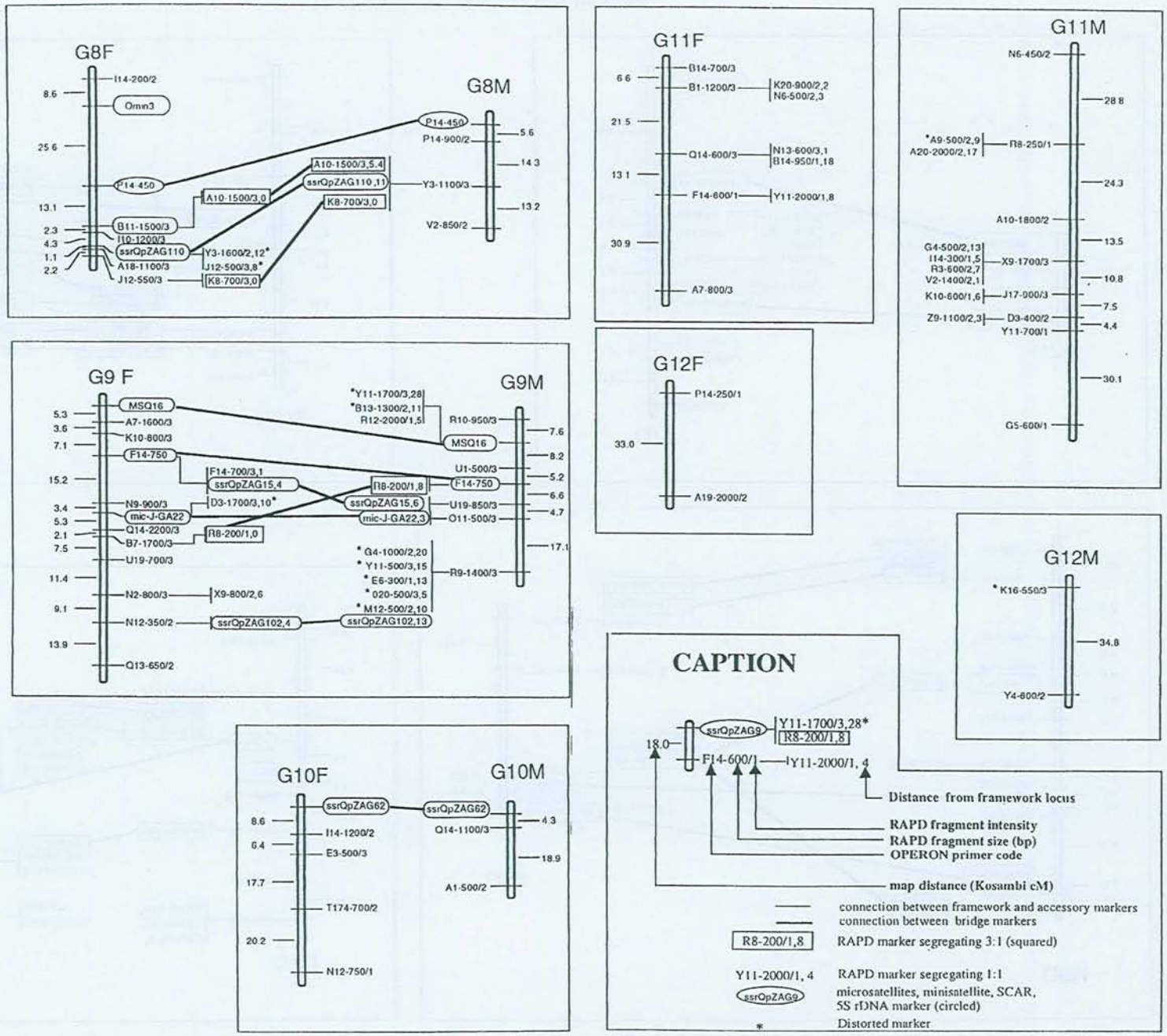



**Fig. 1** Continued. (See page 1094 for legend)

SSRs, isozymes and SCARs, were used as accessory markers.

The number of linkage groups corresponded to the number of haploid chromosomes of *Quercus*: n = 12, in both maps.

### *Homology between parental maps*

Markers segregating in both parents including intercross markers (50 RAPDs, three isozymes and 2 microsatellites) and loci segregating 1 : 1 : 1 : 1 (three SCARs, 15 microsatellites, one isozyme and one 5S rDNA marker) were used to identify homology between parental maps. These "bridge" markers are connected by bold lines in Fig. 1. Hence, ten homologous linkage groups were identified.

Because of low information content between marker-pairs corresponding to category A, B and A, C (Ritter et al. 1990), the use of codominant markers segregating 1 : 1 : 1 : 1 and therefore segregating 1 : 1 from both parents is a prerequisite for constructing an accurate combined map. Merging the male and female maps was not our primary goal but this would be possible if more microsatellites were available.

### Segregation distortion

At $\alpha = 0.05$, segregation distortions from the Mendelian expectation of testcross markers (category A and D markers) were detected at 25 marker loci in the female map and 24 marker loci in the male map. The distortion level was quite similar in both parents (18%). At $\alpha = 0.01$, the number of distorted markers decreased to 13 in the female map and 11 in the male map. Sixty percent of category A-distorted RAPD markers in the male map (66% in the female) displayed an excess heterozygote genotypes.

Fifteen intercross RAPD markers showed strong deviations from Mendelian expectations, but only 2 of these were associated with a LOD score $\geq 4.5$ to a framework marker in the map. Finally, a total of 61 markers (20%) showed segregation distortions at $\alpha = 0.05$. This distortion level is higher than expected by chance alone (15 distorted markers at $\alpha = 0.05$). Markers exhibiting segregation distortion were not randomly distributed in the genome. They tended to be preferentially grouped in a few clusters as shown on Fig. 1.

### Genome length and map coverage

The estimated total map length (average of three values) for the female parent amounted to E(G) = 1192 cM and for the male parent to E(G) = 1235 cM. These values are the first to be reported for oak. The different estimates, obtained from a range of Z values, were very close (data not shown) and were higher than the observed genome length of 893.2 cM for the female and 921.7 cM for the male map. This indicated that the mapped loci did not provide full coverage of the oak genome. Although the number of linkage groups corresponded to the number of haploid chromosomes (n = 12), 2 small linkage groups (LG12) were identified for both parental maps. In addition, linkage group 1 of the female map was homologous to 2 unlinked linkage groups (1a and 1b) of the male map. This gap could be filled in by additional markers. Following Lange and Boehnke (1982), the minimal number, N, of randomly distributed markers required to cover a proportion of $P = 95\%$ of a genome of size of L = 1200 cM at a maximum distance between markers of 2d = 20 cM is: $N = \ln(1 - p)/\ln(1 - 2d/L) = 178$. This expectation was higher than the number of markers actually used in the initial step of map construction (136 markers for both the male and female map). According to this formula, 136 markers would provide 90% coverage for a 20-cM map, a percentage coverage that was very close to that obtained following Bishop et al. (1983) (85% for both maps).

The assumption of a random distribution of markers in the genome seemed to fit our experimental data mainly based on RAPDs. The observed and expected distributions of markers were compared for 20-cM intervals. A chi-square test for departure from a Poisson distribution was computed for six classes or groups of classes containing at least five observations. The mean of the Poisson distribution was set to the mean number of markers per 20-cM interval length. No significant departure ($\alpha = 0.01$) from the Poisson expectation was observed in this goodness-of-fit test. A random distribution of RAPD markers was also observed in maritime pine (Plomion et al. 1995).

## Discussion

### Pseudo-test-cross mapping in *Q. robur*

The *Q. robur* maps presented here were constructed following the "two-way pseudo-testcross" strategy. Two maps of similar size (893 cM in the female parent, 922 cM in the male parent) were obtained for each parent, each one represented by 12 linkage groups corresponding to the 12 haploid chromosomes identified in oaks (Ohri and Ahuja 1990). The two maps are furthermore equally balanced in number of framework markers, mean size of linkage groups and mean distance between framework markers. As oaks exhibit high levels of heterozygozity in natural populations (Kremer and Petit 1993; Zanetto et al. 1994), 18% of the RAPD markers were found to be heterozygous in both parents (these are referred to as intercross loci segregating in the 3 : 1 ratio) compared to the 82% that

**Table 3** Map information on some woody angiosperms and annual plants

| Species (chromosome number) ((pg/C)) | Population type | Number of maps | Marker type | Number of markers | Observed map length | Map length coverage | Estimated map length[b] | References |
|---|---|---|---|---|---|---|---|---|
| *Citrus* (2n = 18) ((0.62 pg/C)) | Intergeneric $BC_1$ | 1 | RFLP isozymes | 69 | 553 cM | – | 900–1350 cM | Durham et al. 1994 |
| | Intergeneric $BC_1$ | 1 | RAPD, RFLP | 312 | 1192 cM | 70–80% | | Cal et al. 1994 |
| | Intergeneric $BC_1$ | 2 | RAPD, RFLP isozymes | 100 | 633.1 cM (female)<br>1503 cM (male) | –<br>– | | Luro et al. 1995 |
| | Intergeneric $F_1$ | 1 | RFLP, SSR isozymes | 56 | 410 cM | – | | Kijas et al. 1997 |
| *Eucalyptus* (2n = 22) ((0, 6 pg/C)) | Interspecific $F_1$ | 2 | RAPD | 491 | 1552 cM (female)<br>1101 cM (male) | 95.80%<br>95.20% | 1100–1650 cM | Grattapaglia and Sederoff 1994 |
| | Interspecific $F_1$ | 2 | RAPD | 505 | 1331 cM (female)<br>1415 cM (male) | 95–99%<br>95–99% | | Verhaegen and Plomion 1996 |
| | Intraspecific $G_2$ | 1 | RAPD, RFLP isozymes | 339 | 1462 cM | – | | Byrne et al. 1995 |
| *Persea* (2n = 24) ((0, 9 pg/C)) | Interspecific $F_1$ | 2 | RAPD, SSR minisatellites | 90 | 352.6 cM | – | 1200–1800 cM | Sharon et al. 1997 |
| *Malus* (2n = 34) ((0.8 pg/C)) | $F_1$ | 1 | RAPD, RFLP isozymes | 409 | 950 cM | – | 1700–2550 cM | Hemmat et al. 1994 |
| *Populus* (2n = 38) ((0.6 pg/C)) | Interspecific $F_1$ | 1 | RAPD, RFLP isozymes, STS | 343 | 1261 cM | 48.50% | 1900–2850 cM | Bradshaw et al. 1994 |
| *Prunus* (2n = 16) ((0.3 pg/C)) | Interspecific $F_2$ | 1 | RFLP, isozymes, morphological | 118 | 800 cM | – | 800–1200 cM | Foolad et al. 1995 |
| *Prunus persica* | Interspecific $F_2$ | 1 | RAPD, RFLP | 65 | 332 cM | < 50% | | Rajapakse et al. 1995 |
| | Intraspecific $F_2$ | 1 | RAPD, isozymes, morphological | 99 | 396 cM | – | | Chaparro et al. 1994 |
| | Intraspecific $F_2$ | 1 | RFLP, RAPD, AFLP, IMA, morphological | 270 | 712 cM | ~ 100% | | Dirlewanger et al. 1998 |
| *Prunus amygdalus* | Intraspecific $F_1$ | 2 | RFLP, isozymes | 127 | 393 cM (female)<br>394 cM (male) | – | | Viruel et al. 1995 |
| *Quercus robur* L. (2n = 24) ((0.9 pg/C)) | Intraspecific $F_1$ | 2 | RAPD, SCAR, SSR, isozymes 5S rDNA | 307 | 893.2 cM (female)<br>921.7 cM (male) | 85%<br>85% | 1200–1800 cM | Barreneche et al. 1998 (this study) |

**Table 3** Continued

| Species (chromosome number) ((pg/C)) | Population type | Number of maps | Marker type | Number of markers | Observed map length | Map length coverage | Estimated map length[b] | References |
|---|---|---|---|---|---|---|---|---|
| *Theobroma cacao* L. (2n = 20) ((0.4 pg/C)) | Intraspecific $F_1$ | 1 | RAPD, RFLP, isozymes | 193 | 759 cM | – | 1000–1500 cM | Lanaud et al. 1995 (this study) |
| | Intraspecific $BC_1$ | 1 | RAPD, RFLP | 138 | 1068 cM | – | | Crouzillat et al. 1996 |
| *Vitis* (2n = 38) ((0.5 pg/C)) | Interspecific $F_1$ | 2 | RAPD, RFLP, isozymes | 523 | 1196 cM (female) | – | 1900–2850 cM | Lodhi et al. 1995 |
| | | | | | 1477 cM (male) | – | | |
| *Arabidopsis thaliana* (2n = 10) ((0.15 pg/C)) | RIL | 1 | RAPD, microsatellites | 302 | 675 cM | ~ 100% | 500–750 cM | Reiter et al. 1992 |
| | | | | 30 | | | | Bell et al. 1994[c] |
| *Hordeum vulgare* (2n = 14) ((5.50 pg/C)) | DH | 1 | RAPD, RFLP, isozymes | 295 | 1250 cM | ~ 100% | 700–1050 cM | Kleinhofs et al. 1993 |
| *Lycopersicon esculentum* (2n = 24) ((0.98 pg/C)) | $F_2$ | 1 | RFLP, isozymes, morphological | 1030 | 1276 cM | ~ 100% | 1200–1800 cM | Tanksley et al. 1992 |
| *Zea mays* (2n = 20) ((2.5 pg/C)) | RIL, $F_2$ | 1 | RFLP, isozymes, molecular markers | 275 | 1765 cM | ~ 100% | 1000–1500 cM | Causse et al. 1996 |

[a] $F_1$, Two-generation outbred pedigree; $F_2$, three-generation inbred pedigree; $G_2$, three-generation outbred pedigree, RIL, recombinant inbred lines; DH, doubled-haploid lines; $BC_1$, first-generation backcross

[b] 100–150 cM per chromosome according to Rajapakse et al. (1995)

[c] See http://cbil.humgen.upenn.edu/Natgc/SSLP_info/SSLP_map.html

segregated in a testcross configuration. Reported corresponding values in other species using the"two-way pseudo-testcross" are quite different: the number of RAPDs segregating in 3:1 proportions varied between 2% and 5% in *Eucalyptus* studies (Grattapaglia and Sederoff 1994; Verhaegen and Plomion 1996) and amounted to 5% in pines (Kubisiak et al. 1995). These differences can likely be imputed to the differences in levels of heterozygozity between these species. The higher proportion of heterozygote RAPD loci in both parents also allowed the assignment of homologous linkage groups of the two parents as suggested by Hemmat et al. (1994). However, homologies between groups became more evident with codominant markers: microsatellites, minisatellite, SCARs, isozymes and the 5S rDNA marker. Among the 12 linkage groups identified in the two parents, 10 pairs could clearly be associated because of the presence of several common markers. A complete merging of the two maps into one single consensus map could be accomplished only if the ordering of markers were more precisely defined and if cosegregation analysis were conducted in larger families.

Within the intraspecific $F_1$ cross of *Q. robur*, a large number of markers (18% in both parental maps) exhibited segregation distortion regardless of the type of marker. In other examples of "pseudo-testcross" mapping, distorted markers are present in lower proportions (9% in *Eucalyptus urophylla*, 8% in *Eucalyptus grandis* Verhaegen and Plomion 1996, 15% in *Pinus palustris* and 14% in *Pinus elliottii* Kubisiak et al. 1995). As in pines, eucalypts and pedunculate oak, there is a general trend towards a clustering of distorted markers on the linkage groups (linkage groups 2 and 4 in the female map, and linkage groups 4 and 9 in the male map, Fig. 1).

The large number of distorted markers in *Q. robur* may be related to the high genetic load observed in this species. Interestingly, the variation of diversity among eucalypts, pines and oaks as illustrated by the number of loci being heterozygous in the two parents follows the same pattern as the variation in the number of distorted loci. Although quantitative data on inbreeding depression are not available in oaks, observations of seedlings originating from control selfing crosses reveal that they experience several detrimental characters during their early development (poor germination, chlorophyll deficiency, reduced growth) (Kleinschmit and Kleinschmit 1996; and personal observations.).

## Genetic map of *Q. robur* and other woody angiosperms and annual plants

To our knowledge this is the first reported genetic map of an oak species and of any member of the *Fagaceae* family. Other linkage studies based on a reduced number of isozyme loci and on several full-sib families have indicated that 4 loci (*Aap*, *Lap*, *Idh*, *Acp*) are located on the same linkage group in *Q. robur* (Zanetto et al. 1996) and *Q. petraea* (Müller-Starck et al. 1996). Interestingly these 4 loci are distributed within linkage group 2 of our genetic map (Fig. 1).

Genetic maps in other genera of broadleaved forest trees are currently only available in *Eucalyptus* (Grattapaglia and Sederoff 1994; Verhaegen and Plomion 1996; Byrne et al. 1995) and in *Populus* (Bradshaw et al. 1994). The comparison with woody angiosperms is therefore extended to fruit trees having related species in natural forests and to a few annual plants for documentary purposes (Table 3). With a few exceptions (*Eucalyptus* and *Quercus*) most existing maps of woody angiosperms cover less than 80% of the genome. Based on the data of Table 3, the physical genome size of woody angiosperms varies between 0.27 pg DNA/C (*Prunus persica*) and 0.9 pg DNA/C (*Quercus robur*), lying in the range of variation between *Arabidopsis* (0.15 pg DNA/C) and *Lycopersicon esculentum* (0.98 pg DNA/C) (Arumuganathan and Earle 1991). Most woody angiosperms have lower physical and genetic genome sizes than annual plants. *Quercus robur* has the largest physical genome among the angiosperm woody species so far investigated, but its size is not greater than the genome of tomato. Curiously, the number of chromosomes and the genetic length of the genome in tomato and oak are also of a similar magnitude. Given that 1 pg represents $0.915 \times 10^9$ bp (Bennet and Smith 1976), the physical distance per unit of genetic distance in *Q. robur* amounts to $0.77 \times 10^6$ bp/cM, which represents only 3.21 times that of *Arabidopsis thaliana* (Table 3).

## Microsatellites and consensus maps in oaks and *Fagaceae*

As illustrated in our example where two maps were constructed (Fig. 1), homologies between linkage groups became obvious as codominant markers segregating in 1:2:1 or 1:1:1:1 or dominant markers segregating in 3:1 could be mapped in both parents. Because the standard error of recombination fraction between markers segregating 1:1:1:1 (recoded 1:1) and markers segregating 1:1 is lower than for the others (Ritter et al. 1990), their position on a linkage group is more precise. Hence, they are preferential candidates as anchor markers and for comparisons across maps.

Microsatellite primers are transportable across oak species as shown by the experiment conducted by Steinkellner et al. (1997b). Among the 18 microsatellites mapped in *Q. robur*, which were originally developed from *Q. petraea*, polymerase chain reaction (PCR) amplification and polymorphism were tested for 14 loci in other oak species and other genera of the *Fagaceae*. All amplified products for the 14 loci were polymorphic in oak species belonging to the same botanical section (*Lepidobalanus*.) The number of successful loci was

reduced to 64% in oak species of section *Cerris* and to 47% in oak species of section *Erythrobalanus*. These results are reinforced by our own experience with microsatellite loci developed from *Quercus macrocarpa* (Dow et al. 1995) and *Quercus salicina* (Kawahara T, unpublished data) and that were applied for mapping purposes in the present contribution. However, successful amplification of oak microsatellites in related genera of the family *Fagaceae* was reduced to 24% in *Fagus* and to 47% in *Castanea*. Even though the proportion of amplified fragments and the level of polymorphism decreased with increasing evolutionary distance, microsatellites can be regarded as preferential anchor and consensus markers across different oak species in the *Quercus* genus.

In the future these maps will be saturated using additional markers like amplified fragment length polymorphism (AFLP, Vos et al. 1995) and further integrated into a single consensus map when more microsatellites can be localised. This consensus map will be of great value in identifying quantitative trait loci of interest and particularly of those involved in adaptive traits of oaks.

**Acknowledgements** The authors gratefully acknowledge Dr. T. Kawahara for sharing primers of the MicJ-AG22 locus, Anne Zanetto and Alexis Ducousso for their contribution in the controlled cross and Bernard Montoussé for field technical assistance. The research was supported by an EU research project (B104 CT96 0706) "Development, optimization and validation of molecular tools for assessments of biodiversity in forest trees".

## References

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 35:1439–1445

Becker M, Levy G (1990) Le point sur l'écologie comparée du chêne sessile et du chêne pédonculé. Rev For Fr 42:148–154

Bell CJ, Ecker JR (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Genomics 19:137–144

Bennet MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. Philos Trans R Soc London 216:179–199

Besendorfer V, Zoldoš V, Peškan T, Krnik-Rasol M, Littway T, Papeš D (1996) Identification of potential cytogenetical and biochemical markers in bioindication of common oak forest. Phyton (Horn, Austria), Special issue 36:139–146

Bishop DT, Cannings C, Skolnick M, Williamson JA (1983) The number of polymorphic DNA clones required to map the human genome. In: Wier BS (ed) Statistical analysis of DNA sequence data. Marcel Dekker, New York, pp 181–200

Bodénès C, Laigret F, Kremer A (1996) Inheritance and molecular variations of PCR-SSCP fragments in pedunculate oak (*Quercus robur* L.). Theor Appl Genet 93:348–354

Bodénès C, Joandet S, Laigret F, Kremer A (1997) Detection of genomic regions differentiating two closely related oak species *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. Heredity 78:433–444

Bradshaw HD Jr, Villar M, Watson BD, Otto KG, Stewart S, Stettler RF (1994) Molecular genetics of growth and development in *Populus*. III. A genetic linkage map of a hybrid poplar composed of RFLP, STS, and RAPD markers. Theor Appl Genet 89:167–179

Butorina AK (1993) Cytogenetic study of diploid and spontaneous triploid oaks, *Quercus robur*. L. Ann Sci For 50:144s–150s

Byrne M, Murrell JC, Allen B, Moran GF (1995) An integrated genetic linkage map for eucalypts using RFLP, RAPD and isozyme markers. Theor Appl Genet 91:869–876

Cai Q, Guy CL, Moore GA (1994) Extension of the linkage map in *Citrus* using random amplified polymorphid DNA (RAPD) markers and RFLP mapping of cold-acclimation-responsive loci. Theor Appl Genet 89:606–614

Camus A (1954) Les Chênes – Monographie du genre *Quercus*. Editions Le Chevalier, Paris

Causse M, Santoni S, Damerval C, Maurice A, Charcosset A, Deatrick J, de Vienne D (1996) A composite map of expressed sequences in maize. Genome 39:418–432

Chakravarti A, Lasher LA, Reffer JE (1991) A maximum likelihood method for estimating genome length using genetic linkage data. Genetics 128:175–182

Chaparro JX, Werner DJ, O'Malley D, Sederoff SS (1994) Targeted mapping and linkage analysis of morphological, isozyme, and RAPD markers in peach. Theor Appl Genet 87:805–815

Crouzillat D, Lerceteau E, Petiard V, Morera J, Rodriguez H, Walker D, Phillips W, Ronning C, Schnell R, Osei J, Fritz P (1996) *Theobroma cacao* L.: a genetic linkage map and quantitative trait loci analysis. Theor Appl Genet 93:205–214

Dirlewanger E, Proner V, Parvery C, Rothan C, Guye A, Monet R (1998) Genetic linkage map of peach (*Prunus persica* (L.) Batsch) using morphological, RFLP, isoenzyme, RAPD and AFLP markers. TAG (in press)

Dow BD, Ashley MV (1996) Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. Mol Ecol 5:615–627

Dow BD, Ashley MV, Howe HF (1995) Characterisation of highly variable $(GA/CT)_n$ microsatellites in the bur oak, *Quercus macrocarpa*. Theor Appl Genet 91:137–141

Durham RE, Liou PC, Gmitter FG, Moore GA (1994) Linkage of restriction length polymorphisms and isozymes in *Citrus*. Theor Appl Genet 84:39–48

Favre JM, Arnould MF (1996) Nuclear DNA content and interspecific ribosomal variation between *Q. robur* and *Q. petraea*. In: Kremer A, Mühs H (eds) Inter- and intra-specific variation in European oaks: evolutionary implications and practical consequences. European Commission, EUR 16717EN, Luxembourg, pp 41–51

Favre JM, Brown S (1996) A flow cytometric evaluation of the nuclear DNA content and GC percent in genomes of European oak species. Ann Sci For 53:915–917

Foolad MR, Aruselkar S, Becerra V, Bliss FA (1995) A genetic map of *Prunus* based on an interspecific cross between peach and almond. Theor Appl Genet 91:262–269

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *E. urophyulla* using a pseudo-testcross mapping strategy and RAPD markers. Genetics 137:1121–1137

Hemmat M, Weeden NF, Manganaris AG, Lawson DM (1994) Molecular marker linkage map for apple. J Hered 84:4–11

Jacobs JME, Van Eck HJ, Arens P, Verkerck-Bakker B, te Lintel Hehhert B, Bastiaanssen HJM, El-Kharbotly A, Pereira A, Jacobsen E, Stiekema WJ (1995) A genetic map of potato (*Solanum tuberosum*) integrating molecular markers, including transposons and classical markers. Theor Appl Genet 91:289–300

Johnsson H (1946) Chromosome numbers of twin plants of *Quercus robur* and *Fagus silvatica*. Hereditas 32:469–472

Kleinschmit JRG, Kleinschmit J (1996) Artificial hybridization between *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. In: Kremer A, Mühs H (eds) Inter- and intra-specific variation in European oaks: evolutionary implications and practical consequences. European Commission, EUR 16717EN, Luxembourg, pp 69–87

Kijas JMH, Thomas MR, Fowler JCS, Roose ML (1997) Integration of trinucleotide microsatellites into a linkage map of *Citrus*. Theor Appl Genet 94:701–706

Kleinhofs A, Kilian A, Shagai-Maroof MA, Biyashev RM, Hayes P, Chen FQ, Lapitan N, Fenwick A, Blake TK, Kanazin V, Ananiev E, Dahleen L, Kudrna D, Bollinger J, Knapp SJ, Liu B, Sorrells M, Heun M, Franckowiack JD, Hoffman D, Skadsen R, Steffenson BJ (1993) A molecular, isozyme and morphological map of barley (*Hordeum vulgare*) genome. Theor Appl Genet 86: 705–712

Kosambi DD (1944) The estimation of map distance from recombination values. Ann Eugen 12:172–175

Kremer A, Petit RJ (1993) Gene diversity in natural populations of oak species. Ann Sci For 50:186s–202s

Kremer A, Savill PS, Steiner KC (1993) Genetics of oaks. Ann Sci For 50:1s–472s

Kubisiak TL, Nelson CD, Nance WL, Stine M (1995) RAPD linkage mapping in a longleaf pine × slash pine $F_1$ family. Theor Appl Genet 90:1119–1127

Lanaud C, Risterucci AM, N'Goran AKJ, Clement D, Flament MH, Laurent V, Falque M (1995) A genetic linkage map of *Theobroma cacao* L. Theor Appl Genet 91:987–993

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Lange K, Boehnke M (1982) How many polymorphic genes will it take to span the human genome? Am J Hum Genet 34:842–845

Lodhi MA, Daly MJ, Ye G-N, Weeden NF, Reisch BI (1995) A molecular marker based linkage map of *Vitis*. Genome 38:786–794

Luro F, Lorieux M, Laigret F, Bové JM, Ollitrault P (1995) Cartographie du génome des agrumes á l'aide de marqueurs moléculaires et distortions de ségrégation. In: Bervillé A, Tersac M (eds) Techniques et utilisation des marqueurs moléculaires. INRA Editions, Paris, pp 69–82

Moreau F, Kleinschmit J, Kremer A (1994) Molecular differentiation between *Q. petraea* and *Q. robur* assessed by random amplified DNA fragments. For Genet 1:51–64.

Müller-Starck G, Zanetto A, Kremer A, Herzog S (1996) Inheritance of isoenzymes in sessile oak (*Quercus petraea* (Matt.) Liebl.) and offspring from interspecific crosses. For Genet 3:1–12

Ohri D, Ahuja MR (1990) Giemsa C – banded karyotype in *Quercus* L. (Oak). Silvae Genet 39:216–219

Plomion C, O'Malley DM, Durel C-E (1995) Genomic analysis in maritime pine (*Pinus pinaster*). Comparison of two RAPDs maps using selfed and open-pollinated seeds of the same individual. Theor Appl Genet 90:1028–1034

Rajapakse S, Belthoff LE, He G, Estager AE, Scorza R, Verde I, Ballard RE, Baird WV, Callahan A, Monet R, Abbott AG (1995) Genetic linkage mapping in peach using morphological, RFLP and RAPD markers. Theor Appl Genet 90:503–510

Reiter RS, Williams JGK, Feldmann KA, Rafalski JA, Tingey SV, Scolnik PA (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

Ritter E, Gebhardt C, Salamini F (1990) Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. Genetics 125:645–654

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. Proc Natl Acad Sci USA. 81:8014–8018

Sharon D, Cregan PB, Mhammed S, Kusharska M, Hillel J, Lahav E, Lavi U (1997) An integrated genetic linkage map of avocado. Theor Appl Genet 95:911–922

Specht T, Szymanski M, Barciszewska MZ, Barciszewski J, Erdmann VA (1997) Compilation of 5S rRNA and 5S rRNA gene sequences. Nucleic Acids Research 25:96–97

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. Plant J 3: 739–744

Steinkellner H, Fluch S, Turetschek E, Lexer C, Streiff R, Kremer A, Burg K, Glössl J (1997a) Identification and characterization of $(GA/CT)_n$ – microsatellite loci from *Quercus petraea*. Plant Mol Biol 33:1093–1096

Steinkellner H, Lexer C, Turetschek E, Glössl J (1997b) Conservation of $(GA)_n$ microsatellite loci between *Quercus* species. Mol Ecol 6:1189–1194

Streiff R, Labbé T, Bacilieri R, Steinkellner H, Glössl J, Kremer A (1998) Within population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. Mol Ecol 7:317–328

Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Miller JC, Miller L, Paterson AH, Pineda O, Röder MS, Wing RA, Wu W, Young ND (1992) High density molecular linkage maps of the tomato and potato genomes. Genetics 32:1141–1160

Verhaegen D, Plomion C (1996) Genetic mapping in *Eucalyptus urophylla* and *Eucalyptus grandis* using RAPD markers. Genome 39:1051–1061

Viruel MA, Messeguer R, de Vicente MC, Garcia-Mas J, Puigdoménech P, Vargas F, Arus P (1995) A linkage map with RFLP and isozymes markers for almond. Theor Appl Genet 91: 964–971

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Williams JGK, Kubelik AR, Livak KJ, Rafalski A, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535

Zanetto A, Roussel G, Kremer A (1994) Geographic variation of inter-specific differentiation between *Quercus robur* L. and *Quercus petraea* (Matt.) For Genet 1:111–123

Zanetto A, Kremer A, Müller-Starck G, Hattemer HH (1996) Inheritance of isozymes in Pedunculate oak (*Quercus robur* L.) J Hered 87:364–370

Zoldoš V, Papeš D, Brown S, Panaud O, Šiljak-Yakovlev S (1998) Genome size and base composition of seven *Quercus* species: inter- and intra-population variation. Genome (in press)

université HENRI POINCARÉ
NANCY 1

Monsieur **TRONTIN Jean-François**

**DOCTORAT de l'UNIVERSITE HENRI POINCARE, NANCY-I**

en BIOLOGIE FORESTIERE

VU, APPROUVÉ ET PERMIS D'IMPRIMER

Nancy, le 29 DEC. 2000 n° 483

Le Président de l'Université

UNIVERSITÉ HENRI POINCARÉ, NANCY I
LE PRÉSIDENT

Cl. BURLET

Université Henri Poincaré, Nancy-I
24-30 rue Lionnois - B.P. 3069 - 54013 NANCY CEDEX
Tél. : 03 83 85 48 00 - Fax : 03 83 85 48 48

**Characterization and variation of the nuclear 5S ribosomal DNA multigene family from four forest tree species belonging to genera *Larix* M. (*Pinaceae*) and *Quercus* L. (*Fagaceae*).**

In higher plants, the main part of the nuclear genome (50-90%) exhibited a complex pattern of organization involving multiple families of tandem repeats or scattered repeated elements, with either well-identified (ribosomal DNA, histone genes, *etc.*) or unknown function (satellites, mobile elements, *etc.*). It is generally accepted that non-repeated sequences evolved under natural selection and/or genetic drift. In the case of many families of genes and noncoding sequences, however, fixation of mutations within a population may proceed more rapidly as a consequence of specific molecular mechanisms of turnover within the genome (concerted evolution).

In this work, the structural organization and evolution of nuclear 5S ribosomal DNA (5S rDNA), a universal and dominant multigenic family coding for a small component of eukaryotic ribosomes (5S ribosomal RNA), has been investigated in four long-lived forest tree species belonging to genera *Quercus* L. (the sympatric pedunculate and sessile oaks) and *Larix* M. (the allopatric European and Japanese larches). As in most higher plants, the 5S rDNA is organized into clusters of tandem repeats in these species. Each tandemly repeated unit consists of a highly conserved (between species) 120 bp coding region (5S gene) and a more variable intergenic spacer. A unique class of units (~ 350 bp) composed of 1000-2000 copies per genome has been detected in pedunculate and sessile oaks. In contrast, two highly divergent unit size classes, approximately 650 bp (short class) and 870 bp (long class) in length, were found in European and Japanese larches. The short and long units appeared to contribute about equally to the global copy number (3000-6000 copies per genome), as expected from their main alternating organization within array. A similar composite structure of tandem arrays was evidenced in four others American or European *Larix* species, suggesting that the elementary functional repeat which is homogenized and subjected to natural selection in larch consists of an unusual 1520 bp dimeric association of the short and long units. For most PCR-amplified 5S rDNA markers, our studies of intraspecific (pedunculate oak) and interspecific (European x Japanese larch) crosses revealed a conventional pattern of mendelian inheritance apparently governed by one major locus. In pedunculate oak, the 5S rDNA locus has been mapped on linkage groupe G5 located on the second metacentric pair of chromosomes.

The length evolution of 5S rDNA units involved tandem duplication, insertions and deletion mainly located in the central part of the intergenic spacer region. Small insertions and deletions induced by replication slippage are indeed responsible for spacer size variations in pedunculate and sessile oaks (197-233 bp) as well as in the short (514-546 bp) and long (738-771 bp) spacer classes in European and Japanese larches. Large tandem duplications and deletions (≥ 50 bp), probably resulting from unequal exchanges, have also been identified in the long larch spacer class and in spacer clones from several *Quercus* species. Comparatively, reduced length variations have been registered in the 5S gene region for pedunculate and sessile oaks (119-120 bp) and for European and Japanese larches (120-122 bp). In contrast to length variations, substitutions were found to affect about equally the 5S gene and spacer regions, thus contributing to the emergence of numerous pseudogenes within genome.

Except in inheritance studies, allelic variation between chromosomes (heterozygosity or *trans* variation) could not be distinguished from the substantial allelic variation within chromosome (*cis* variation) in order to formulate genotypes from phenotypes. Therefore, the total allelic variation (*cis* + *trans*) in oaks and larches was more readily deduced from sequence analyses. Our calculations of intraspecific nucleotide diversity for 5S gene (0.005-0.066), pedunculate and sessile oak spacers (0.047-0.057), European and Japanese larch short (0.098-0.067) or long spacers (0.015-0.037) were found to be similar to estimates published for many herbaceous species. In these short-lived species, the intraspecific spacer polymorphism is generally highly correlated with interspecific divergence. We were therefore surprised to discover that no qualitative differences could be identified in the spacer region between either pedunculate and sessile oak or European and Japanese larch. As already demonstrated for the 5S gene region in many plants, levels of intraspecific polymorphism and interspecific divergence may be decoupled in the long larch spacers. Unlike the short larch spacers or the oak spacers, the long larch spacers indeed contain a putative gene element conserved in the the *Pinaceae* family and possibly submitted to natural selective pressure. However, a more valid explanation could be that the number of generation since divergence of pedunculate and sessile oak or European and Japanese larch was not sufficient to promote accumulation of interspecific differences. In such an hypothesis, 5S rDNA array heterogeneity could be mainly attributed to the rapid chromosomes turnover in short-lived species and to the low frequency of concerted homogenizing forces in long-lived species. In this way, the rate of interrepeat homogenization was estimated to be lower or equal to the rate of speciation in *Quercus* and *Larix*. The taxonomic information provided by the 5S rDNA spacer region in long-lived forest tree species appeared to be useful for phylogenetic studies at the genus (variation between species) or family (variation between genera) levels. This was confirmed by a preliminary phylogenetic study of 9 *Quercus* species based on 5S rDNA spacer variation.

**Keywords**: 5S rDNA, multigene family, tandem repeats, *Quercus petraea*, *Quercus robur*, *Larix decidua*, *Larix kaempferi*, molecular diversity, inheritance, phylogeny, concerted evolution

**Caractérisation et variation d'une famille multigénique, l'ADN ribosomique 5S nucléaire, chez quatre espèces forestières des genres *Larix* M. (*Pinaceae*) et *Quercus* L. (*Fagaceae*).**

**Résumé** - La plus grande partie du génome nucléaire des plantes supérieures (50 à 90 %) est composée de multiples familles de séquences répétées en tandem ou dispersées, fonctionnelles (ADN ribosomique, gènes des histones, *etc.*) ou de fonction inconnue (satellites, éléments mobiles, *etc.*) qui sont agencées selon une architecture particulièrement complexe. Par rapport aux séquences uniques, seulement soumises à la sélection naturelle et à la dérive génétique, les familles de séquences répétées évoluent plus rapidement car elles subissent également une homogénéisation intragénomique (évolution concertée).

Ce travail concerne l'une de ces familles, l'ADN ribosomique 5S nucléaire (ADNr 5S) codant pour les petits ARN ribosomiques 5S, pour laquelle nous avons étudié la structure, l'organisation et les variations chez deux arbres forestiers sympatriques du genre *Quercus* L. (les chênes pédonculé et sessile) et chez deux arbres forestiers allopatriques du genre *Larix* M. (les mélèzes d'Europe et du Japon). Comme chez la plupart des plantes supérieures, cette famille multigénique est organisée en longues séries d'unités répétées en tandem qui sont chacune constituée d'un gène 5S très conservé entre les espèces (120 pb) et d'un espaceur intergénique de taille variable. Une seule classe d'unités (~ 350 pb) représentée par 1000-2000 copies par génome a été détectée chez les chênes pédonculé et sessile. En revanche, deux classes d'unités très divergentes, les unes courtes (~ 650 pb) et les autres longues (~ 870 pb), comportant chacune 1500-3000 copies par génome, ont été observées chez les mélèzes d'Europe et du Japon. Ces unités courtes et longues alternent majoritairement au sein de séries composites dont l'établissement est sans doute très ancien car le même type d'organisation a été retrouvé chez 4 autres espèces de mélèzes. Le maintien de cette structure composite originale peut s'expliquer si l'on admet que c'est le motif dimérique de 1520 pb formé d'une unité courte et d'une unité longue, et non chacune d'elles, qui constitue l'unité fonctionnelle soumise à la sélection naturelle. L'analyse d'un croisement de chênes pédonculés et d'un croisement de mélèzes d'Europe et du Japon a montré que les marqueurs fournis par l'ADNr 5S présentent généralement un mode d'hérédité mendélien gouverné par un unique locus. Dans le cas du chêne pédonculé, le locus a pu être cartographié sur le groupe de liaison G5 correspondant à la paire de chromosomes N°2.

Les unités d'ADNr 5S évoluent en longueur par duplication en tandem, insertions et délétions surtout localisées dans la zone centrale des espaceurs. De petites insertions et délétions liées à des phénomènes de glissements de brins sont ainsi responsables des variations de taille des espaceurs observées chez les chênes pédonculé et sessile (197-233 pb) et dans les espaceurs courts (514-546 pb) ou longs (738-771 pb) des mélèzes d'Europe et du Japon. De même, des duplications en tandem et délétions de grande taille (≥ 50 pb) probablement provoquées par des crossing-over inégaux ont été détectées dans les espaceurs longs de mélèzes ainsi que dans les espaceurs de plusieurs espèces de chênes. Les variations de longueur enregistrées dans les gènes 5S des chênes pédonculé et sessile (119-120 pb) et des mélèzes d'Europe et du Japon (120-122 pb) sont comparativement plus faibles. En revanche, les substitutions, affectent apparemment autant les gènes 5S que les espaceurs et contribuent à l'émergence de nombreux pseudogènes au sein de ces génomes.

En raison de l'existence de la variation allélique intrachromosomique (*cis*), il n'a pas été possible de décrire la variation allélique interchromosomique (*trans*) en dehors de l'étude de croisements. La quantification de la variation allélique totale (*cis* + *trans*) a donc été abordée chez les 4 espèces étudiées à partir d'analyses de séquences. Les estimations de diversité nucléotidique intraspécifique des gènes 5S (0,005-0,066), des espaceurs de chênes pédonculé et sessile (0,047-0,057), des espaceurs courts (0,098-0,067) ou longs (0,015-0,037) des mélèzes d'Europe et du Japon sont comparables à celles obtenues pour les espèces herbacées à cycle court. Chez ces dernières, le niveau de divergence interspécifique des espaceurs est bien corrélé au polymorphisme intraspécifique. Il est donc surprenant de constater qu'aucune différence qualitative n'a pu être enregistrée dans cette région entre chêne pédonculé et sessile ou entre mélèze d'Europe et du Japon. Comme pour les gènes 5S, il est possible que le polymorphisme intraspécifique et la divergence interspécifique des espaceurs longs de mélèzes soient découplés car ils contiennent un élément conservé commun avec d'autres *Pinaceae* qui est peut-être soumis à la pression de la sélection naturelle. Une telle explication n'est cependant pas valable dans le cas des espaceurs courts de mélèzes et des espaceurs de chênes. Il semble plus probable que le nombre de générations écoulées depuis la divergence des chênes pédonculé et sessile ou des mélèzes d'Europe et du Japon soit insuffisant pour que des différences interspécifiques aient pu apparaître. Sous cette hypothèse, l'hétérogénéité des séries d'ADNr 5S reflète surtout l'efficacité du brassage chromosomique chez les espèces à cycle court tandis qu'elle traduit plutôt la faible fréquence des événements d'homogénéisation intragénomique chez les espèces à cycle long. L'évolution concertée se réalise ainsi apparemment à une vitesse plus faible que les événements de spéciation chez les chênes et les mélèzes. Le niveau d'information taxinomique fourni par les espaceurs de l'ADNr 5S chez ces espèces forestières à long cycle a été approximativement situé entre l'échelle du genre (variations entre espèces) et l'échelle de la famille (variations entre les genres). Un test prospectif appliqué au genre *Quercus* (9 espèces) confirme l'intérêt de l'ADNr 5S pour les études phylogénétiques.

**Mots clés** : ADNr 5S, familles multigéniques, répétitions en tandem, *Quercus petraea*, *Quercus robur*, *Larix decidua*, *Larix kaempferi*, diversité moléculaire, héritabilité, phylogénie, évolution concertée