

Analyse du génome et gestion des ressources génétiques forestières

D. Prat, P. Faivre Rampant et E. Prado



Analyse du génome et gestion des ressources génétiques forestières

D. Prat, P. Faivre Rampant
et E. Prado

Table des matières

L'évaluation des ressources génétiques : l'analyse du polymorphisme	1
Conservation des ressources génétiques forestières	1
Classification des arbres forestiers	4
Organisation des génomes des arbres forestiers	6
Le génome nucléaire	8
■ Caryotype et ploïdie	10
■ Taille du génome	12
■ Organisation et expression d'un gène	12
■ Familles de gènes	15
■ ADN répété	20
Le génome chloroplastique	25
Le génome mitochondrial	32
L'évolution des génomes	35
Variabilité génétique des arbres forestiers	43
Ampleur de la variabilité	43
Évolution de la variabilité au cours des générations	44
Flux de gènes	46
L'intérêt des marqueurs génétiques	47
Informations attendues des marqueurs	47
Paramètres décrivant la variabilité génétique	49
■ Structure intrapopulation	50
<i>Paramètres de diversité</i>	50
<i>Déséquilibre de liaison</i>	52
<i>Régime de reproduction</i>	53
<i>Structure spatiale</i>	55
<i>Taille de population</i>	55
■ Structure interpopulation	56
<i>Différenciation entre sous-populations</i>	56
<i>Flux géniques</i>	60
<i>Distances génétiques</i>	61
<i>Phylogéographie</i>	66
Relations entre marqueurs génétiques et caractères quantitatifs	67
■ Carte génétique	67

■ Localisation des gènes majeurs	69
Les outils d'analyse	73
Les marqueurs génétiques.....	73
Principes et généralités	73
Marqueurs morphologiques	74
Marqueurs biochimiques.....	75
■ Produits du métabolisme secondaire.....	75
■ Isoenzymes	76
■ Protéines totales	77
Marqueurs moléculaires	81
■ Mise en évidence du polymorphisme	81
<i>Restriction par des endonucléases</i>	82
<i>Hybridation</i>	84
<i>Amplification génique</i>	85
<i>Structure de l'ADN</i>	87
<i>Séquence d'ADN</i>	90
■ Description des marqueurs moléculaires.....	92
<i>RFLP</i>	92
<i>Marqueurs dérivés de la PCR</i>	95
<i>AFLPTM et techniques dérivées</i>	113
<i>Marqueurs semi-spécifiques</i>	116
<i>Autres marqueurs</i>	117
■ Marqueurs moléculaires de séquences exprimées	118
■ Polymorphisme de nucléotide unique (SNP).....	121
Reproductibilité	124
■ Effet du prélèvement ou du tissu analysé.....	124
■ Difficultés techniques.....	125
■ Transfert entre laboratoires	128
Hérédité.....	129
■ Marqueurs du génome nucléaire	129
<i>Déterminisme génétique</i>	129
<i>Dominance</i>	133
■ Marqueurs des génomes cytoplasmiques.....	135
Automatisation et analyses	139
■ Extraction et électrophorèse	140
■ Lecture de gels : analyse d'images.....	144
■ Puces à ADN	146
<i>Principe</i>	146
<i>Applications</i>	148
■ Programmes d'analyse de la diversité	151
<i>Diversité et structure génétique</i>	152

<i>Taille de population</i>	157
<i>Régime de reproduction</i>	157
<i>Simulation</i>	158
<i>Phylogénie</i>	158
<i>Cartographie génétique</i>	160
<i>Détection et localisation de QTL</i>	162
Cytologie.....	164
Cytométrie.....	164
■ Cytophotométrie.....	165
■ Cytométrie en flux	166
Cytogénétique	168
■ Hybridation <i>in situ</i> (HIS).....	169
■ Hybridation de gènes (<i>Fluorescent in situ hybridization</i> , FISH).....	170
<i>Structure chromosomique</i>	170
<i>Gènes ribosomiques</i>	171
<i>Gènes de séquence unique</i>	171
<i>Amorçage in situ</i>	172
■ Hybridation de génome (<i>Genome in situ hybridization</i> , GISH)	172
Applications et choix des marqueurs	175
Échantillonnage du génome	176
Transfert des marqueurs entre espèces	177
Adéquation marqueur - objectif.....	181
Critères de choix.....	182
Utilisations et choix des marqueurs.....	192
■ Identification génétique.....	197
■ Phylogénie	201
■ Analyse et structuration de la variabilité génétique	206
<i>Structure et différenciation</i>	206
<i>Flux de gènes</i>	213
<i>Phylogéographie</i>	217
■ Cartographie génétique et ses applications	217
L'apport des marqueurs dans la gestion des ressources génétiques	223
Systématique et phylogénie.....	223
Histoire des populations	224
Introgression	225
Influence de grandes variations environnementales anciennes.....	225
Gestion des ressources génétiques	229
Évaluation des ressources génétiques	230
■ Variabilité et adaptation	230
■ Flux de gènes	235

Gestion de réseaux de populations <i>in situ</i>	238
■ Influence de la sylviculture	238
■ Constitution du réseau	240
<i>Choix des populations et contraintes climatiques</i>	240
<i>Principes de gestion</i>	244
<i>Illustrations</i>	246
■ Espèces disséminées	248
■ Espèces menacées	249
Ressources génétiques <i>ex situ</i>	251
■ Échantillonnage	252
■ Choix des sites	253
■ Gestion	255
 Conclusions	 259
 Annexes	 269
 Références bibliographiques	 341
 Glossaire	 433
 Index	 453

Avant-propos

Cet ouvrage a été préparé à la demande de la Commission des ressources génétiques pour lui apporter des informations aussi complètes que possible sur l'utilisation des marqueurs moléculaires dans la gestion des ressources génétiques forestières. Les arbres forestiers sont des espèces encore assez peu influencées par l'homme. Leur caractère naturel en fait un matériel de choix pour la mise en place de réseaux de protection des ressources génétiques. L'intérêt de ce document débord largement les espèces forestières bien qu'elles servent aussi souvent que possible d'illustration.

Les arbres forestiers appartiennent à des groupes taxonomiques différents : les Angiospermes et les Gymnospermes. Ces taxons ont divergé depuis longtemps et possèdent des particularités de leur génome qui sont présentées en préambule aux outils d'analyse du génome.

Les différents types de marqueurs génétiques en usage et d'utilisation prévisible à ce jour sont décrits de façon à présenter leurs avantages et inconvénients ainsi que leurs applications. Les techniques de marquage moléculaire évoluent sans cesse, et aujourd'hui elles tendent à cibler au mieux le gène. Nous avons également inclus les analyses à l'échelle du génome complet qui apportent aussi des informations utiles. Le lecteur pourra ainsi choisir les marqueurs les mieux appropriés à son objectif grâce aux analyses critiques des différents types de marqueurs. Les choix proposés ne sont que des recommandations. Les auteurs se sont attachés à faire le bilan des connaissances et des évolutions prévisibles au moment de la rédaction.

Nous remercions la Direction de l'espace rural et de la forêt qui nous a fourni les moyens d'initier cet ouvrage (convention 61.21.12/97). Nous tenons également à rendre hommage à toutes les personnes qui ont contribué à la préparation de cet ouvrage, en particulier Michel Arbez, Catherine Bodènès, Nathalie Frascaria-Lacoste, Alain Ghesquière, Antoine Kremer, François Lefèvre, Marie-Claude Lesage, Roseline Lumaret, Jean-Claude Mounolou, Rémy Petit, Bernard Romant-Amat, Sylvain Santoni et Giuseppe Vendramin, pour leur soutien à ce projet, pour les informations ou éléments qu'ils ont fournis et pour les améliorations qu'ils ont suggérées. Nous sommes également très reconnaissants à Gerald Tuskan et Stephen DiFazio qui nous ont fait part des données sur le génome de *Populus trichocarpa*, dont le séquençage est achevé.

L'évaluation des ressources génétiques : l'analyse du polymorphisme

Conservation des ressources génétiques forestières

La protection du patrimoine génétique des espèces végétales est reconnue comme une priorité, soutenue par les instances politiques nationales et internationales, dans un souci de conservation de la richesse génétique, mais aussi de protection de la nature et de l'environnement. Les ressources génétiques ont suscité plusieurs réunions ministérielles internationales. La première réunion gouvernementale internationale importante à l'échelle européenne a été la conférence ministérielle européenne pour la protection des forêts qui s'est tenue à Strasbourg en décembre 1990 et qui a réuni les délégations de 31 pays européens. La deuxième des 6 résolutions porte sur les ressources génétiques forestières : *L'ensemble des États d'Europe qui veulent s'associer à une défense commune de la forêt et de son environnement et souhaitent mobiliser les moyens nécessaires, s'engagent à mettre en œuvre dans leurs pays, selon les modalités qui leur paraissent les plus appropriées, une politique de conservation des ressources génétiques forestières.*

La conférence des Nations-Unies sur l'environnement et le développement et la forêt de Rio de Janeiro en juin 1992 a débouché sur une convention portant sur la protection de la biodiversité. La conservation *in situ* de la diversité biologique y est préconisée (Barthod, 1993). Les États, en fonction de leurs moyens, sont chargés d'élaborer des stratégies, plans ou programmes pour assurer la conservation et l'utilisation durable de la diversité biologique. Ces orientations sont confirmées et précisées dans la résolution H2 de la conférence interministérielle d'Helsinki de juin 1993 (Barthod et Touzet, 1994) et lors des réunions suivantes.

Les autorités politiques sont donc sensibles à la protection des ressources génétiques forestières. En effet, les espèces forestières sont des espèces sauvages qui recèlent encore l'essentiel de leur variabilité naturelle. Il est alors temps de protéger cette ressource que les activités humaines menacent de plus en plus.

Sous la pression démographique l'homme a réduit la surface boisée au profit de l'agriculture et de l'urbanisation. L'activité humaine exerce ainsi une pression sur

le milieu environnant qui s'est considérablement accrue au cours des dernières décennies ; cette pression menace l'intégrité de la biodiversité et les ressources génétiques forestières en particulier. L'homme est intervenu aussi sur le milieu forestier par la sylviculture conduite au profit de la production de bois en favorisant souvent un nombre réduit d'espèces au détriment des autres. L'homme a également occasionnellement déplacé des populations et modifié les structures génétiques des massifs forestiers.

D'autres facteurs contribuent à la dégradation de la diversité génétique des espèces forestières : les facteurs environnementaux et notamment les grands changements climatiques, telles les périodes glaciaires qui ont contraint les espèces à se restreindre en populations de faible effectif dans des zones refuges. C'est à partir de ces zones que les espèces survivantes ont recolonisé leur aire de répartition actuelle. Le réchauffement attendu de la planète laisse entrevoir d'autres risques et menaces d'extinction pour diverses espèces vivantes dont des arbres forestiers. Les conditions d'adversité exceptionnelles (pullulation d'insectes ravageurs, prolifération de pathogènes, pollution atmosphérique, etc.), favorisées ou non par l'activité humaine, font peser des menaces sur la survie des populations et des espèces forestières.

L'objectif de la conservation des ressources génétiques forestières est de préserver le potentiel d'adaptation des espèces ou des populations sur le long terme. Les populations doivent donc pouvoir évoluer pour répondre et s'adapter aux changements futurs de leur environnement. Ce ne sont pas tant les peuplements actuels que les générations futures qu'ils engendreront qui sont visés par les mesures de conservation des ressources génétiques. Nos choix doivent être guidés par le succès des générations d'arbres à venir. L'objectif majeur de la conservation des ressources génétiques forestières est de préserver des gènes qui permettront aux populations de faire face aux aléas futurs et inconnus de leur environnement et ainsi de contribuer au maintien des populations et espèces forestières (Arbez, 1994 ; Namkoong, 1995). Les caractères adaptatifs de ces espèces sont donc particulièrement importants. En effet, la longévité de ces espèces expose les populations et les individus aux diverses fluctuations et variations de l'environnement. Le potentiel adaptatif élevé des espèces forestières réside vraisemblablement dans leur forte variabilité génétique (Behm *et al.*, 1997). Cette variabilité doit être préservée afin de maintenir ce potentiel (Gregorius, 1991).

Dans une première étape, les espèces concernées par la mise en place de réseaux de conservation des ressources génétiques forestières sont des espèces de valeur économique ou qui nécessitent des mesures d'urgence. À l'échelle française, les quatre premières sont le sapin pectiné (*Abies alba*), le hêtre (*Fagus sylvatica*), l'orme (*Ulmus* : menace de la graphiose) et le merisier (*Prunus avium* : risque d'introduction de provenances étrangères appartenant parfois à une autre espèce). Les réseaux européens *EUFORGEN* se sont focalisés au début sur le peuplier noir (*Populus nigra* : risque d'introgression par les formes cultivées), des feuillus

précieux, le chêne liège (*Quercus suber*) et l'épicéa commun (*Picea abies* : menace en Europe centrale) (Turok et Frison, 1995). Les actions de conservation des ressources génétiques forestières se sont ensuite organisées autour de cinq réseaux européens : les feuillus sociaux (*Fagus*, *Quercus*), les chênes méditerranéens, les feuillus précieux, les peupliers européens et les conifères. Les réseaux préexistants se sont élargis à d'autres espèces. Ces réseaux sont sous la tutelle d'organismes internationaux (IPGRI, International Plant Genetic Resources Institute). D'autres réseaux existent en Asie-Pacifique (*APFORGEN*) et en Afrique au sud du Sahara (*SAFORGEN*, qui se préoccupe aussi d'arbres fruitiers). Pour la plupart de ces espèces, la conservation des ressources génétiques est envisagée *in situ*, sauf en cas de menace évidente des populations naturelles afin de maintenir sur une longue période un nombre élevé de génotypes et aussi de soumettre ces génotypes aux pressions évolutives naturelles.

La conservation des ressources génétiques a besoin d'une gestion adaptée, elle a aussi un coût économique. Une gestion rationnelle se base sur la connaissance de la diversité génétique de l'espèce, de sa distribution spatiale et dans la mesure du possible, de l'influence des forces évolutives pouvant modifier cette structure dans le temps.

La conservation *in situ*, qui consiste en la protection des populations dans leur milieu naturel, est largement préconisée pour les espèces forestières en raison de la longue durée de vie des individus, du maintien des pressions évolutives naturelles et d'une plus grande facilité de gestion.

La conservation *ex situ*, qui consiste au maintien des ressources génétiques en dehors du milieu naturel, prend des formes plus variées (collections d'individus, banques de graines, banques de pollen, banques de tissus, etc.). Elle est conduite lorsque la conservation *in situ* est trop difficile, comme chez l'orme (menace très forte *in situ* en raison d'un problème sanitaire). Le renouvellement des générations est alors la principale difficulté. Les règles de gestion de ces collections sont également difficiles à élaborer pour assurer leur entretien et leur pérennité.

Quels sont la taille et le nombre de populations à préserver pour assurer la meilleure représentation de la variabilité génétique existante et garantir au mieux la pérennité des populations ? Comment choisir les populations à préserver ? Comment assurer la régénération de ces populations pour le très long terme ? Telles sont les principales questions posées par la gestion des ressources génétiques. Pour espérer répondre même partiellement à ces questions, il est nécessaire d'évaluer au mieux la variabilité des espèces considérées en s'intéressant aussi à la répartition spatiale. Les pressions évolutives effectives et les menaces qui pèsent sur les populations doivent également être identifiées et estimées.

Avant de décrire les outils permettant l'évaluation de la variabilité génétique et la compréhension des mécanismes qui la régissent, nous présentons ici un état des connaissances sur les génomes des arbres forestiers de façon à mettre en évidence certaines de leurs particularités.

Classification des arbres forestiers

Les espèces forestières appartiennent à deux groupes taxonomiques différents, les Gymnospermes et les Angiospermes. Les Gymnospermes sont d'origine plus ancienne : elles se sont largement développées au Trias et sont probablement apparues au Carbonifère, soit il y a près de 300 millions d'années (Meyer-Berthaud *et al.*, 1999), alors que les Angiospermes ne se sont répandues qu'au Crétacé (Crane *et al.*, 1995), les plus anciens fossiles trouvés datant du Jurassique, soit il y a environ 150 millions d'années (Sun *et al.*, 1998). Les diverses lignées des Angiospermes se seraient différenciées dès l'émergence de ce groupe ; leur berceau se situerait en région tropicale (Barale et Lemoigne, 2000). Les Monocotylédones ont ensuite divergé et se sont diversifiées au début du Crétacé (Bremer, 2000). Les Angiospermes et les Gymnospermes se différencient par de nombreux caractères botaniques et génétiques.

La classification du monde vivant est établie à partir de divers caractères morphologiques qui sont partagés par les espèces voisines. Une proximité génétique est ainsi attendue entre espèces appartenant à des genres ou familles voisins. Cette information sert donc aussi à prédire le transfert éventuel, d'une espèce vers une autre, d'une information génétique telle que l'organisation du génome, la possibilité d'analyser le polymorphisme avec un ensemble de marqueurs génétiques, etc.

La plupart des Gymnospermes sont des plantes ligneuses à port arborescent et produisent donc du bois. Les forêts de Gymnospermes se rencontrent surtout sous un climat tempéré et boréal. Dans les régions tropicales, les Gymnospermes, sont présentes essentiellement en altitude où elles sont fréquemment associées aux Angiospermes. Les espèces forestières gymnospermes, souvent groupées sous le terme de résineux ou de conifères, sont réparties dans quelques familles : les Araucariacées, Céphalotaxacées, Cupressacées, Pinacées, Podocarpacees, Taxacées et Taxodiacees (souvent considérées comme faisant partie des Cupressacées). La famille des Pinacées, particulièrement importante, regroupe la majeure partie des conifères utilisés pour la production de bois. Au sein de cette famille, les données recueillies sur une espèce modèle sont plus facilement transférables aux autres espèces.

Les espèces forestières angiospermes, souvent appelées feuillus par les forestiers, constituent la presque totalité des forêts tropicales et sont aussi très abondantes sous un climat tempéré. Les espèces forestières angiospermes, qui retiennent l'attention de l'IUFRO (International Union of Forest Research Organizations), appartiennent à 117 genres qui se répartissent dans 52 familles. La plupart de ces familles appartiennent aux ordres des Rosales, Célastrales, Fabales, Fagales, Malpighiales, Malvales, Myrtales et Sapindales qui sont classés au sein des Rosidées (fig. 1). Les Rosidées constituent l'une des principales divisions des Eudicotylédones et comportent la majorité des espèces forestières feuillues. Dans ce groupe des Rosidées figurent les Brassicales qui comportent entre autres les

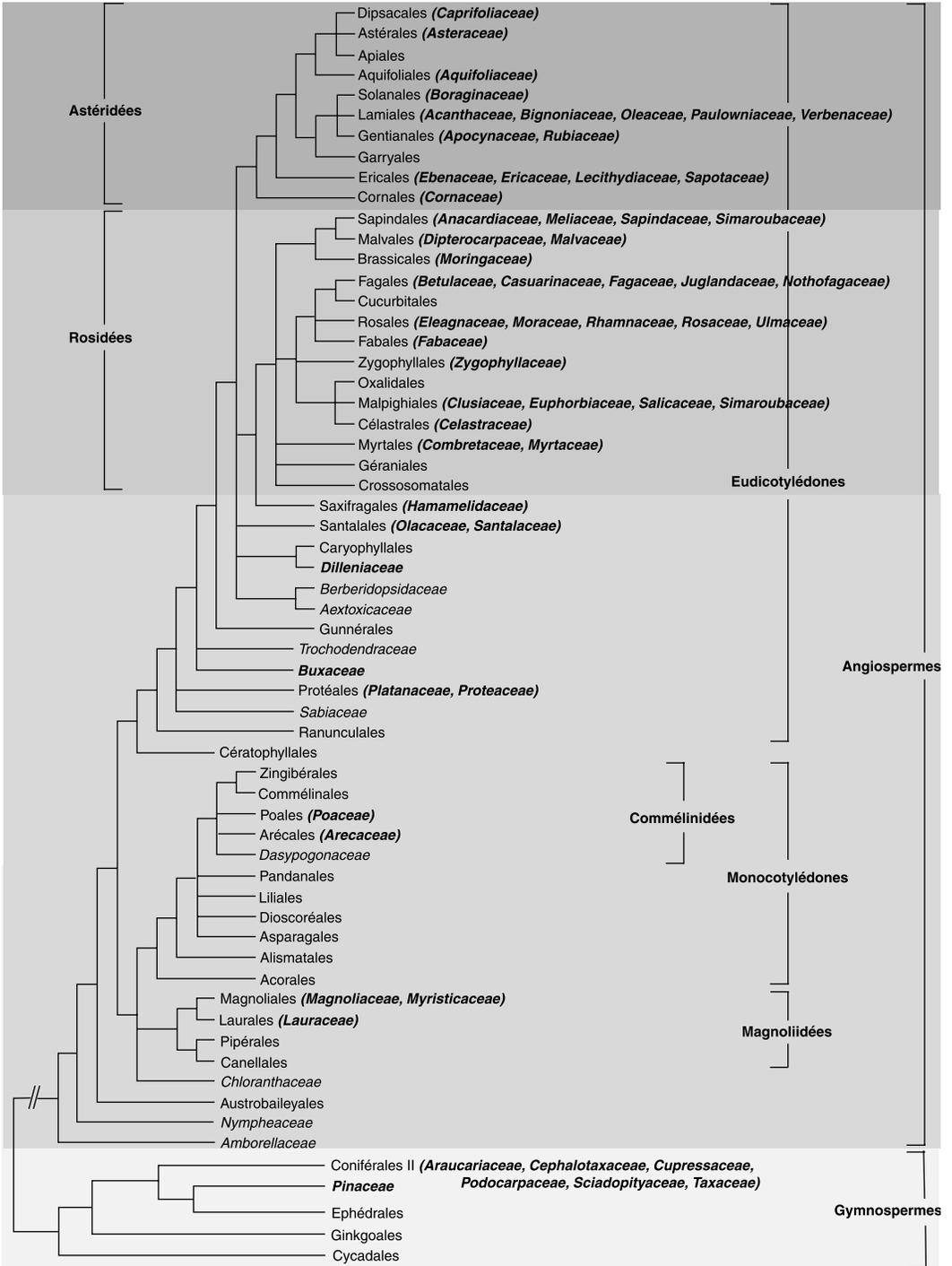


Figure 1 – Position systématique des principales familles comportant des espèces forestières importantes dans la classification des plantes supérieures (d'après Bowe *et al.*, 2000 ; Soltis *et al.*, 2000 ; The Angiosperm Phylogeny Group, 2003).

Brassicacées avec la plante de référence *Arabidopsis thaliana*, dont le génome a été entièrement séquencé (The *Arabidopsis* Genome Initiative, 2000). Malgré des traits opposés à ceux des arbres forestiers comme son cycle court (1 mois), son port herbacé, cette espèce au génome de petite taille peut servir de référence à de nombreux arbres forestiers pour un certain nombre de caractéristiques grâce à sa proximité systématique et phylogénétique.

Organisation des génomes des arbres forestiers

L'activité cellulaire des espèces végétales résulte de l'interaction de trois génomes : le génome nucléaire, le génome chloroplastique et le génome mitochondrial. Ces trois génomes sont caractérisés par leur localisation cellulaire. Chacun de ces génomes possède aussi ses particularités : hérédité, taille, nombre de copies, organisation des régions codantes et non codantes, expression et régulation des gènes, mode d'évolution. Le génome nucléaire est le plus important par la taille et la quantité d'informations. Il est porté par les chromosomes et est présent en deux exemplaires dans les cellules somatiques diploïdes, en plusieurs exemplaires dans les cellules polyploïdes et un seul exemplaire dans les cellules haploïdes (gamétophytes et gamètes). Il est redistribué par recombinaison génétique lors de la méiose. Les génomes chloroplastiques et mitochondriaux localisés respectivement au sein des chloroplastes et des mitochondries sont beaucoup plus petits, le nombre de copies par organelle et donc par cellule peut être très élevé (centaines ou milliers de copies par cellules).

Arabidopsis thaliana est la première espèce pour laquelle le génome nucléaire comme les génomes chloroplastiques et mitochondriaux ont été complètement séquencés (The *Arabidopsis* Genome Initiative, 2000). La comparaison des trois génomes est possible : elle montre des différences importantes sur la taille des trois génomes, leur contenu en gènes, les proportions de base G/C, la présence de pseudogènes (tabl. 1). La proportion du génome occupée par des gènes codant des protéines est faible pour le génome nucléaire (3,9 %) malgré le grand nombre de gènes. C'est le génome chloroplastique qui possède la plus grande part de séquences codantes des protéines (46,0 %). Le génome du riz, *Oryza sativa*, de 430.10⁶ pb, est également très largement séquencé (Yu *et al.*, 2002 ; Goff *et al.*, 2002). Des informations provenant des plantes herbacées sont utilisées dans cet ouvrage pour compléter celles souvent moins abondantes venant des espèces forestières.

Un arbre forestier est particulièrement bien connu, il s'agit de *Populus trichocarpa* dont le génome est maintenant complètement séquencé et en cours d'analyse. Le séquençage de son génome de 480.10⁶ pb, soit environ quatre fois plus gros que celui d'*Arabidopsis*, a été réalisé en moins de deux ans et s'est terminé fin 2004. Les premiers résultats concernant l'ensemble du génome de cette espèce sont utilisés ici (Brunner *et al.*, 2004 ; Sterky *et al.*, 2004, DiFasio *et al.*, 2005). L'analyse des séquences et l'identification des gènes demandent encore du temps

Tableau 1 – Caractéristiques des trois génomes présents chez les plantes, exemple d'*Arabidopsis thaliana* (d'après Unseld *et al.*, 1997 ; Sato *et al.*, 1999 et The Arabidopsis Genome Initiative, 2000).

Caractéristique	Génome nucléaire	Génome chloroplastique	Génome mitochondrial
Taille du génome (une copie) (pb)	125.10 ⁶	154 478	366 924
Copies/cellule	2	> 400	> 20
Taux de redondance (%)	60	17	10
Proportion de bases G/C (%)	34,7	36,3	44,8
Proportion de bases G/C (séquences codantes ^a) (%)	44,1	36,9	44,3
Gènes codant une protéine identifiée	25 498	67	30
Espace moyen occupé pour un gène (pb)	4 500	1 200	6 250
Taille moyenne de la séquence codante (pb)	1 900	900	860
Proportion du génome codant des protéines (%)	3,88	46,0	13,6
Gènes possédant des introns (%)	79	18	12
Pseudogènes / gènes (%)	3	0	20-50
Transposons (% taille du génome) (%)	14	0	4

a : y compris les ORF.

pour être achevées. La transcription des gènes doit être confirmée par d'autres analyses moléculaires. Des données sont également disponibles en particulier pour *Eucalyptus* et *Pinus*. Les particularités des *Populus* et *Eucalyptus* (taille réduite du génome, aptitudes à la multiplication végétative, aptitudes à la transformation génétique, évaluation rapide due à une croissance rapide, intérêt économique) en font un matériel de choix. Diverses informations sur le génome des arbres forestiers sont disponibles sur le site internet Dendrome : <http://dendrome.ucdavis.edu/index.html>. Des séquences de gènes sont disponibles et peuvent être comparées à d'autres par recherche d'homologie de séquences (site internet <http://www.ncbi.nlm.nih.gov/Database/index.html>). La taille des génomes chloroplastiques et mitochondriaux des arbres forestiers n'est connue que pour quelques espèces (tabl. 2). Les études de variabilité génétique peuvent porter sur chacun des trois génomes, mais les informations qu'ils apportent ne sont pas équivalentes.

Par convention, les noms des gènes sont écrits en italique, ceux des organites cytoplasmiques commencent par une minuscule, alors que ceux du génome nucléaire commencent par une majuscule (Commission on Plant Gene Nomenclature, 1993).

Tableau 2 – Taille des génomes chloroplastiques et mitochondriaux de quelques espèces forestières.

Espèce	Taille du génome chloroplastique (kpb)	Taille du génome mitochondrial	Référence
Gymnospermes			
<i>Cryptomeria japonica</i>	133		Tsumura <i>et al.</i> , 1993
<i>Ginkgo biloba</i>	158		Palmer, 1985
<i>Pinus contorta</i>	121		Lidholm et Gustafsson, 1991b
<i>Pinus koraiensis</i>	117		Noh <i>et al.</i> , non publié
<i>Pinus radiata</i>	120		Strauss <i>et al.</i> , 1988
<i>Pinus sylvestris</i>	120		Karpinska et Karpinski, 1993
<i>Pinus thunbergii</i>	120		Wakasugi <i>et al.</i> , 1994b
<i>Pseudotsuga menziesii</i>	121		Strauss <i>et al.</i> , 1988
Angiospermes			
<i>Eucalyptus nitens</i>	143		Steane <i>et al.</i> , 1991
<i>Eucalyptus nitens</i>	151		Byrne <i>et al.</i> , 1993
<i>Eucalyptus regnans</i>	143		Steane <i>et al.</i> , 1991
<i>Eucalyptus ovata</i>	143		Steane <i>et al.</i> , 1991
<i>Populus trichocarpa</i>	157		http://genome.ornl.gov/poplar_chloroplast
<i>Populus</i> sp.		420 - 450 kpb	Radetzky, 1990

Le génome nucléaire

La plupart des caractères génétiques sont codés par ce génome qui comporte la presque totalité des gènes. L'information génétique est portée par les chromosomes dont la forme et le nombre (caryotype) sont des caractéristiques de l'espèce. Les locus ont une répartition identique sur les deux chromosomes de la même paire chez les espèces diploïdes. Les deux formes occupant ces deux locus homologues sont appelées allèles. Il a cependant été montré par analyse de séquences chez *Zea mays* que des fragments homologues ne possèdent pas strictement les mêmes gènes et que cela ne résulte pas simplement d'une inversion locale (Fu et Dooner, 2002).

La répartition des paires de bases AT et GC varie selon les espèces : le pourcentage de GC est très stable chez les Gymnospermes (37-38 % : Miksche et Hotta, 1973) ; il est plus faible pour *Populus deltoides* (34 %, Dhillon, 1987) et plus élevé chez des espèces du genre *Quercus* (38-40 % : Zoldos *et al.*, 1998 ; 41-42 % : Favre et Brown, 1996). Chez *Arabidopsis thaliana*, les bases GC représentent de 33,4 à 35,5 % du génome selon le chromosome, leur pourcentage atteint 44,3 % dans les séquences codantes (tabl. 1) : les séquences codantes sont plus

riches en paires de bases GC que les séquences non codantes. La proportion des paires de bases AT et GC change aussi le long du chromosome. Les régions centromériques, constituées surtout d'hétérochromatine, renferment des séquences répétées diverses. Les séquences codantes y sont rares, elles sont localisées surtout dans l'euchromatine et sont réparties le long des bras chromosomiques, mais plus abondantes vers les extrémités des chromosomes (Schwarzacher, 2003 ; fig. 2a, voir planche 1).

Le génome nucléaire comporte les séquences codantes de quelques dizaines de milliers de gènes : 25 500 pour *Arabidopsis* (The *Arabidopsis* Genome Initiative, 2000), 42 000 pour *Populus* (les premières estimations faisaient état de plus de 58 000 gènes, <http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>), 45 000 chez *Oryza* (<http://rice.tigr.org>), 59 000 chez *Zea* (Messing *et al.*, 2004). *Populus* ne se distingue donc pas par le nombre de gènes. Pour une large majorité de gènes (près de 80 %, tabl. 1), les séquences codantes, appelées exons, sont entrecoupées de séquences non codantes dénommées introns. Chez *Arabidopsis*, les gènes comportent en moyenne 5 exons d'une longueur de 250 pb et les introns sont plus courts. Les séquences codantes n'occupent qu'une faible part de ce génome (moins de 4 %, tabl. 1).

Le génome nucléaire le mieux connu est celui d'*Arabidopsis thaliana* qui a été intégralement séquencé et dont l'analyse a commencé depuis quelques années (The *Arabidopsis* Genome Initiative, 2000). Son génome de 125.10^6 pb, l'un des plus petits du règne végétal, est réparti sur 5 chromosomes et comporte environ 25 500 gènes. La fonction d'un peu plus de la moitié des gènes est déterminée. Une part équivalente des EST (*Expressed Sequence Tags*, séquences du génome traduites en protéines identifiées à partir des ARN messagers) de *Populus* est identifiée (Sterky *et al.*, 2004). Les gènes sont classés en différentes catégories. La catégorie de gènes la plus abondante concerne le métabolisme cellulaire : 22 % chez *Arabidopsis*, 25 % chez *Oryza* (Goff *et al.*, 2002), 32 % chez *Populus* (Sterky *et al.*, 2004). Ceci est observé chez divers organismes. Les gènes impliqués dans la transcription du génome représentent environ 17 % des gènes identifiés chez *Arabidopsis* ; cette part est plus faible chez d'autres organismes dont *Oryza* et *Populus* (9 %, Sterky *et al.*, 2004). Les gènes de la croissance et de la division cellulaire représentent près de 12 % des gènes d'*Arabidopsis*, et 9 % des EST identifiés de *Populus*. Les gènes de défense, mort cellulaire et vieillissement sont légèrement moins abondants chez *Arabidopsis* (11 %). Les gènes d'adressage de protéines et de communication cellulaire représentent environ 10 % ; les derniers sont moins abondants chez *Populus* (environ 6 %). Les autres catégories de gènes sont moins abondantes chez *Arabidopsis* (transport intracellulaire : 8 %, transport : 5 %, synthèse protéique : 4 %) et *Populus*, excepté les gènes impliqués dans le métabolisme énergétique qui représentent 12 % des EST identifiés de *Populus*. Les gènes des différentes fonctions (métabolisme cellulaire, transcription, gènes de défense, transport...) sont portés dans des proportions très semblables par les cinq chromosomes, il n'y a donc pas de répartition des fonctions entre chromosomes.

Beaucoup de gènes réalisent la même fonction. Ils existent en un nombre variable de copies plus ou moins dispersées dans le génome, il s'agit de familles de gènes. Le nombre de copies de ces gènes et leur répartition dans le génome varient entre espèces. Les gènes ribosomiques sont un cas particulier dont les duplications très nombreuses sont distribuées en tandem (les unes à la suite des autres).

Une partie de cette redondance du génome a une autre origine. Le génome d'*Arabidopsis* montre une duplication d'origine ancienne (fig. 2b, voir planche 1). De grandes régions du génome sont retrouvées avec une organisation similaire à un autre endroit du génome. Les zones dupliquées peuvent être portées par un même chromosome ou un autre. Des parties du chromosome 1 d'*Arabidopsis* sont équivalentes à des portions des chromosomes 2, 3, 4 ou 5. À part les chromosomes 2 et 5 qui ne partagent pas de régions semblables, les autres chromosomes ont tous des portions en commun, parfois plusieurs, même éventuellement d'orientation différente et plus ou moins distantes. Cette redondance est distribuée en mosaïque. Bien que ce génome soit l'un des plus petits génomes végétaux, il présente une redondance importante. Ceci n'est pas une particularité d'*Arabidopsis thaliana*, une situation similaire a été observée pour des céréales.

Cette redondance structurelle du génome est particulièrement marquée chez *Populus trichocarpa* (Difazio *et al.*, 2005). Il s'agit encore de portions d'un chromosome qui sont retrouvées de façon similaire sur un chromosome d'une autre paire. Les parties homologues représentent une part importante du chromosome ; un chromosome donné ne présente d'homologie généralement qu'avec un seul autre chromosome, toutefois les chromosomes I et IV ont des régions homologues avec 4 autres chromosomes. Les chromosomes VIII et X sont très semblables. Cette structure correspond vraisemblablement à une autopolyploïdisation ou à une allopolyploïdisation impliquant des génomes proches qui a eu lieu anciennement et qui a ensuite évolué avec des redistributions de fragments chromosomiques. *Populus trichocarpa* semble donc être un polyploïde ancien, avec un fonctionnement actuel diploïde. Les régions communes aux paires de chromosomes actuels correspondent probablement, au moins pour les plus importantes, aux chromosomes ancestraux.

De très nombreuses plantes ont une origine polyploïde, cette situation n'est donc pas exceptionnelle. Ces duplications de génome, par autopolyploïdisation ou allopolyploïdisation conduisent à des séquences très semblables qui sont distribuées sur des chromosomes différents. Quand les gènes dupliqués ont gardé la même fonction et sont exprimés, le fonctionnement observé est identique à celui d'un polyploïde.

Caryotype et ploïdie

Le nombre de chromosomes est très stable dans des familles comme celle des Pinacées (Annexe 1), alors qu'il varie très largement dans certains genres comme le genre *Alnus*, voire dans certaines espèces. Les Gymnospermes ont de grands chromosomes, ce qui facilite les études de cytogénétique. Ainsi, le caryotype des

Pinacées comporte 24 chromosomes métacentriques ou submétacentriques sauf chez *Pseudotsuga menziesii* (22 chromosomes métacentriques ou submétacentriques et 4 télacentriques) et chez *Pseudolarix kaempferi* (40 chromosomes télacentriques et 4 submétacentriques) (Hizume, 1988). Les chromosomes télacentriques proviennent vraisemblablement de la fission des deux bras de chromosomes métacentriques mais les analyses d'Hizume et Kondo (1992) n'en ont pas apporté les preuves. Un nombre chromosomique plus élevé n'est donc pas nécessairement associé à une plus grande redondance de l'information génétique.

Quelques espèces forestières sont dioïques. Toutefois, le déterminisme du sexe est rarement lié au caryotype. Chez *Podocarpus macrophyllus*, les arbres mâles ont un chromosome en moins ($2n = 37$) que les arbres femelles ($2n = 38$) : c'est un rare cas de déterminisme chromosomique du sexe chez un arbre (Hizume *et al.*, 1988b). L'hérédité même du déterminisme sexuel d'une espèce dioïque est parfois difficile à comprendre, comme c'est le cas chez *Salix* (Alstrom-Rapaport *et al.*, 1997).

Des chromosomes surnuméraires ou chromosomes B ont été observés chez quelques espèces forestières dont : *Cupressus glabra* (Hunziker, 1961), *Picea glauca* (Teoh et Rees, 1977), *Picea glehnii* (Hizume *et al.*, 1988a), *Picea sitchensis* (Fox, 1987), *Salix sp.* (Büchler, 1986), *Quercus petraea*, *Q. robur* et *Q. rubra* (Ohri et Ahuja, 1990).

Des variations du niveau de ploïdie (nombre de copies du nombre haploïde n de chromosomes) sont observées au sein d'une même espèce ou entre espèces proches. De telles séries polyploïdes sont décrites dans les genres *Alnus*, *Betula*, *Salix* : dans le genre *Alnus* le nombre de chromosomes varie de $2n = 14$ chez *A. inokumae* (Chiba, 1962) à $2n = 112$ chez *A. firma* (Kodama, 1967). À l'intérieur de l'espèce *Alnus glutinosa* des variations du niveau de ploïdie sont aussi rapportées : $2n = 28$, $3n = 42$, $4n = 56$; le caryotype diploïde est cependant très largement répandu (Kammacher et Zygomala, 1986 ; Zygomala 1987). Les Gymnospermes présentent de rares cas de polyploïdie : *Sequoia sempervirens* est hexaploïde, mais son origine autopolyploïde ou allopolyploïde n'est pas élucidée (Schlarbaum et Tsuchiya, 1984 ; Rogers, 1997).

De nombreuses espèces végétales résultent de croisements interspécifiques (Rieseberg, 1997 ; Rieseberg et Carney, 1998) qui conduisent à l'association de deux génomes différents et donc à une redondance de l'information génétique. La méiose des hybrides ne donne généralement de gamètes viables qu'après doublement chromosomique. Les espèces tétraploïdes qui résultent de ce doublement du stock chromosomique après spéciation alloploïde sont alors des amphidiploïdes et se comportent finalement souvent comme des diploïdes. Les éventuels croisements en retour avec des individus des espèces parentes aboutissent à des situations plus complexes présentant des triploïdes ; leur stérilité n'est pas absolue (Ramsey et Schemske, 1998). L'évolution ultérieure du génome chez les allopolyploïdes peut conduire à une redistribution des génomes initiaux entre les chromosomes, comme chez le maïs (Wilson *et al.*, 1999).

Plusieurs génomes définis par une organisation et un caryotype sont parfois décrits dans un même genre pour caractériser des espèces ou groupes d'espèces. Ainsi, Liljefors (1955) a identifié quatre génomes dans le genre *Sorbus*. Beaucoup d'espèces d'origine hybride associent, selon ces observations, des génomes différents.

Taille du génome

La taille du génome haploïde, exprimée en nombre de paires de bases (pb) ou en masse d'ADN (pg), montre des fluctuations importantes entre espèces forestières et aussi une très nette différence entre les Gymnospermes ($10\text{-}100.10^9$ pb) et les Angiospermes ($0,2\text{-}1.10^9$ pb ; Annexe 1). Les Gymnospermes ont des tailles de génome parmi les plus grandes connues (Dhillon, 1987). Chez les Angiospermes, ce sont les Liliales et en particulier certains genres comme *Fritillaria* qui présentent les plus gros génomes nucléaires (Soltis *et al.*, 2003). La taille de génome de ces monocotylédones dépasse parfois largement celle des Gymnospermes. La grande taille du génome n'est donc pas une caractéristique des arbres forestiers ni des Gymnospermes.

La taille du génome montre aussi des variations intraspécifiques. Des fluctuations importantes sont observables d'une publication à une autre pour la même espèce. Les estimations en masse d'ADN et en nombre de paires de bases ne sont pas toujours concordantes. Même si les techniques utilisées et les erreurs expérimentales sont sources de fluctuations importantes, certaines variations intraspécifiques sont dues aussi au matériel végétal analysé (génotype et tissu).

Des tailles de génome plus grandes sont observées aux latitudes et altitudes les plus élevées chez *Picea glauca* et *Pinus banksiana* (Miksche, 1968), chez *Pseudotsuga menziesii* (El-Lakany et Sziklai, 1971). Les variations observées sont supérieures à un facteur 2. Chez *Pinus resinosa*, la latitude n'est pas corrélée aux fluctuations de taille de génome (Dhir et Miksche, 1974).

La taille du génome peut être considérée comme un élément de régulation de l'activité mitotique et de la taille des cellules pour Cavalier-Smith (1978). Les fluctuations observées sont donc susceptibles d'avoir une signification adaptative à des facteurs environnementaux. Les corrélations entre les variations de taille du génome de plusieurs espèces de *Pinus* et des paramètres écologiques (température, précipitation, sécheresse) sont en faveur de cette signification adaptative (Wakamiya *et al.*, 1993; 1996). La taille du génome semble fluctuer également en fonction du tissu, de l'âge de l'arbre, de la date de prélèvement pour un même tissu végétal aussi bien pour des cellules interphasiques que pour des cellules en mitose (Mellerowicz *et al.*, 1989 ; Zhong *et al.*, 1995 ; Wyman *et al.*, 1997).

Organisation et expression d'un gène

Un gène est un segment de la molécule d'ADN qui occupe un emplacement déterminé sur le chromosome, appelé locus. Il assure une fonction précise en