

Diversité génétique et histoire des populations d'arbres forestiers

Rémy J. Petit

▶ **To cite this version:**

Rémy J. Petit. Diversité génétique et histoire des populations d'arbres forestiers. Génétique des populations [q-bio.PE]. Université de Paris-Sud Orsay, 1999. ⟨tel-04725121⟩

HAL Id: tel-04725121

https://hal.science/tel-04725121v1

Submitted on 11 Oct 2024

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.



HAL Authorization

ORSAY
N° D'ORDRE :

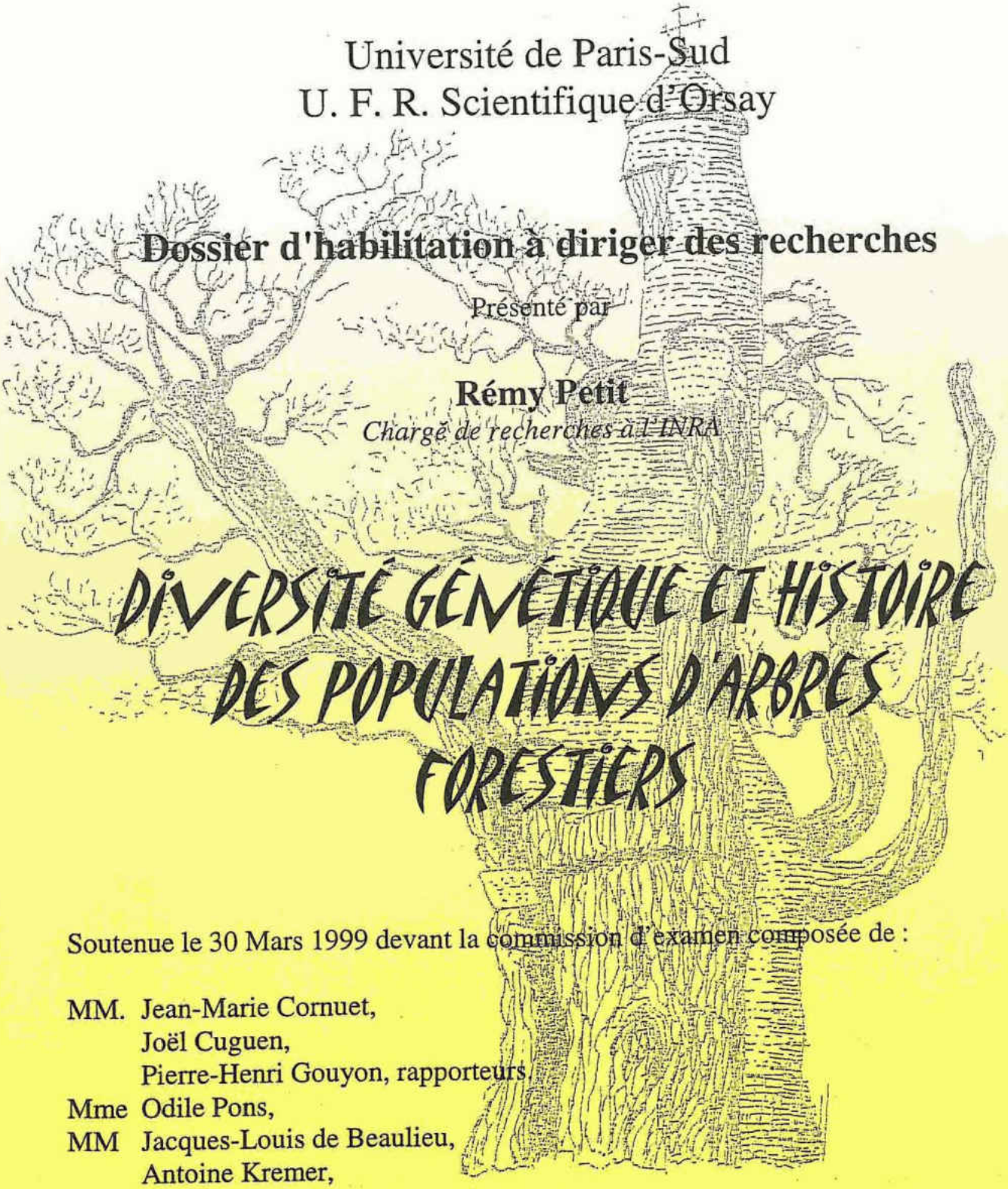
Université de Paris-Sud
U. F. R. Scientifique d'Orsay

Dossier d'habilitation à diriger des recherches

Présenté par

Rémy Petit
Chargé de recherches à l'INRA

DIVERSITÉ GÉNÉTIQUE ET HISTOIRE DES POPULATIONS D'ARBRES FORESTIERS



Soutenue le 30 Mars 1999 devant la commission d'examen composée de :

MM. Jean-Marie Cornuet,
Joël Cuguen,
Pierre-Henri Gouyon, rapporteurs,
Mme Odile Pons,
MM Jacques-Louis de Beaulieu,
Antoine Kremer,
Dominique de Vienne, examinateurs.

ORSAY
N° D'ORDRE :

Université de Paris-Sud
U. F. R. Scientifique d'Orsay

Dossier d'habilitation à diriger des recherches

Présenté par

Rémy Petit
Chargé de recherches à l'INRA

DIVERSITÉ GÉNÉTIQUE ET HISTOIRE DES POPULATIONS D'ARBRES FORESTIERS

Soutenue le 30 Mars 1999 devant la commission d'examen composée de :

MM. Jean-Marie Cornuet,
Joël Cuguen,
Pierre-Henri Gouyon, rapporteurs.
Mme Odile Pons,
MM Jacques-Louis de Beaulieu,
Antoine Kremer,
Dominique de Vienne, examinateurs.

Remerciements

A Brigitte Demesure, Sylvie Dumolin-Lapègue, thésardes d'hier; A Marie-Hélène Pemonge et Sabine Dumas, expertes ès électrophorèse : Je mesure tout ce que je vous dois, ce travail est largement le vôtre. A Delphine Grivet et Marie-France Deguilloux, thésardes d'aujourd'hui, un peu délaissées ces derniers temps... A tous les autres étudiants avec qui j'ai travaillé, et dont la diversité a été une richesse: DEA, stagiaires, thésards étrangers de passage...

A Antoine Kremer pour sa confiance et son aide, pour les orientations décisives et les discussions passionnées. A Michel Arbez pour son intérêt bienveillant. A Alexis Ducousso, pour tout ce qu'il m'a appris sur la forêt et pour son amitié.

A Odile Pons, pour tout le travail effectué ensemble et pour la très grande patience dont elle a fait preuve à mon égard. A Silvia Fineschi, Beppe Vendramin et Birgit Ziegenhagen, mes alter ego en Italie et en Allemagne... A Christian Burban, Joëlle Chat, Bernard Comps, Nasser Bahrman, qui m'ont fait confiance en m'associant à leurs travaux. A Pierre Taberlet et Ludovic Gielly, qui nous ont ouvert la voie à plusieurs reprises.

Aux membres du jury, qui ont accepté d'évaluer ce travail ; plusieurs ont suivi d'une manière ou d'une autre ces travaux et m'ont encouragé à diverses reprises.

A toute l'équipe de Pierroton, avec une mention spéciale pour Guy et Mimi, qui viennent me chercher tous les mercredis pour un tour de piste parmi les pins...

A Ellen enfin qui m'a encouragé et supporté pendant la rédaction de ce mémoire, à Elsa, Maryse et Annika qui auraient préféré que je leur laisse l'ordinateur plus souvent pour jouer à Adibou...

Table des matières

Remerciements 0

Table des matières 1

Introduction 4

Origine des études de diversité génétique utilisant les marqueurs biochimiques et moléculaires 4
Diversité génétique des arbres forestiers 5
L'étude de la diversité génétique en question 5
La diversité génétique : un ou plusieurs concepts ? 6
Un recentrage des études de diversité génétique autour de l'histoire des populations 6
Etudier des marqueurs neutres : en attendant mieux ? 8
Objectifs de ce travail 8

Chapitre 1 10

Mesure de la diversité et de la différenciation génétique: apports méthodologiques

INTRODUCTION 10

DIVERSITÉ DES DIVERSITÉS : LES MESURES POSSIBLES 10

RELATIONS ENTRE PARAMÈTRES DE DIVERSITÉ CHEZ LES ARBRES : ÉTUDE BIBLIOGRAPHIQUE 13

GÉNÉTIQUE DES POPULATIONS THÉORIQUE ET ANALYSE DE LA DIVERSITÉ 14

Théorie de l'échantillonnage des allèles 14
Indices de fixation : F_{ST} et cie. 15
Estimation des flux de gènes 16

IMPORTANCE DU CHOIX DES MARQUEURS 16

CONTRIBUTION À L'ESTIMATION DE LA DIVERSITÉ 17

Etude analytique, allèles non-ordonnés 18
Etude empirique, allèles non-ordonnés 19
Prise en compte de la distance entre allèles 19
Etude de la richesse allélique et introduction de la technique de raréfaction en génétique 20

Chapitre 2 23

Structuration de la diversité génétique de l'ADN chloroplastique et mitochondrial chez les plantes

INTRODUCTION 23

HISTORIQUE DE L'ÉTUDE DE LA VARIABILITÉ DES GÉNOMES CYTOPLASMIQUES CHEZ LES PLANTES 23

Hérédité des génomes cytoplasmiques des plantes 23
Compatibilité nucléo-cytoplasmique 24
Cytologie et hérédité cytoplasmique 25
Le début des années 70 : les bases de la biologie moléculaire des génomes cytoplasmiques 25

CONSÉQUENCES DE L'HISTOIRE RÉCENTE DES FORÊTS SUR LA STRUCTURATION GÉNÉTIQUE DES ESPÈCES 57
Paysage forestier et structuration de la diversité chloroplastique chez les chênes 57
Variabilité génétique de l'ADN chloroplastique dans la forêt domaniale de Compiègne 58
LA PALÉOGÉNÉTIQUE, UN NOUVEL OUTIL POUR ÉTUDIER L'HISTOIRE DES ARBRES FORÊSTIERS 59
Chapitre 4 61
Perspectives en matière de gestion et de conservation des ressources génétiques
INTRODUCTION 61
DIFFÉRENCIATION, FLUX DE GÈNES ET CONSERVATION 62
Différenciation génétique et flux de gènes 62
Identification des populations pour la conservation 64
GÉNÉTIQUE MULTI-ESPÈCES, ZONES REFUGES ET PRIORITÉS EUROPÉENNES 64
MIGRATIONS ET CHANGEMENTS CLIMATIQUES 65
CONSERVATION ET GESTION DES FORÊTS ANCIENNES 66
Conclusion 69
Références 71
Liste des publications 79

Introduction

« L'homme soucieux du devenir du monde doit réentendre la parole des feuilles. Elle en sait long sur l'obstination, la fragilité, la diversité fertile du vivant. Faite de mots témoins des origines, elle se prête encore à l'édification d'une syntaxe compréhensible à notre temps. »

Pierre Lieutaghi 1998. *La plante compagne. Pratique et imaginaire de la flore sauvage en Europe occidentale*. Actes Sud, Arles.

Origine des études de diversité génétique utilisant les marqueurs biochimiques et moléculaires

La diversité intraspécifique des espèces sauvages ou domestiques avait frappé les naturalistes du XIXème siècle. Toutefois, seule une partie de cette variabilité est attribuable au patrimoine génétique, le reste étant dû à l'action du milieu ou à son interaction avec les caractéristiques héréditaires. La possibilité d'accéder directement à cette diversité génétique fut rendue possible par l'apparition, au milieu des années 1960, des marqueurs biochimiques, au déterminisme génétique facilement interprétable (Harris 1966 ; Hubby et Lewontin 1966 ; Lewontin et Hubby 1966 ; figure 1). Désormais, les chercheurs n'étaient plus contraints de « stagner à la surface des phénotypes » (Powell 1994). Les premières études sur la structure génétique à l'aide de polymorphismes moléculaires sont en fait bien antérieures (Landsteiner 1900), mais le déterminisme génétique de ces marqueurs n'était pas connu.

L'ampleur du polymorphisme qui fut mis en évidence en 1966 et dans les années qui suivirent surprit la plupart des généticiens des populations, à l'exception de quelques chercheurs qui avaient pressenti ce résultat. Selon Muller (1950), en effet, tous les locus sont essentiellement homozygotes, et les mutations, généralement délétères, sont vite éliminées. Les mutations bénéfiques envahissent rapidement les populations et deviennent de nouveaux 'allèles sauvages'. Dobzhansky pensait au contraire qu'un individu devait être hétérozygote à la plupart de ses locus (Dobzhansky 1955). La notion d'allèle sauvage n'avait donc pas lieu d'être. Il ne s'agissait pas seulement d'une différence technique, mais bien de deux visions de l'humanité (Gouyon et coll. 1997). En effet, si la diversité au sein des populations –ou des races– est faible (à l'exception de quelques 'tarés', porteurs de mutations délétères), c'est que les différences entre races sont fixes et importantes (Muller était l'un des porte-parole du mouvement eugéniste). Si par contre la diversité intrapopulation est beaucoup plus élevée, les différences entre races n'en sont que le prolongement relativement modeste. On voit d'emblée l'importance des concepts de diversité et de différenciation et de leur relation, encore un peu flous à l'époque. Il faudra attendre le travail de Lewontin (1972) intitulé « *The apportionment of human diversity* », pour que soit quantifiée la part de la diversité liée

à la différenciation entre 'races'. A la suite de Dobzhansky et Lewontin, plusieurs généticiens s'attacheront à vulgariser le résultat principal de ces études, c'est à dire la fantastique diversité génétique, le caractère unique de chaque individu et la faible différenciation entre les populations humaines.

Après 1966, une nouvelle génération de chercheurs entreprit de caractériser la diversité génétique naturelle de nombreuses espèces (et plus seulement des espèces modèles que constituaient l'homme, la souris ou la drosophile), armés d'une batterie de *marqueurs génétiques* d'abord limitée -les protéines enzymatiques-, puis de plus en plus étendue (ADN visualisé par hybridation moléculaire avec des sondes homologues marquées, puis amplifié par la technique PCR et étudié à l'aide d'enzymes de restriction, voire directement séquencé).

Diversité génétique des arbres forestiers

L'application de la technique des isozymes aux plantes fut très rapide : la figure 2 illustre des résultats obtenus sur le blé dès 1967. Chez les arbres forestiers, après l'étude du contrôle génétique (Bartels 1971 ; Conkle 1971), les premières études de génétique des populations apparaissent. Sakai et coll. (1971) s'écrient : « *It seems that by introducing this new method we may attack the problem of measuring genetic variability in natural forests without spending years and space* ». En effet, à cette époque, de nombreux arbres forestiers faisaient déjà l'objet d'études de variabilité génétique détaillées, sous forme de tests de comparaison de provenances ou de descendances, dans des dispositifs de plusieurs hectares. Il semble que l'objectif consistait surtout, pour ces généticiens forestiers désormais rivés à la paillasse de leur laboratoire, à repérer des différences entre populations, plutôt qu'à comparer les niveaux de diversité. Les résultats seront assez décevants. Les différences escomptées ne sont pas toujours au rendez-vous : « *The resulting gene frequency distributions indicate highly polymorphic patterns within populations but only minor differentiation between individual populations* » (Bergmann 1973). Mais surtout, la différenciation moléculaire observée n'est généralement pas reliée aux caractères qui intéressaient initialement le forestier.

L'étude de la diversité génétique en question

Même si les généticiens se rendirent (plus ou moins) rapidement compte que les marqueurs biochimiques étudiés ne pouvaient généralement être reliés individuellement à des caractères phénotypiques particuliers, ils espéraient au moins, avec ces marqueurs, situer géographiquement les régions où l'espèce étudiée est différenciée ou particulièrement variable, pour y concentrer les efforts de conservation et d'amélioration génétique. Cela suppose que soient clairement définies la diversité génétique –pour les marqueurs-, et la variabilité -pour les caractères phénotypiques (Kremer 1994)-, et qu'une corrélation positive existe entre cette diversité moléculaire et la variabilité phénotypique. Des observations récentes sur des populations ayant subi de brusques réductions de leur effectif (par effet fondation), donc ayant nécessairement perdu de la diversité génétique, mais montrant une augmentation de la variance pour certains caractères phénotypiques (Carson 1990), illustrent la difficulté de ces comparaisons entre marqueurs et caractères adaptatifs, même si d'autres travaux sont moins pessimistes (Bataillon et coll. 1996). Le lien entre adaptabilité et diversité génétique reste donc difficile à mettre en évidence. Selon

Holsinger (1996), « *Surveys of molecular markers provide important insight into population structure, ecology, and systematics, but the levels of diversity they reveal are unlikely to provide reliable indicators of population viability or even reliable indicators of patterns of variability for traits subject to different selection pressures* ». Certains auteurs ont même prétendu que le maintien de la diversité (moléculaire) ne devrait pas constituer l'objet principal de la génétique de la conservation (Milligan et coll. 1994).

Cette position qui peut sembler surprenante n'est pas très éloignée de celle de Lewontin (1991) qui estime ainsi que « *the immense outpouring of data on genetic variation has also been a millstone around our necks. Its first effect was a considerable depauperization of the diversity of empirical work in evolutionary genetics. The second reason that electrophoresis is a fardel that we bear is that the result it has generated is so rich and so general, yet not, in itself, rich enough to solve the riddle of its own existence* ». Avise (1994, chap. 2) explique que historiquement les marqueurs moléculaires furent d'abord utilisés pour tenter de répondre à des questions très fondamentales (liées par exemple au débat neutralisme/sélectionisme des variants protéiques révélés par électrophorèse), avant d'être utilisés pour répondre à des questions plus concrètes, centrées autour de la biologie et l'histoire propre à chaque espèce étudiée. C'est parce que certains étaient allés trop loin en voulant tout expliquer avec ces marqueurs que ce rééquilibrage semble aujourd'hui nécessaire. C'est également parce qu'ils ont été fréquemment mal utilisés : la facilité qui consiste, à partir d'une étude utilisant des marqueurs moléculaires, aussi limitée soit-elle, à calculer de multiples paramètres génétiques et à interpréter longuement chaque valeur ainsi obtenue, sans trop s'interroger sur leur signification statistique, a conduit trop souvent à dévaloriser les études de diversité.

La diversité génétique : un ou plusieurs concepts ?

D'une manière surprenante, alors que la (bio)diversité, pour un écologiste, recouvre plusieurs notions et qu'il dispose pour la caractériser d'une pléthore de paramètres recouvrant différents concepts, ***la*** diversité génétique est le plus souvent définie comme la probabilité de tirer deux allèles différents (*hétérozygotie théorique*). C'est le succès de l'application en génétique des populations de l'indice de Simpson (1949) par Nei (1973) - indice rebaptisé au passage *gene diversity* et symbolisé par la lettre *H*- qui est à l'origine de cette simplification abusive. Les autres paramètres utilisés en génétique (nombre moyen d'allèles par locus *A*, pourcentage de locus polymorphes...) eurent beaucoup moins de succès. En outre, malgré une étude détaillée des propriétés statistiques du paramètre *H*, et les mises en garde répétées sur la nécessité de disposer de très nombreux locus pour pouvoir dégager des tendances générales (Nei 1978, Archie 1985), de nombreux résultats basés sur un petit nombre de locus et de populations furent publiés et interprétés sans précautions. Même si les progrès dans le domaine de l'analyse statistique des données de génétique des populations sont réels, de nombreux problèmes demeurent.

Un recentrage des études de diversité génétique autour de l'histoire des populations

Il serait sans doute instructif d'étudier l'évolution des causes principales mises en avant par les généticiens, depuis les années 1930, pour rendre compte des différences de

diversité génétique (ou plus précisément d'hétérozygotie) entre populations d'une même espèce (qu'elles soient ou non statistiquement fondées). Parmi les causes les plus souvent avancées, on distingue : (1) la diversité écologique de l'habitat ; l'importance de ce facteur a été particulièrement mise en avant, chez les plantes, par Allard (Allard et coll. 1978) ; (2) l'isolement par distance ; (3) la dérive dans des petite populations ; (4) les conséquences d'une utilisation par l'homme, ou (5) l'histoire des populations (zones refuges, zones secondaire de rencontre de populations d'origine différentes, effets fondation ou goulots d'étranglement). On peut plus simplement distinguer les explications faisant référence à des modèles de génétique des populations théorique à l'équilibre, de celles faisant appel à des événements historiques particuliers, plus ou moins récents.

Cette deuxième catégorie d'explications, faisant référence à des événements historiques ponctuels, est plus récente. Il me semble qu'il s'agit là d'une véritable révolution dans le domaine de la génétique des populations, liée à une nouvelle manière de considérer l'évolution des populations, même si la transition avec les vues précédentes fut particulièrement lente. Bennett, dans son livre récent « *Evolution and Ecology. The pace of life* », propose également de replacer l'évolution dans son contexte historique de changements climatiques permanents, liés aux variations d'orbite de la terre (figure 3), qui n'ont pas affecté la seule époque Quaternaire. Il souligne que l'étude des conséquences évolutives des changements climatiques du Quaternaire constitue une base appropriée pour la compréhension de l'évolution, intra- et interspécifique. Il semblerait que les évolutionnistes soient désormais prêts à considérer que l'histoire particulière à chaque organisme, avec toute la contingence qui y est associée (maintien d'une petite population dans une zone refuge, événement exceptionnel de colonisation à longue distance, goulot d'étranglement suite à la chute drastique des effectifs, événement *a priori* improbable d'hybridation interspécifique...), ait déterminé de manière majeure les phases ultérieures de l'évolution des espèces. Toutefois, de telles interprétations historiques ne peuvent être testées, si elles sont déduites exclusivement des données moléculaires. Elles doivent être confrontées à des données historiques obtenues indépendamment, rarement disponibles. Les arbres forestiers, de grande taille, souvent très abondants et produisant en quantité du pollen (pour les espèces anémophiles), constituent une heureuse exception, grâce aux études des paléobotanistes.

D'autres facteurs ont contribué à recentrer les études de génétique des populations autour de l'histoire particulière des populations constituant les espèces. A titre d'exemple, même si l'étude théorique des goulots d'étranglement n'est pas particulièrement récente (James 1971, Nei et coll. 1975, Watterson 1984), de nouveaux travaux et la mise au point de tests statistiques pour les détecter entraînent désormais un regain d'intérêt pour l'étude de ces *bottlenecks* (Cornuet et Luikart 1976). L'apparition de nouvelles catégories de marqueurs moléculaires permet également des avancées dans l'étude de l'histoire des populations. Les microsatellites, étudiés depuis le milieu des années 1990 chez les plantes, permettent de révéler un grand nombre d'allèles, valorisant ainsi les mesures de richesse allélique, souvent plus pertinentes pour comparer le niveau de diversité des populations et étudier les conséquences d'accidents démographiques.

L'étude de la variabilité de l'ADN chloroplastique, initiée au milieu des années 1980, permit les premières reconstructions phylogénétiques des variants : la chronologie de l'apparition des mutations pouvait désormais être établie au sein des espèces. Parce que la transmission est généralement uniparentale (maternelle chez la plupart des angiospermes, et paternelle chez une partie des gymnospermes), la recombinaison est sans effet, toutes les copies de l'ADN chloroplastique étant identiques au sein des cellules. L'existence de lignées asexuées au sein d'organismes fondamentalement sexués a permis l'essor de la phylogéographie : une lecture spatio-temporelle de l'évolution récente des espèces devient possible, comme l'a montré pour la première fois Avise (1979), en étudiant la variabilité de l'ADN mitochondrial de rongeurs nord américains. L'ordre d'apparition des mutations peut être en effet identifié, permettant d'accéder directement à la dimension historique. La perspective d'étudier génétiquement des restes fossiles, comme cela a été fait de manière systématique pour l'ADN mitochondrial du lapin par une équipe française (Hardy et coll. 1995) devrait permettre de progresser plus encore dans ce recentrage de la génétique des populations autour de l'histoire.

Etudier des marqueurs neutres : en attendant mieux ?

Dans ce contexte, étudier des marqueurs neutres n'est plus un handicap, puisqu'au contraire ils permettent une lecture plus directe de la trame historique qui a servi de base à l'adaptation des organismes. L'identification des gènes directement responsables de caractères adaptatifs progressant rapidement, ils pourront dans un proche avenir être utilisés pour caractériser les populations naturelles. Une référence 'neutre' permettant d'évaluer l'intensité des flux de gènes, l'origine des populations, leur âge, et l'existence d'accidents démographiques qui les ont affectées deviendra encore plus indispensable.

De plus, les caractères qu'il convient de conserver ne sont pas forcément connus à l'avance (par exemple des gènes de résistance pour des maladies mineures aujourd'hui, des propriétés pharmaceutiques encore inconnues, etc.). Baser une stratégie de conservation sur la distribution et la dynamique de la diversité génétique neutre (évaluée par des marqueurs appropriés) n'est donc pas forcément un pis-aller.

Objectifs de ce travail

Le présent document reprendra une partie des interrogations évoquées dans cette introduction. Il combinera des synthèses bibliographiques avec une présentation d'une sélection des travaux auxquels j'ai participé depuis mon recrutement à l'INRA fin 1988 : travaux d'étudiants que j'ai encadrés, collaboration avec des collègues du laboratoire de Génétique et Amélioration des Arbres Forestiers ou d'autres laboratoires. Des travaux déjà publiés mais aussi d'autres initiés plus récemment seront présentés. Les références à ces travaux sont données dans des notes de bas de page. Les autres références citées, ainsi qu'une liste de mes publications, figurent en fin de volume.

Au sein du laboratoire de génétique et d'amélioration, de nombreux autres travaux portent sur la diversité moléculaire, la variabilité de caractères adaptatifs, l'étude du système de reproduction et des flux de gènes, la cartographie et l'identification des bases moléculaires de l'adaptation des arbres à leur milieu. Même pour les chênes pédonculé et sessile (figure 4), qui ont constitué mon principal modèle d'étude pendant ces dix ans à l'INRA, on ne trouvera

donc pas ici un panorama complet des résultats disponibles au laboratoire. Parmi ceux-ci, on peut souligner : plusieurs études de diversité basées sur des marqueurs moléculaires (isozymes, marqueurs ADN), d'importants tests de provenance, l'obtention de familles issues de croisements contrôlés intra- et interspécifique, puis multipliées par bouturage, des études détaillées du système de reproduction (recherches de paternité et d'hybridation), une carte génétique très complète, et bientôt les premières recherches de marqueurs moléculaires liés à des caractères adaptatifs (dont ceux distinguant les deux espèces). Le site web du laboratoire (www.pierroton.inra.fr/genetics/) donne un aperçu général des travaux effectués et la liste des travaux publiés par mes collègues.

Le plan suivi ici sera le suivant : les problèmes méthodologiques liés à la mesure de la diversité et à l'interprétation de sa dynamique sont présentés dans le premier chapitre. Une revue bibliographique partielle de la littérature sur la variabilité des isozymes chez les arbres forestiers illustre les relations entre les différents paramètres proposés. Le deuxième chapitre est consacré à l'étude des génomes cytoplasmiques (historique de leur étude chez les plantes, présentation des travaux auxquels j'ai participé, et revue bibliographique (pratiquement exhaustive) des études de structuration génétique effectuées à l'aide de ces marqueurs. Le troisième chapitre tente de démontrer qu'il convient d'accorder une place centrale à l'histoire (Quaternaire ou même plus récente, liée à l'action de l'homme) pour interpréter la structuration génétique actuelle des arbres forestiers. La dernière partie aborde les conséquences de ces travaux pour la conservation des ressources génétiques, et présente plusieurs perspectives de recherches dans ce domaine.

Chapitre 1

Mesure de la diversité et de la différenciation génétique: apports méthodologiques

Lorsqu'on veut appliquer la géométrie et le calcul à des objets dont nous ne connaissons pas assez les propriétés pour pouvoir les mesurer, *« on est obligé de faire des suppositions toujours contraires à la nature, de dépouiller le sujet de la plupart de ses qualités, d'en faire un être abstrait qui ne ressemble plus à l'être réel, et lorsqu'on a beaucoup raisonné et calculé sur les rapports et les propriétés de cet être abstrait on croit avoir trouvé quelque chose de réel, et on transporte ce résultat idéal dans le sujet réel, ce qui produit une infinité de fausses conséquences et d'erreurs ».*
Buffon, De la manière d'étudier et de traiter l'histoire naturelle, Paris, 1749.

INTRODUCTION

Dans ce chapitre j'examine quelques aspects méthodologiques liés à la quantification de la diversité génétique à l'aide de marqueurs moléculaires. Après une présentation des mesures qui peuvent être utilisées, une analyse de données issues de la littérature sur la diversité génétique des arbres forestiers est effectuée. Les relations entre paramètres de diversité sont plus particulièrement étudiées. Ensuite, afin d'illustrer certaines interprétations qui peuvent être faites de ces paramètres descriptifs en génétique des populations, je présente très succinctement quelques développements théoriques dans ce domaine. Je décris ensuite deux travaux qui illustrent l'importance du choix des marqueurs pour mesurer la diversité du pin maritime, effectués collaboration avec Nasser Bahrman. Enfin, je résume les travaux portant sur l'analyse et l'estimation de la diversité, effectués pour l'essentiel en collaboration avec Odile Pons.

DIVERSITE DES DIVERSITES : LES MESURES POSSIBLES

Bien que la diversité semble être une notion familière (de même que la variation, le polymorphisme, l'hétérogénéité, la multiplicité, la complexité..), les quelques citations qui suivent expriment la difficulté d'arriver à un consensus sur la question de sa mesure :

« The recent literature on species diversity contains many semantic, conceptual, and technical problems. It is suggested that, as a result of these problems, species diversity has become a meaningless concept, that the term be abandoned, and that ecologists take a more critical approach to species-number relations... ». Hurlbert 1971.
« Despite extensive literature on diversity-related issues, a formal definition and logical development of diversity as a concept

and of its measurement have been lacking. » Patil et Taillie 1982.
« *We badly need a more-or-less consistent conceptual framework and a more-or-less usable measure of the value of diversity...* ». Weitzman 1992.

Il est clair que de nombreuses lacunes subsistent, d'où beaucoup de perspectives de recherche, en partie dues au développement des techniques moléculaires qui fournissent des données qualitativement et quantitativement plus riches que par le passé (plus grande richesse allélique, possibilité de mesurer les similarités entre allèles, voire de déduire des relations généalogiques entre les variants). Le premier problème est de définir ces paramètres de diversité et de différenciation ; ensuite, des estimateurs de ces paramètres et de leur variance pourront être obtenus.

Quand on considère la diversité génétique au niveau de séquences homologues, une possibilité est d'énumérer le nombre de séquences différentes. Toutefois la mesure obtenue dépend du nombre de séquences analysées. Une définition plus précise est donc le nombre de séquences différentes dans un échantillon de taille donnée n : S_2, S_3, S_4... S_n. Afin que la diversité soit nulle quand toutes les séquences sont identiques, on peut soustraire un à la valeur obtenue. Toutefois, si la séquence est longue, la mesure peut 'saturer', si toutes les séquences analysées diffèrent ($S_n = n - 1$) (autant de séquences différentes que de séquences étudiées). D'autre part, on ne distingue pas avec une telle mesure les séquences quasi identiques de celles qui diffèrent en de nombreux nucléotides. Pour corriger cela, on peut introduire une distance entre les séquences. Pour une paire de séquences, il pourra s'agir de la proportion de nucléotides qui diffèrent, ou d'une autre mesure de distance *ad hoc*. Pour un triplet, un quadruplet, etc. de séquences, la pondération pourra prendre en compte la longueur totale de l'arbre reliant les séquences (*minimum spanning tree*), et non plus les distances deux à deux (exemple : Weitzman 1992 ; Faith 1992 ; voir figure 1).

On peut illustrer facilement ceci dans le cas de l'indice de Simpson (1949) (*alias* polymorphic index, Marshall et Jain (1969), *alias* diversité de Nei (1973), *alias* hétérozygotie théorique, que j'appellerai ici '*diversité d'ordre 2*').

Considérons une population subdivisée en un grand nombre de populations parmi lesquelles I séquences (allèles) ségrègent. Soit p_i la fréquence de la séquence (allèle) i dans la population.

La diversité d'ordre 2 dans la population est :

$h = 1 - \sum_i p_i^2$, qui s'écrit aussi : $h = \sum_{i \neq j} p_i p_j$.

Introduire une pondération (distance entre la séquence i et j) est simple ; soit π_{ij} la distance entre les séquences i et j. On a : $v = \sum_{i \neq j} \pi_{ij} p_i p_j$ (ex : Pons et Petit 1996[1]).

En ce qui concerne l'étude de la diversité génétique, les mesures de diversité d'ordre supérieur à deux n'ont guère été développées ; le raisonnement est donc presque toujours basé sur la comparaison de paires de séquences, d'allèles ou d'haplotypes, ou sur des mesures de richesse allélique non standardisées, ou enfin sur des

[1] Pons O et R J Petit 1996 Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* **144**, 1237-1245.

mesures mixtes et peu précises (pourcentage de locus polymorphes : le 'polymorphisme' étant fixé à un seuil arbitraire, 5% ou 1% en général). Avant de chercher à développer des mesures alternatives, une question importante est de savoir si ces mesures sont très corrélées entre elles. Peu de données empiriques existent à ce sujet. DeBenedictis (1973) a étudié les relations entre divers indices utilisés en écologie, et fait remarquer que des dépendances mathématiques sont présentes, et qu'une corrélation positive est donc généralement attendue. La figure 2 montre toutefois que l'utilisation de la diversité d'ordre 2 (H) et d'ordre n (richesse allélique A) peut aboutir à un classement différent des populations. Des revues ont abordé la corrélation entre l'hétérozygotie théorique moyenne H et la richesse allélique A (sans standardisation), au niveau interspécifique. Par exemple, Hamrick et coll. (1979) trouvent une corrélation de 0,80 sur un ensemble de 100 espèces de plantes pour lesquelles A et H ont été mesurés. Il y a par contre peu d'études qui ont examiné cette corrélation entre populations d'une même espèce (*cf.* Brown et Schoen 1994).

Quand on étudie la diversité de plus d'une population, il y a deux niveaux hiérarchiques ; on peut définir la diversité moyenne intrapopulation V_S et la diversité totale V_T, calculée sur l'ensemble des populations. La différence entre les deux niveaux $V_T - V_S$ est la différenciation (ou diversité interpopulation). Les écologistes parlent de diversité α et β, il s'agit des mêmes concepts. Le coefficient de différenciation (souvent appelé abusivement différenciation) est le rapport $(V_T - V_S)/V_T$. Parce qu'il est sans unité, ce coefficient se prête bien à des comparaisons (entre locus, entre espèces analysées différemment, entre classes de marqueurs, etc.).

Cette décomposition de la diversité entre niveaux hiérarchiques différents a été surtout appliquée à la diversité de Nei (G_{ST}), même si l'une des premières applications, par Lewontin (1972), était basée sur l'indice de Shannon (indice de diversité issu de la théorie de l'information). Cette décomposition peut parfaitement s'appliquer à n'importe quelle autre mesure, par exemple le nombre d'espèces après standardisation (ou la richesse allélique en génétique). S'il est bien connu désormais que la subdivision de la diversité peut dépendre des marqueurs utilisés (caractères quantitatifs *versus* marqueurs moléculaires, marqueurs cytoplasmiques *versus* marqueurs nucléaires), on peut se demander dans quelle mesure cette subdivision dépendra de la mesure de diversité utilisée. Ceci sera également illustré plus loin.

Une autre manière de quantifier la répartition de la diversité consiste à étudier la variation du niveau de diversité d'une population à l'autre (Brown 1978 ; Schoen et Brown 1991). En effet, il est possible d'imaginer par exemple des populations ayant exactement le même niveau de diversité, mais totalement différentes dans leur composition allélique. Une manière d'évaluer l'importance de ce phénomène est d'estimer le coefficient de variation de la mesure de diversité considérée. La figure 3 illustre l'aspect complémentaire de ces deux coefficients.

Les études de diversité génétique étant assez codifiées, de nouveaux paramètres, même pertinents, comme ce coefficient de variation du niveau de diversité, mettent beaucoup de temps à être assimilés. Simultanément, des paramètres d'intérêt plus discutable, comme le pourcentage de locus polymorphes, continuent à être utilisés largement.

RELATIONS ENTRE PARAMETRES DE DIVERSITE CHEZ LES ARBRES : ETUDE BIBLIOGRAPHIQUE

Une recherche bibliographique a été effectuée pour étudier les relations entre paramètres de diversité, à partir des études utilisant les isozymes chez les arbres forestiers[2]. Un total de 62 études (dont 38 concernant des gymnospermes et 24 des angiospermes) a été utilisé (une espèce peut être représentée plusieurs fois, plusieurs espèces peuvent être représentées dans une même publication) (Table 1).

Les critères de sélection étaient les suivants : au moins 9 locus polymorphes et 10 populations étudiées (un nombre plus faible de locus polymorphes a été toléré chez les Fagacées, plus particulièrement étudiées au laboratoire). Le nombre moyen d'individus diploïdes échantillonnés par population (ou d'équivalents diploïdes, dans le cas d'études basées sur les mégagamétophytes haploïdes des conifères) a été calculé ; quand ce nombre était trop hétérogène d'une population à l'autre, les populations d'effectifs extrêmes ont été éliminées de l'analyse, pour que la comparaison des niveaux de richesse allélique garde un sens. Les paramètres de diversité estimés sont les suivants: pourcentage de locus polymorphes ; nombre moyen d'allèles par locus au niveau de l'espèce (A_{SP}) ; diversité H ; nombre moyen d'allèles par locus et par population (A_P) ; corrélation entre H et A_P ; coefficient de variation de H et de A_P ; coefficient de différentiation G_{ST}. Les résultats sont donnés dans la table 1. Le nombre moyen de populations étudiées est 27 (20 sans le hêtre et le chêne sessile) ; le nombre moyen de locus polymorphes est 17 et le nombre moyen d'individus (diploïdes) par population est de 48. En ce qui concerne les mesures de diversité, on trouve les valeurs moyennes suivantes : $A_{SP} = 2,75$, $A_P = 2,03$, $H = 0,17$, $P = 0,84$. Le coefficient de variation de H vaut 0,17, celui de A_P vaut 0,20, et $G_{ST} = 0,10$. Enfin, la corrélation entre H et A_P est de 0,50. Ces valeurs peuvent être comparées à celles données par Hamrick et coll. (1992), à partir de 584 études de plantes : $A_{SP} = 1,97$, $A_P = 1,52$, $H = 0,11$, $G_{ST} = 0,23$. Pour les 196 plantes ligneuses pérennes étudiées, les valeurs se rapprochent plus des nôtres : $A_{SP} = 2,22$, $A_P = 1,76$, $H = 0,15$, et $G_{ST} = 0,08$. Les arbres sont donc comparativement variables et peu différenciés.

Les coefficients de variation n'avaient pas été étudiés par Hamrick et coll. Toutefois, Schoen et Brown (1991) trouvent une valeur moyenne de 0,64 pour les espèces autogames, et de 0,12 pour les allogames. La valeur trouvée ici pour les arbres forestiers (tous allogames) est de 0,17. La corrélation moyenne intraspécifique entre les estimateurs des deux paramètres A_P et H n'avait jamais été évaluée. La figure 4 illustre la distribution des valeurs observées ; la valeur médiane est d'environ 0,70, mais un nombre non négligeable de valeurs sont beaucoup plus faibles, voire négatives (ex : le chêne sessile *Quercus petraea*, le hêtre *Fagus sylvatica*).

La matrice des corrélations entre les caractéristiques de l'échantillonnage et les estimateurs des paramètres de diversité est donnée dans la table 2. L'objectif est en effet d'évaluer si l'échantillonnage peut être cause de biais dans l'estimation de ces paramètres, notamment du fait que les mesures de richesse allélique ne sont pas standardisées. Ensuite, il s'agit d'examiner les relations entre paramètres de diversité. Toutes les

[2] Petit R J et B Comps. Résultats non publiés.

valeurs supérieures à 0,25 sont significatives (P<0,05).

On peut remarquer que le coefficient de corrélation entre A_P et H (r(H, A_P)) est négativement lié au nombre d'individus par population (-0,31) : cela traduit la forte relation entre la diversité d'ordre 2 et la diversité d'ordre n quand n se rapproche de 2, et démontre la dépendance de la richesse allélique vis-à-vis de la taille de l'échantillon. Par contre, la corrélation entre A_S et le nombre de populations est plus surprenante.

Les mesures soulignant l'hétérogénéité du niveau de diversité ou des fréquences alléliques entre populations (CV_H et G_{ST}) sont positivement corrélées avec $r(H,A_P)$ (0,34 et 0,36 respectivement). Par ailleurs, quand la richesse allélique est élevée (grand nombre d'allèles par locus, donc présence de nombreux allèles rares), les deux mesures de diversité deviennent plus indépendantes (corrélation entre r(H,A_P) et A_P significativement négative). (Quand le nombre d'allèles est réduit à 2, la perte d'un allèle entraîne aussi la perte d'hétérozygotie, d'où cette relation).

Une corrélation élevée entre H et A_P existe au niveau interspécifique (0,68), déjà mise en évidence par Hamrick et coll. (1979). La valeur est plus élevée qu'au niveau intraspécifique, de population à population (corrélation moyenne de 0,50 à ce niveau). En outre, la corrélation entre H et CV_H ou CV_A est fortement négative (-0,50 et –0,45) ; la corrélation est également négative entre A_P et CV_H ou CV_A : les arbres qui maintiennent les plus hauts niveaux de diversité ou de richesse allélique intrapopulation sont ceux qui la répartissent de manière homogène. Les espèces concernées (*Cunninghamia lanceolata*, *Pinus sylvestris*, *Robinia pseudoacacia*, *Fagus sylvatica*, *F. crenata*, *Larix decidua*, *Quercus petraea*...) sont caractérisées par une aire de répartition large et souvent continue facilitant les flux de gènes. La corrélation négative entre H et G_{ST} avait été décrite par Hamrick et coll. (1992) chez les plantes. Les résultats présentés ici indiquent que le coefficient de variation de H permet de mieux prédire que le G_{ST} le niveau de diversité des populations.

D'une manière générale, les mesures de diversité étudiées (H, A_P, et A_{SP}) ainsi que celles mesurant la distribution de cette diversité (CV_H, CV_A, et G_{ST}) covarient positivement, mais ne sont pas redondantes.

GENETIQUE DES POPULATIONS THEORIQUE ET ANALYSE DE LA DIVERSITE

Dans le riche corpus de la génétique des populations théoriques, avec ses modèles sous-jacents d'évolution des populations, interviennent de nombreux paramètres étroitement liés à ceux plus descriptifs utilisés ci-dessus. J'en donnerai quelques exemples.

Théorie de l'échantillonnage des allèles

Ewens (1972) a étudié le problème de l'échantillonnage des allèles sélectivement neutres. Il calcule notamment l'espérance et la variance de A (noté k dans son article), le nombre d'allèles dans un échantillon de taille n. Un modèle stochastique d'évolution des populations, adapté du modèle standard de Wright (1931), et incluant la dérive, la mutation (nombre d'allèles infini), des générations discrètes, une taille efficace constante etc. est utilisé. L'hypothèse qu'un équilibre (état stationnaire) est atteint (pour la probabilité F que deux gènes tirés au hasard de la

population aient le même allèle) est utilisée. La figure 5 illustre la relation entre A et H pour différentes valeurs de $\theta = 4\ N_e u$ (où N_e est la taille efficace de la population et u le taux de mutation vers de nouveaux allèles). On constate que la relation n'est pas linéaire (elle l'est toutefois pour $0{,}0 < H < 0{,}6$). Au delà, la pente s'infléchit et la relation entre A et H devient plus forte. Comme on a vu plus haut que pour de faibles valeurs de diversité, la corrélation entre H et A est artificiellement augmentée, ces résultats montrent que la dépendance entre ces deux paramètres sera la moins forte pour des hétérozygoties intermédiaires.

Là où les écologistes (Hurlbert 1971) proposaient la technique de raréfaction -qui sera décrite plus loin- pour estimer le nombre d'espèces dans un échantillon de taille donnée, sous forme d'un simple calcul probabiliste sur un échantillon observé, les généticiens des populations étudiaient dans le cadre d'un modèle très circonscrit la théorie de l'échantillonnage *ad hoc* des allèles, très complexe.

Chakraborty et Fuerst (1979) élaborent encore ce modèle de Ewens pour tenir compte du taux variable de mutation entre protéines (qui était supposé constant précédemment). Ils obtiennent de nouvelles formulations pour l'espérance et la variance de k. De plus, ils étudient la corrélation entre A et H. Ils notent que A est plus sensible au taux de mutation. Comme l'indique la figure 6, tirée de leur étude, la corrélation dépend bien de l'échantillon considéré ($n = 50$, 300 et 500 dans la figure). En outre, la relation est inversée selon que le taux de mutation est homogène ou non (modélisé suivant une distribution gamma, Nei et coll. 1976). L'analyse de Chakraborty et Fuerst prédit des coefficients de corrélation de 0,725 entre H et A en cas de taux de mutation constant, pour $n \approx 100$ et $H \approx 0{,}15$ (chiffres s'approchant des moyennes obtenues dans la revue bibliographique de la diversité chez les arbres forestiers présentée ci-dessus). Par comparaison, les valeurs observées étaient $r(H,A) = 0{,}50$ (niveau intraspécifique) et 0,68 (niveau interspécifique). La première valeur est nettement inférieure à celle prédite. Toutefois, la médiane de la distribution (*cf.* figure 4) n'est pas tellement éloignée de la valeur prédite, indiquant que ce n'est que pour un petit nombre d'espèces que l'adéquation avec la théorie est mauvaise. Chakraborty et Fuerst trouvaient aussi une bonne adéquation du coefficient de corrélation prédit et observé au niveau interspécifique.

Indices de fixation : F_{ST} et cie.

Un autre exemple est fourni par les nombreuses interprétations, définitions et estimations de l'indice de fixation de Wright (F_{ST}), du coefficient de différentiation de Nei (1973) (G_{ST}) ou de Pons et Petit (1995), ou enfin du f de Weir et Cockerham (1984). Parmi les papiers qui discutent de cette question, on peut noter ceux de Nei (1986) et de Chakraborty et Danker-Hopfe (1991). Les principales différences sont liées à l'existence ou non d'un modèle d'évolution sous-jacent, dont dépend le modèle d'échantillonnage des populations (fixe ou aléatoire), et à l'utilisation d'un modèle d'échantillonnage multinomial ou d'analyse de variance. Dans le modèle de Weir et Cockerham, tel qu'il est exposé, les populations se sont différenciées au cours du temps. Une analyse de variance des fréquences alléliques est effectuée, et les populations sont considérées comme aléatoires. Dans le modèle de Pons et Petit, une population totale est subdivisée en un grand nombre de populations indépendantes (modèle aléatoire également). Un modèle

d'échantillonnage multinomial des individus dans les populations et des populations est utilisé pour les estimations. Malgré la différence entre les modèles, **les estimateurs obtenus sont les mêmes** (au moins quand la taille des échantillons est identique dans toutes les populations ; en cas d'hétérogénéité, le même poids est donné aux populations dans le traitement de Pons et Petit, alors que dans l'analyse de variance le poids de la population dépendra de son effectif). Par contre, ces estimateurs diffèrent de celui de Nei et Chesser (1983) (la différence est minime quand le nombre de populations est grand ; toutefois, quand le nombre de populations est petit ou que le G_{ST} est utilisée comme une mesure de distance génétique entre paires de populations, la différence est loin d'être négligeable). En fait, l'hypothèse d'indépendance des populations, jugée intenable par Nei (1986) et Weir et Cockerham (voir aussi Robertson (1975) dans sa critique du test de Lewontin et Krakauer (1973) ; figure 7), est utilisée par tous ces auteurs de manière implicite.

Estimation des flux de gènes

L'objectif de certains de ces travaux est d'estimer indirectement les flux de gènes : certains paramètres et/ou estimateurs seraient plus appropriés que d'autres pour cela (Cockerham et Weir 1993 ; Rousset 1997). Le danger est de confondre précision de l'estimateur de différenciation et précision dans l'estimation des flux de gènes ($\hat{N}m$).

Un papier récent de Wakeley (1998) propose une nouvelle manière d'estimer les flux de gènes dans une population subdivisée (modèle en île). Wakeley remarque tout d'abord que le nombre de sites variables dans une séquence (S, analogue à la richesse allélique) est un descripteur utile de la variabilité d'une population. Il utilise la théorie de la coalescence des allèles (Kingman 1982ab ; Tajima 1983 ; Hudson 1983) pour étudier l'espérance et la variance de S et le rapport de S intra- et interpopulation. Il démontre que le taux de migration est mieux estimé à l'aide de ce rapport, plutôt qu'en utilisant les méthodes basées le F_{ST} (c'est à dire en considérant des paires de séquences). Je montrerai plus loin que la répartition inter- et intrapopulation de la richesse allélique peut être évaluée grâce à la technique de raréfaction.

IMPORTANCE DU CHOIX DES MARQUEURS

La démonstration de la très forte diversité des protéines, en 1966, dont il a été question dans l'introduction générale, a donné jour a toute une littérature destinée à comprendre pourquoi telle protéine sera plus variable qu'une autre (pour une revue voir Ward et coll. 1992). Parmi les facteurs susceptibles d'affecter la variabilité des protéines, on peut citer les suivants : la taille (nombre d'acide aminés) ; la structure quaternaire ; la fonction ; la localisation dans la cellule ; le mode d'expression ; l'abondance ; etc. Ces questions peuvent sembler un peu éloignées des autres préoccupations de ce travail ; toutefois, un problème qui peut se poser quand on souhaite estimer la diversité génétique est la représentativité des marqueurs étudiés. De plus, ces informations de base seront utiles pour comprendre l'évolution des protéines, à un niveau très général.

La technique d'électrophorèse bidimensionnelle permet de révéler un très grand nombre de protéines, avec toutefois un biais en faveur de celles qui sont les plus

abondantes. Elle est particulièrement adaptée à l'étude de ce genre de questions.

La comparaison des protéines provenant de trois organes différents (pollen, aiguilles, bourgeons) de 18 pins maritimes a été effectuée. Les protéines les plus lourdes ont plus de variants alléliques (changements de charge) (Bahrman et Petit 1995[3]). Par contre, elles ne montrent pas plus de variation quantitative, tendant à démontrer que ces variations de quantité sont le fait de la variation d'autres protéines régulatrices.

En outre, les protéines organe-spécifiques sont beaucoup plus variables que les protéines ubiquistes (elles montrent deux fois plus de variation allélique, et cinq fois plus de variation quantitative ; figure 8).

Il est possible que les contraintes fonctionnelles réduites, dans le cas de protéines exprimées dans un seul environnement cellulaire, expliquent ce nombre plus grand d'allèles. De même, leur régulation plus complexe (comparée aux protéines exprimées dans tous les organes), en augmentant le nombre de sites de contrôle et en multipliant les possibilités de mutation, pourrait expliquer l'augmentation des variations quantitatives (de Vienne et coll. 1988 ; Bahrman et coll. 1997[4]).

Ces résultats devraient être présents à l'esprit des chercheurs étudiant les bases moléculaires de caractères adaptatifs : les protéines généralistes, largement prépondérantes (90% des protéines détectées dans un organe donné) sont les moins variables à tout point de vue, donc celles qui répondront le moins à la sélection. Les enzymes étudiées dans les études de diversité sont en général exprimées dans tous les organes, donc, d'après les résultats qui viennent d'être présentés, potentiellement moins variables que d'autres protéines plus spécifiques.

Dans une autre étude, toujours chez le pin maritime, la différentiation a été comparée pour trois types de marqueurs (terpènes, isozymes et protéines totales révélées par électrophorèse bidimensionnelle) (Petit et coll. 1995)[5].

Pour ces trois marqueurs, le déterminisme génétique avait été étudié. Il s'agissait d'ailleurs de la première étude de génétique des populations utilisant les protéines totales chez les plantes. Une hypothèse était que les terpènes, plus soumis à sélection, seraient plus différenciés que les marqueurs protéiques. En effet, les différences géographiques observées dans l'aire du pin maritime pour certains au moins de ces terpènes avaient été interprétées comme étant probablement adaptatives (Baradat et Marpeau-Bézard 1988). Les résultats ont montré que la différenciation était remarquablement similaire pour les trois classes de marqueurs (estimation proche de 0,17 dans chaque cas).

CONTRIBUTION A L'ESTIMATION DE LA DIVERSITE

Comme on l'a vu au travers de la présentation rapide de certains développements théoriques, anciens ou plus récents, il reste encore de nombreux progrès à réaliser pour disposer de définitions et d'estimations de paramètres de diversité très généraux (qui ne soient pas basés sur des

[3] Bahrman N et R J Petit. 1995. Genetic polymorphism in maritime pine (*Pinus pinaster* Ait.) assessed by two-dimensional gel electrophoresis of needle, bud, and pollen proteins. *J. Mol. Evol.* **41**, 231-237.

[4] Bahrman N, C Plomion, R J Petit et A Kremer. 1997. Contribution of two-dimensional electrophoresis of proteins to maritime pine genetics. *Ann. Sci. For.* **54**, 225-236.

[5] Petit R J, N Bahrman et Ph Baradat. 1995. Comparison of genetic differentiation in maritime pine (*Pinus pinaster* Ait.) estimated using isozyme, total protein and terpenic loci. *Heredity* **75**, 382-389.

modèles d'évolution trop restrictifs) mais qui permettent néanmoins des comparaisons intéressantes avec les modèles théoriques qui se développent. En collaboration avec Odile Pons, l'étude statistique des paramètres de diversité classiques, d'ordre deux, a été entreprise (génétique des paires), en tenant compte de la similarité des séquences ou allèles. J'ai essayé également de prospecter quelques pistes en ce qui concerne l'étude de la diversité d'ordre n ($n > 2$), en appliquant en génétique la technique de raréfaction des écologistes.

Etude analytique, allèles non-ordonnés

Nous avons proposé (Pons et Petit 1995[6]) de nouvelles définitions et estimations non biaisées des indices de diversité dans une population subdivisée.

Le cas haploïde a été étudié en premier, puis Pons et Chaouche (1995) ont étudié le cas diploïde. Les définitions et estimations supposent explicitement l'indépendance des populations, considérées de plus comme aléatoires. Les variances (totale, intra- et interpopulation) de ces paramètres (h_S, h_T, G_{ST}) ont été définies et des estimateurs sans biais obtenus (sous forme d'approximations dans le cas de $\mathrm{Var}(\hat{h}_T)$ et $\mathrm{Var}(\hat{G}_{ST})$). Le choix de commencer par le cas le plus simple (haploïde) avant de poursuivre par l'analyse du cas diploïde s'est avéré judicieux, vu la sophistication mathématique de l'analyse, dans le cas des estimateurs des variances et covariances notamment.

Ces travaux ont permis de définir un échantillonnage optimal (répartition de l'effort de l'échantillonnage inter- et intrapopulation qui minimise la variance du G_{ST}). La variance du G_{ST} dépend en effet de n (nombre de populations) et de $\tilde{n}$ (nombre d'individus par population). La dérivée de la fonction $\mathrm{Var}(G_{ST}) = f(n, \tilde{n})$ s'annule quand la variance est minimale. L'effectif 'optimal' peut donc être déduit analytiquement. Les résultats indiquent que : a) pour un nombre de populations donné, la variance de G_{ST} baisse rapidement quand le nombre d'individus par population augmente, puis un plateau est très vite atteint, au delà duquel il ne sert plus à grand chose d'augmenter l'échantillonnage ; b) pour un effectif par population donné, augmenter le nombre de populations reste rentable très longtemps (figure 9) ; c) l'effectif optimal est généralement très faible (compris entre 2 et 30 individus haploïdes), d'autant plus que le G_{ST} est grand (figure 10).

Il s'agit là d'un résultat inattendu ; la majorité des études de diversité sont basées sur un nombre réduit de populations (une douzaine en moyenne ; Hamrick et coll. 1992) mais comportent beaucoup d'individus par population (près de 50 individus diploïdes dans notre compilation chez des arbres forestiers, soit 100 copies de chaque gène). Nos études sur les marqueurs cytoplasmiques (chapitres 2 et 3) se sont donc toutes basées sur un grand nombre de populations, avec par contre de petits effectifs par population. Nous avons vérifié à plusieurs reprises qu'augmenter l'échantillonnage intrapopulation ne changeait guère les conclusions (exemple : Petit et coll. 1993[7]).

Le message, peu intuitif, passe difficilement ; par exemple, Wu et coll. (1998) ont récemment analysé la

[6] Pons O et R J Petit. 1995. Estimation, variance and optimal sampling of gene diversity. I. Haploid locus. *Theor. Appl. Genet.* **90**, 462-470.

[7] Petit R J, D B Wagner et A Kremer. 1993. Ribosomal DNA and chloroplast DNA polymorphisms in a mixed stand of *Quercus robur* and *Q. petraea*. *Ann. Sci. For.* **50, Suppl. 1**, 41s-47s.

différenciation entre populations de trois pins californiens sur la base de 3, 4 et 6 populations par espèce, avec des effectifs de 25 à 40 individus par population. Vu la différenciation considérable observée (0,80 à 0,90), il s'agit là d'un échantillonnage particulièrement inadapté ; pourtant, l'étude de la différenciation était l'objectif annoncé de l'étude).

Etude empirique, allèles non-ordonnés

L'étude précédente avait été initiée suite à des essais de rééchantillonnages *bootstrap* (tirages avec remise), pour obtenir la précision des indices de diversité (notamment le G_{ST}). La question que je me posais alors était le type de rééchantillonnage à effectuer (populations, individus dans populations, ou double échantillonnage des populations et individus).

L'étude analytique (estimateurs *bootstrap*) et empirique (par tirages) a permis de démontrer que la méthode la plus appropriée est le tirage des populations.

Un tirage des populations et des individus dans les populations, qui semble 'mimer' la réalité de l'échantillonnage, surévalue en fait la variance (Petit et Pons 1998[8]). Les estimateurs *bootstrap* analytiques donnaient des résultats très proches des simulations *bootstrap*, démontrant la validité des approximations effectuées dans la méthode analytique.

Toutefois, même en ne pratiquant que le tirage des populations, la méthode *bootstrap* est biaisée quand le nombre d'individus par population devient petit. La méthode analytique lui est donc préférable. Des tirages sans remise dans une grande population ont permis en outre de retrouver empiriquement la taille optimale prédite analytiquement (Petit et Pons, non publié).

Prise en compte de la distance entre allèles

Les indices de diversité vus précédemment ne tiennent pas compte des similarités entre allèles. Dans le même cadre théorique que précédemment, des définitions et des estimateurs des paramètres de diversité ont été proposés qui utilisent l'information propre aux séquences ou aux haplotypes (Pons et Petit 1996[9]).

D'autres études statistiques sur ce sujet avaient été publiées auparavant (Nei 1982, Lynch et Crease 1990, Excoffier et coll. 1992). Les particularités de notre étude, outre l'application du modèle original décrit dans Pons et Petit (1995), est la comparaison de G_{ST} et N_{ST}. (N_{ST} est l'analogue du G_{ST} quand on prend en compte la distance entre allèles). Un test $N_{ST}>G_{ST}$ est proposé, qui permet de mesurer si la description des haplotypes permet de préciser la structuration (les haplotypes les plus similaires se retrouvent-ils préférentiellement dans les mêmes populations ?). La figure 11 illustre les différentes relations possibles.

Une autre possibilité consiste à regrouper les haplotypes de proche en proche (en '*remontant l'arbre*'), ce qui permet d'examiner comment se répartit la diversité en fonction de l'âge des lignées prises en compte (Pons et Petit 1996).

Plus récemment, j'ai développé un programme pour comparer N_{ST} et G_{ST} en permutant l'identité des haplotypes (Burban et coll. 1999[10]).

[8] Petit R J et O Pons. 1998. Bootstrap variance of diversity and differentiation estimators in a subdivided population. *Heredity* **80**, 56-61.

[9] Pons O et R J Petit 1996 Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* **144**, 1237-1245.

[10] Burban C, R J Petit, E Carcreff et H Jactel. Rangewide variation of the maritime pine bast scale *Matsucoccus feytaudi* Duc. (Homoptera:

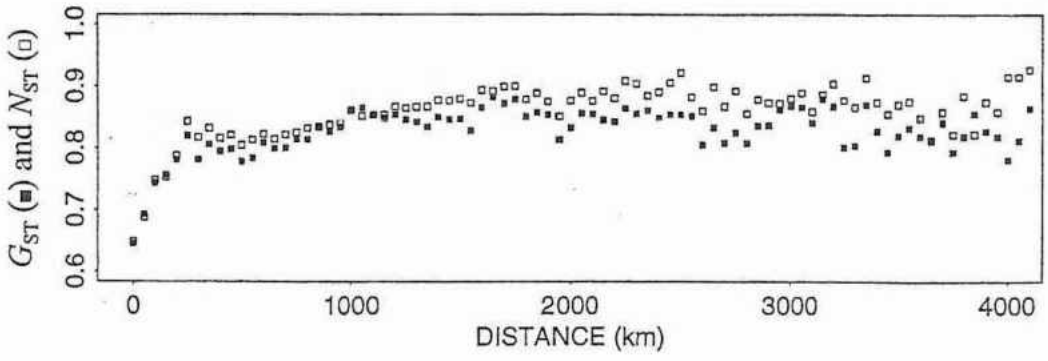


Figure 12 : Différenciation entre populations en fonction de leur éloignement géographique (exemple de la variabilité de l'ADN chloroplastique chez les chênes). Initialement, le G_{ST} et le N_{ST} sont égaux, puis le N_{ST} devient supérieur (effet phylogéographique).

La matrice de distances entre haplotypes, ainsi que leur fréquence, restent inchangées. Après 1000 permutations, la moyenne et la distribution des valeurs de N_{ST} sont obtenues ; comme prévu, la moyenne des N_{ST} 'permutés' est très proche du G_{ST} observé. Si le N_{ST} observé est plus grand que 95% des valeurs après permutations, on déclarera que N_{ST} est significativement plus grand que le G_{ST}. Il reste à évaluer en détail si les résultats obtenus sont semblables à ceux du test analytique.

L'interprétation qui peut être faite quand N_{ST} est plus grand que G_{ST} (structure phylogéographique, comme l'illustre la figure 11) reste incertaine, en l'absence d'étude de génétique des populations théorique (ou de simulations) comparant le comportement de ces deux paramètres. Même si on dispose d'estimateurs sans biais des coefficients de différentiation, il peut persister des 'artefacts' dans les cas extrêmes, aux faibles ou fortes diversités. J'ai décrit un tel artefact (passé apparemment inaperçu jusque là) en démontrant qu'aux faibles diversités G_{ST} était 'biaisé' vers le bas (Petit et coll. 1995), en particulier quand le nombre de populations échantillonnées est faible, et ceci quelque soit l'estimateur utilisé. On peut faire le parallèle avec l'étude de Cuguen et coll. (1988) sur le déficit d'hétérozygotes F_{IS}, également biaisé vers des valeurs négatives quand la diversité est faible (« *in a complete panmictic situation, we did not obtain an equal proportion of samples with a deficit or an excess of heterozygotes* »). Ces situations sont rarement reconnues (voir cependant Nagylaki 1998), et peuvent donc causer des problèmes d'interprétation. De plus, on ignore totalement si une structuration phylogéographique est compatible avec un équilibre entre migration et dérive ($N_{ST}>G_{ST}$), ou si une telle configuration prouve au contraire l'existence d'un déséquilibre. Intuitivement, la mutation crée instantanément une structuration phylogéographique, car les haplotypes nouvellement formés doivent être regroupés initialement avec l'allèle dont ils dérivent par mutation. Toutefois, si la mutation est négligeable devant la migration, tous les haplotypes observés sont relativement anciens et ont dû avoir le temps de se répartir indépendamment de leur phylogénie (pas d'association des allèles apparentés dans les mêmes populations).

Les coefficients de différenciation (G_{ST} ou N_{ST}) ne prennent pas en compte la proximité géographique des populations. Dans le cadre d'une large étude de la variabilité de l'ADN chloroplastique des chênes en Europe, la différenciation génétique a été estimée pour des classes de distance croissantes (Dumolin-Lapègue et coll. 1997a[11]).

La figure 12 montre un effet très net de l'isolement par distance pour les deux paramètres. La valeur moyenne du G_{ST} dans chaque classe est donnée par : $1-\frac{\overline{h}_s}{\overline{h}_T}$, et non par la moyenne des G_{ST} de la classe. Des permutations des populations permettent d'obtenir l'enveloppe correspondant à l'hypothèse nulle d'absence de structuration spatiale.

Etude de la richesse allélique et introduction de la technique de raréfaction en génétique

La richesse allélique est une mesure largement sous-utilisée ou mal employée en génétique des populations. J'ai montré empiriquement dans la première partie que

Matsucoccidae) in relation to the genetic structure of its host. *Mol. Ecol.* (soumis).

[11] Dumolin-Lapègue S, B Demesure, V Le Corre, S Fineschi et R J Petit. 1997. Phylogeographic structure of white oaks throughout the European continent. *Genetics* **146**, 1475-1487.

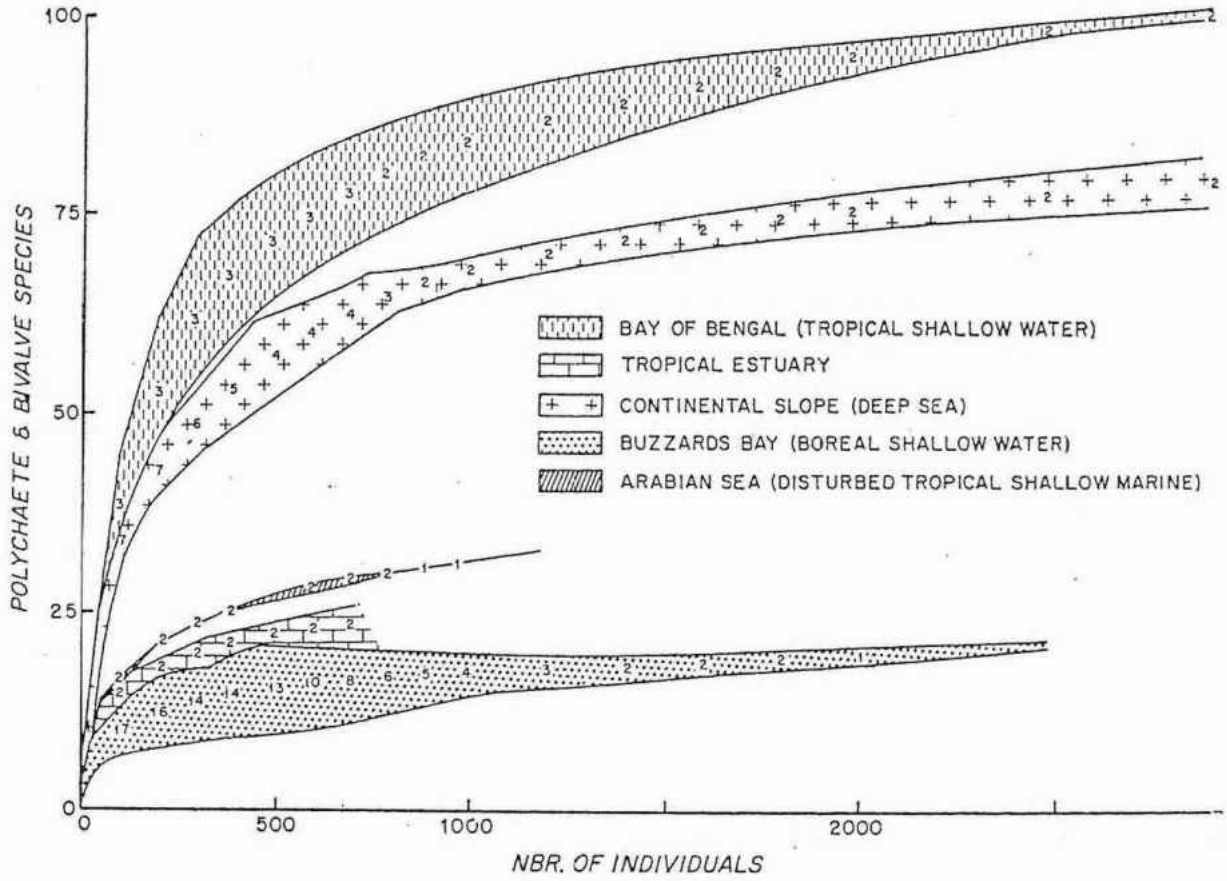


Figure 13 : Courbes de raréfaction pour différents écosystèmes marins. Le nombre d'espèces différentes de polychètes et de bivalves est donné en fonction du nombre d'individus examinés (Sanders 1968). C'est la première application de la méthode de raréfaction, mais l'estimation était encore approximative : elle sera corrigée par Hurlbert (1971).

sa corrélation avec l'hétérozygotie n'est pas totale ($r = 0,5$ en moyenne au niveau intraspécifique chez les arbres). Les travaux théoriques montrent que c'est une mesure hautement informative pour identifier certains processus : (1) goulots d'étranglement (notamment quand elle est confrontée à l'hétérozygotie, Nei et coll. 1975, Cornuet et Luikart 1996, Luikart et coll. 1998) ; (2) contact entre populations divergentes ('*admixture*') (Chakraborty et coll. 1988) ; (3) flux de gènes (Wakeley 1998, *cf.* ci-dessus) ; (4) estimation de θ (Tajima 1983). En outre, c'est une mesure de diversité pertinente (plus que H) dans le domaine de la conservation des ressources génétiques (Marshall et Brown 1975, Kremer 1994, Bataillon et coll. 1996, etc.).

Le seul défaut qui lui est reproché est sa dépendance vis-à-vis de la taille de l'échantillon (*cf.* par exemple Brown et Weir 1983, ou plus récemment Ward et coll. 1992). Pourtant, dès le début des années 1970, les écologistes proposaient une solution pour remédier à ce problème. En effet, Hurlbert (1971), affinant une technique proposée par Sanders (1968 ; figure 13), montre que l'espérance du nombre d'espèces est donné par :

$$E(S_n) = \sum_i \left[1 - \frac{\binom{N - N_i}{n}}{\binom{N}{n}} \right], \qquad (1)$$

avec n le nombre d'individus pris au hasard parmi une collection de N individus ($n<N$), et N_i le nombre d'individus composant l'espèce i. Bien que le travail de Hurlbert ait été immédiatement repris et devienne un classique de la littérature écologique (Simberloff 1972, Peet 1974, Heck et coll. 1975, etc.), il faut attendre 1988 pour que Chakraborty et coll. proposent une procédure d'ajustement de la richesse allélique à la taille de l'échantillon, dans un appendice d'un article sur l'effet de l'agglomération de différentes populations humaines sur la variabilité génétique. La méthode sera reprise dans un autre papier de la même équipe (Neel et coll. 1988) examinant le nombre de variants protéiques parmi un très large échantillon d'habitants des villes d'Hiroshima et Nagasaki au Japon. Le nombre attendu d'allèles, selon ces auteurs, est donné par :

$$E(k_n) = \sum_i \left[1 - (1 - p_i)^n \right] \approx \sum_i \left[1 - \exp\{-np_i\} \right].$$

Ces deux formules sont moins précises que la solution donnée par Hurlbert, notamment pour des petits échantillons. Cependant, manipuler des factorielles est rapidement problématique, même avec les puissances de calcul des ordinateurs disponibles actuellement. Heureusement, j'ai pu montrer que l'équation (1) se simplifie comme suit (Comps et coll. 1999[12]) :

$$E(k_n) = \sum_i \left[1 - \prod_{j=0}^{n-1} \frac{N - N_i - j}{N - j} \right].$$

La première application de la formule de Hurlbert en génétique a consisté à comparer, chez un arbre forestier marocain, l'arganier (*cf.* chapitre 4), la richesse allélique de 10 populations évaluée à l'aide de neuf locus enzymatiques polymorphes (El Mousadik et Petit 1996[13]).

En effet, les tailles d'échantillons étaient fort hétérogènes : ils variaient de 20 à 50 individus diploïdes par population. La technique de raréfaction a été utilisée pour

[12] Comps B, D Gömöry, J Letouzey, B Thiébaut et R J Petit. Imprint of postglacial colonization on the distribution of heterozygosity and allelic richness in the European beech. Soumis à *Genetics*.

[13] El Mousadik A et R J Petit 1996. High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics*, **92**, 832-839.

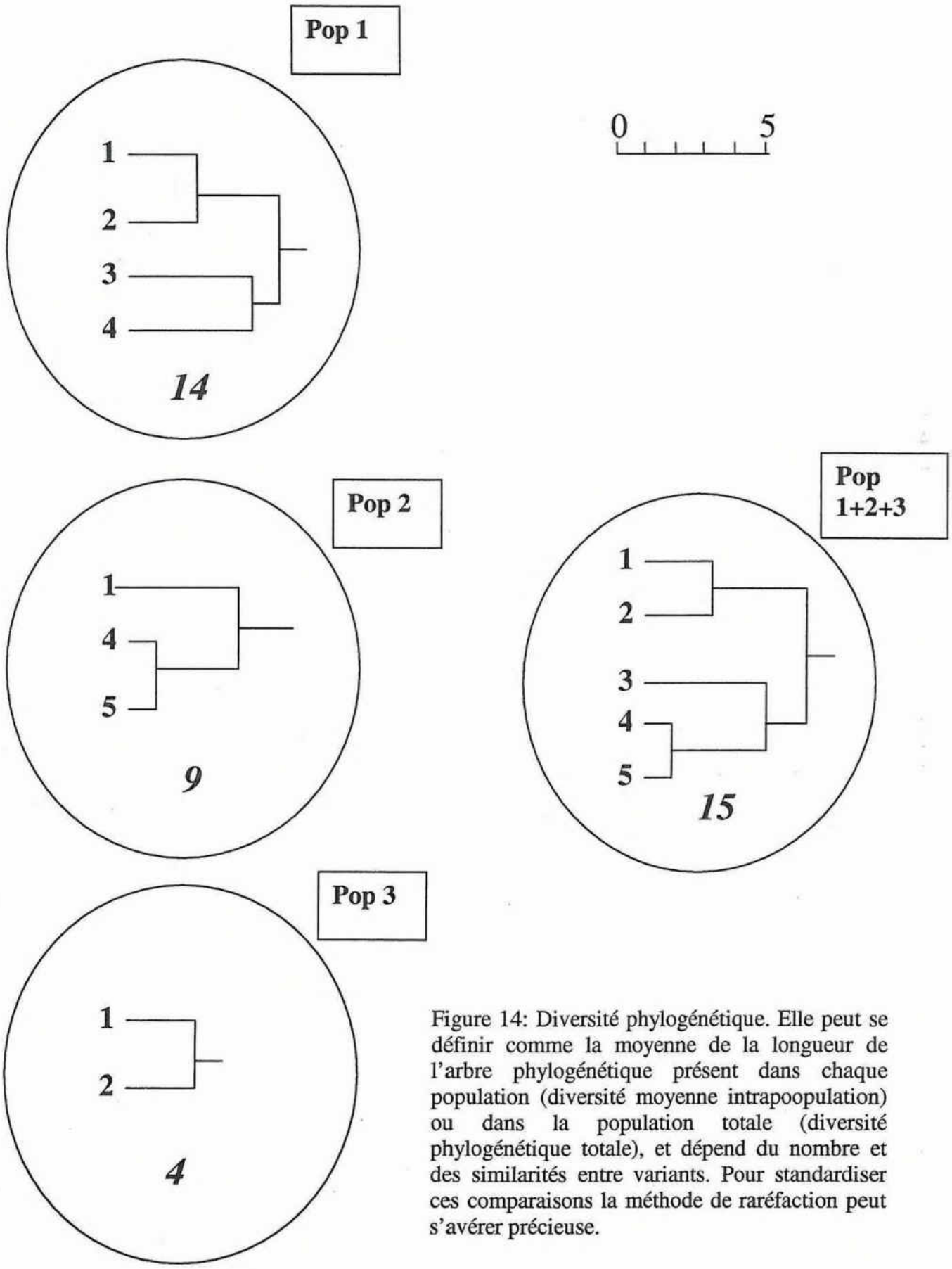


Figure 14: Diversité phylogénétique. Elle peut se définir comme la moyenne de la longueur de l'arbre phylogénétique présent dans chaque population (diversité moyenne intrapoopulation) ou dans la population totale (diversité phylogénétique totale), et dépend du nombre et des similarités entre variants. Pour standardiser ces comparaisons la méthode de raréfaction peut s'avérer précieuse.

estimer la richesse allélique pour des effectifs standardisés à 40 et à 10 gènes par locus. Le rang des populations, classées selon leur richesse allélique, a été modifié par ces standardisations, montrant l'intérêt de ces corrections. La richesse allélique standardisée à 10 est beaucoup plus corrélée à H que celle standardisée à 40 : 0,95 contre 0,70.

J'ai introduit de plus le coefficient de différenciation pour la richesse allélique, en notant qu'il est possible de calculer la richesse allélique sur l'ensemble des populations, et de le contraster avec la richesse allélique moyenne par population, en standardisant à un effectif commun (10 ou 40 en l'occurrence). Si on note A_S la richesse allélique moyenne intrapopulation, et A_T la richesse allélique totale, le coefficient de différentiation s'écrit :

$$A_{ST}[n] = 1 - \frac{A_S[n] - 1}{A_T[n] - 1}.$$

Les résultats donnent, pour l'arganier : $A_{ST}[10] = 0{,}38$, $A_{ST}[40] = 0{,}52$, comparé à $F_{ST} = A_{ST}[2] = 0{,}25$. On constate donc que le coefficient de différentiation dépend très largement de la diversité considérée. Ceci indique que les allèles rares sont répartis de manière plus hétérogène que les allèles plus fréquents chez cette espèce. Depuis, cette tendance a été retrouvée systématiquement, pour d'autres espèces et d'autres marqueurs : la part de la diversité (d'ordre n) distribuée entre les populations augmente avec n.

La méthode a été appliquée depuis à trois reprises (Petit et coll. 1998[14], Vendramin et coll. 1999[15], et Comps et coll. 1999).

Dans la deuxième étude, les données concernaient des microsatellites chloroplastiques chez le sapin (*Abies alba*). Il y avait deux locus avec respectivement 8 et 18 variants de longueur, se combinant en 90 haplotypes. La richesse allélique A était plus stable que H d'un marqueur à l'autre (corrélation de rang de plus de 0,5 entre la richesse allélique évaluée avec le locus 1 et celle évaluée avec le locus 2). Dans la troisième étude, la plus complète, l'emploi de la technique de raréfaction a été systématique, car l'objectif consistait précisément à décrire la variabilité géographique de la richesse allélique. Des courbes complètes de raréfaction ont été obtenues, afin d'étudier à partir de quelle valeur la diversité d'ordre n devient qualitativement différente de celle d'ordre 2, en ce qui concerne la répartition de la diversité.

En conclusion, on peut remarquer qu'il est désormais possible d'analyser des données en tenant compte des similarités des variants ; il est aussi possible de considérer autre chose que des paires (étude de la richesse allélique). Une perspective serait désormais de combiner ces deux méthodes. L'idée consisterait à calculer la longueur moyenne de l'arbre (**longueur totale de la généalogie**) pour un échantillon de n séquences ou haplotypes et de contraster cette mesure au niveau population et espèce (figure 14). Selon moi, ce serait la meilleure manière de valoriser l'information (prise en compte des similarités des séquences et de la richesse allélique). La théorie de la coalescence devrait permettre d'étudier les propriétés d'une telle mesure dans des modèles de génétique des populations variés.

[14] Petit R J, A El Mousadik et O Pons. 1998. Identifying populations for conservation on the basis of genetic markers. *Cons. Biol.* **12**, 844-855.

[15] Vendramin G G, B. Degen, R J Petit, M Anzidei, A Madaghiele et B Ziegenhagen. 1999. High level of variation at Abies alba microsatellite loci in Europe. *Mol. Ecol.* In press.

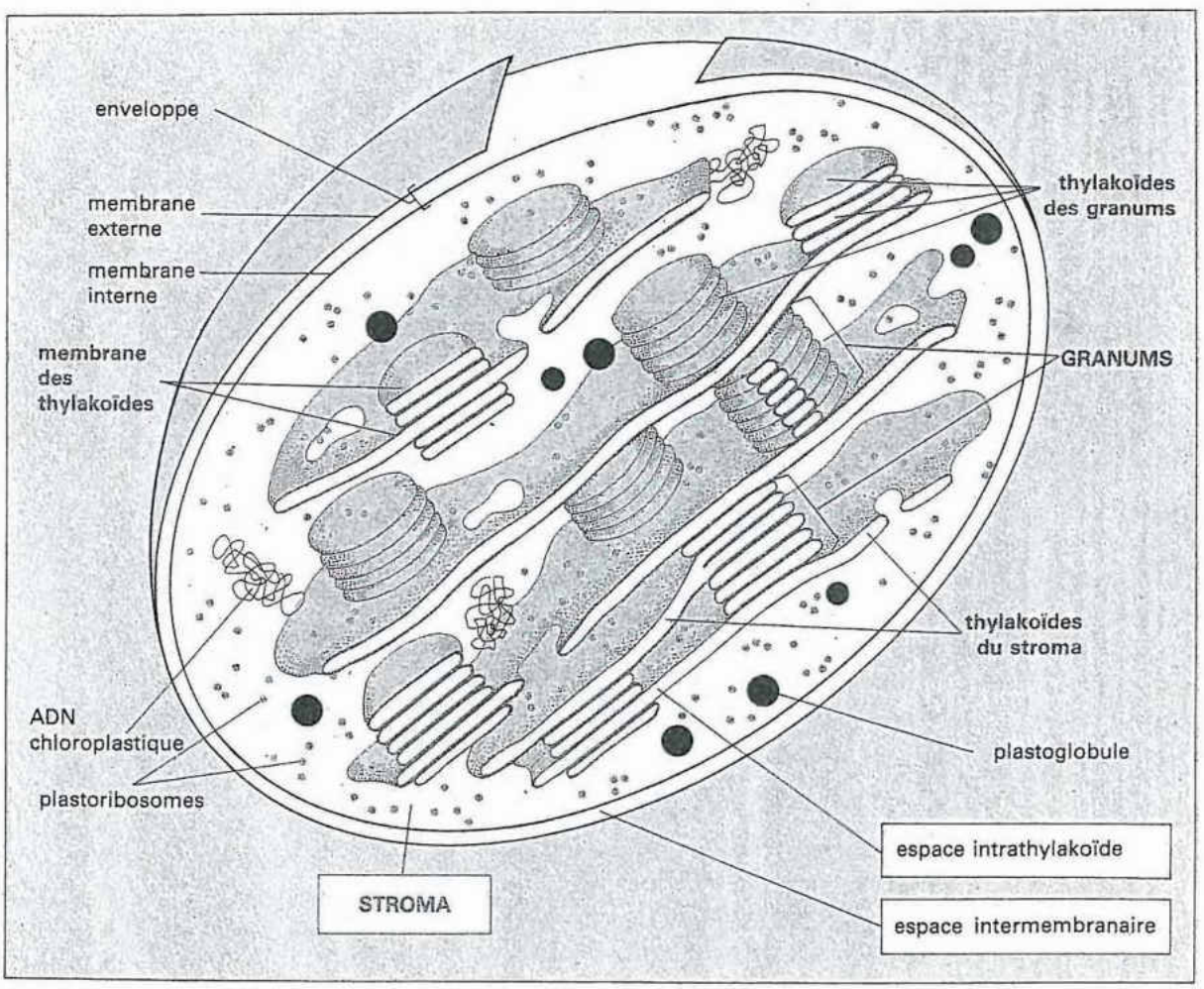


Figure 1 : Structure d'un chloroplaste à granums (d'après Berkaloff et coll. 1981). L'ADN des chloroplastes (ADNcp) est une molécule circulaire d'environ 50 microns (100 millions de daltons). Il existe par cellule plusieurs copies d'ADNcp ; en effet chaque chloroplaste possède 10 à 30 molécules identiques d'ADNcp, par groupes de 3 à 6, formant les nucléoïdes visibles dans le stroma. Dans les cellules de feuilles, qui comptent entre 30 et 500 chloroplastes, on a donc 300 à 15 000 copies de ce génome (contre seulement deux pour le génome nucléaire).

Chapitre 2

Structuration de la diversité génétique de l'ADN chloroplastique et mitochondrial chez les plantes

« It is particularly fascinating that three different genomes, embedded in and serving the same organism and interacting with each other in numerous ways, can evolve differently ».

C. William Birky, Jr. 1988. Evolution and variation in plant chloroplast and mitochondrial genomes. In: Plant Evolutionary Biology, Gottlieb & Jain éditeurs. Chapman & Hall, Londres.

INTRODUCTION

Mon travail de thèse avait porté sur la variabilité de l'ADN chloroplastique chez les chênes et la mise au point de modèles théoriques de génétique des populations appliqués aux génomes cytoplasmiques des plantes (1989-début 1992). Depuis, avec l'aide de plusieurs collègues et étudiants, j'ai continué à perfectionner les méthodes d'analyse des données (*cf.* chapitre précédent) et les outils moléculaires, et à les appliquer à l'étude des génomes cytoplasmiques des arbres forestiers. Si les travaux sur la variabilité de l'ADN chloroplastique et mitochondrial chez les plantes sont en pleine croissance, c'est en partie grâce à ces efforts.

Dans ce chapitre, je tente en premier lieu de replacer les travaux effectués dans le contexte historique du développement de la génétique des génomes cytoplasmiques chez les plantes. La présentation (chronologique) est volontairement orientée vers les aspects touchant le plus à la biologie des populations, sans ignorer toutefois le développement des techniques moléculaires qui ont permis ces études. Je résume ensuite brièvement tous les travaux dont je me suis occupé et qui ont trait à l'étude de la variabilité des génomes cytoplasmiques. Certains résultats seront plus longuement analysés dans le chapitre suivant. Je présente enfin une première version d'une analyse bibliographique sur la variabilité des génomes cytoplasmiques chez les plantes. Le jeu de données comprend 75 espèces pour lesquelles la diversité chloroplastique et/ou mitochondriale intra- et interpopulation a pu être mesurée. Une comparaison avec les marqueurs nucléaires est également effectuée.

HISTORIQUE DE L'ETUDE DE LA VARIABILITE DES GENOMES CYTOPLASMIQUES CHEZ LES PLANTES

Hérédité des génomes cytoplasmiques des plantes

Chez les plantes, l'intérêt de l'étude des génomes cytoplasmiques : ADN chloroplastique (ADNcp) (figure 1) et ADN

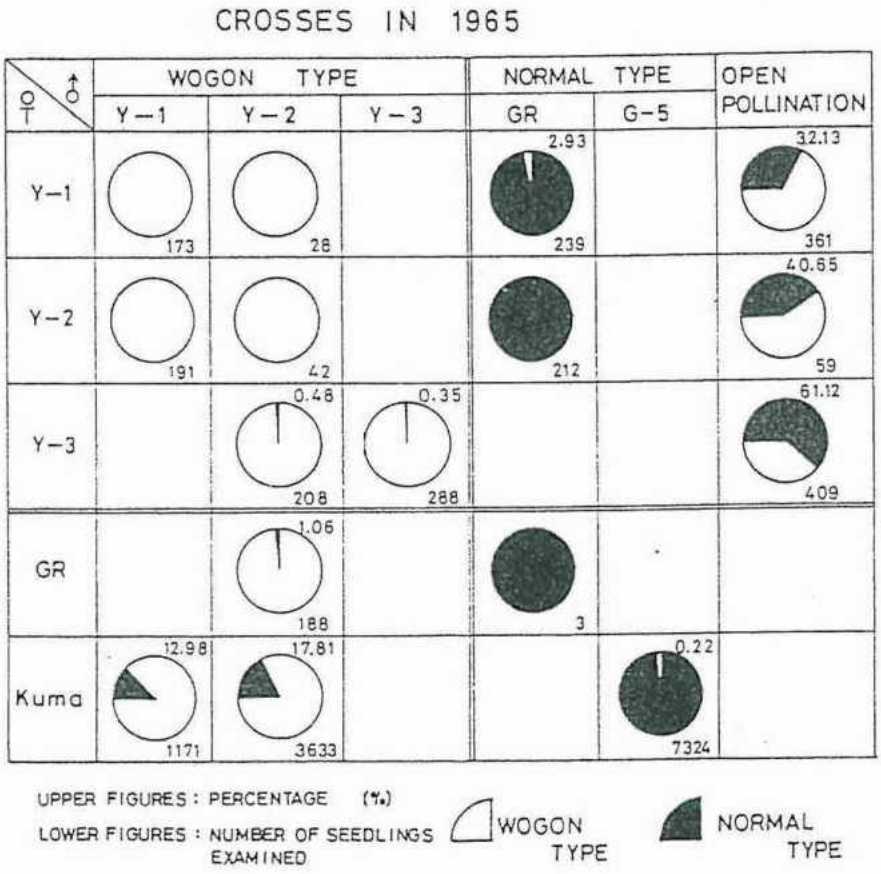


Figure 2 : Résultats des croisements contrôlés chez le conifère Cryptomeria japonica (Ohba et coll. 1971). Le type 'Wogon' a des aiguilles primaires blanches, et non vertes. Ces résultats indiquent que les rtypes Wogon et normaux sont le plus souvent transmis paternellement à leur descendance.

mitochondrial (ADNmt), tient largement à leur mode d'hérédité non mendélienne, souvent maternelle, voire biparentale, mis en évidence indépendamment dès 1909 (donc seulement neuf ans après la redécouverte des lois de Mendel), par Correns et Baur. Correns (un des redécouvreurs de lois de Mendel en 1900), étudiait une variété ornementale panachée de *Mirabilis japala*. Les graines issues des fleurs situées sur une partie albinos de la plante produisaient uniquement des plants albinos, quelque soit l'origine du pollen utilisé pour les féconder ; par contre le pollen de ces mêmes fleurs ne transmettait pas ce caractère à des fleurs normales. Baur, à la même époque (1909), décrivait le premier cas d'hérédité biparentale des plastes, chez *Pelargonium zonale*.

L'année précédente, Correns (1908) avait publié une étude sur la transmission du sexe chez des espèces gynodioïques (telle *Cirsium oleraceum*). Les femelles, pollinisées par des hermaphrodites, ne produisent que des femelles. Là aussi, un caractère cytoplasmique pouvait être en cause (on démontrera plus tard que c'est l'ADN mitochondrial qui est impliqué). Toutefois, si les femelles sont hétérozygotes Aa, et si l'allèle A féminisant est transmis préférentiellement à l'ovule lors de la méiose femelle, il n'est pas nécessaire d'imaginer un déterminant génétique cytoplasmique (East 1910). L'étude de maïs mâle-stériles par Rhoades en 1933 a définitivement démontré la participation de déterminants cytoplasmiques pour ce caractère, en démontrant l'absence de distorsions de ségrégation sur les 10 chromosomes de l'espèce dans la lignée femelle, confirmant ainsi l'hypothèse de Correns.

Un intérêt particulier pour ces phénomènes non conventionnels d'hérédité cytoplasmique (chez les animaux tout comme chez les plantes ou chez les microorganismes) est patent notamment en Europe, souvent en opposition à la position dominante de la génétique mendélienne (Sapp 1987). Les détracteurs anglo-saxons de ces études sur les déterminants héréditaires cytoplasmiques opposent ainsi une génétique nucléaire (*nuclear*) et une génétique '*unclear*' (Picard 1998).

A la suite des premiers travaux de Baur et Correns, beaucoup d'autres cas d'hérédité maternelle de déficiences chlorophylliennes furent mis en évidence chez les plantes (revue dans Kirk et Tilney-Bassett 1978). En 1971, Ohba et coll. démontrent avec cette méthode la transmission paternelle des plastes d'un conifère, *Cryptomeria japonica* (figure 2).

Compatibilité nucléo-cytoplasmique

Outre les études d'hérédité, de nombreux travaux furent effectués sur la génétique cytoplasmique des plantes. En particulier, à partir des années 20 et pendant plus de 50 ans, l'analyse de combinaisons différentes de cytoplasmes et de noyaux, obtenus par *backcross*, a été un sujet d'étude très prisé par de nombreux généticiens. Les travaux sur *Oenothera* furent particulièrement nombreux, notamment en Allemagne (travaux de Otto Renner et de son équipe, cités en détail dans Grun 1976). Dans le cas de *Oenothera*, la nature chloroplastique des caractéristiques observées (panachage des feuilles) semble démontrée par deux arguments différents. Tout d'abord, certaines cellules contiennent à la fois des chloroplastes normaux et d'autres anormaux, visibles au microscope. De plus, la vitesse de ségrégation (dérive intra- et interindividu, ou '*sorting-out*') des caractères semble trop rapide pour que les mitochondries puissent être impliquées (il y a en effet moins de plastes que de

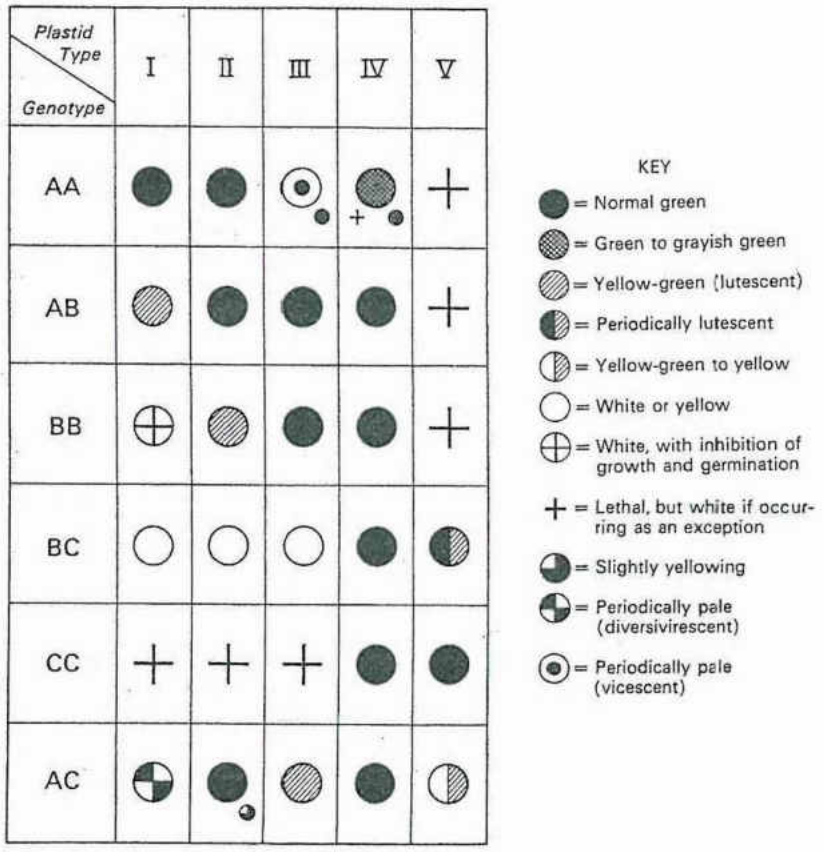


Figure 3 : Relation de compatibilité entre différents génomes nucléaires (AA, AB, etc.) et les types de plastes (I-V) trouvés chez 14 espèces d'*Oenothera* (d'après Stubbe 1964, in Grun 1976).

mitochondries par cellule). Cette vitesse de ségrégation fut étudiée mathématiquement par Correns dès 1922.

Les travaux entrepris à la suite de Renner par Stubbe notamment démontrèrent qu'il y avait un certain nombre de types différents de plastes dans le genre *Oenothera* (notés I-V, voir figure 3). D'autres travaux ont démontré des anormalités morphologiques ou physiologiques quand le cytoplasme d'un individu était associé au génome nucléaire d'un autre individu de la même espèce mais distant géographiquement, démontrant ainsi l'existence de polymorphismes cytoplasmiques intraspécifiques. Les études sur *Epilobium hirsutum* de Michaelis et de son équipe, durant les années 1935-1970 en Allemagne, furent particulièrement détaillées. Elles démontrèrent l'existence de nombreux 'plasmons' différents dans cette espèce, associés à chaque fois à des génomes nucléaires compatibles. Une partie des combinaisons testées entre plantes éloignées géographiquement donnait des aberrations (réduction des feuilles et tiges, gaufrage des feuilles, et stérilité mâle). Toutefois des combinaisons très éloignées pouvaient être compatibles. Vu le nombre de populations testées chez cette espèce présente en Europe, Asie et Afrique, on peut considérer qu'il s'agissait là de la première étude d'envergure de variabilité cytoplasmique chez une plante. La possibilité que les polymorphismes chloroplastiques ou mitochondriaux d'une même espèce ne soient pas neutres n'est plus guère évoquée dans les études actuelles, sans doute à tort dans certains cas.

Cytologie et hérédité cytoplasmique

Dès 1903, Coker montrait que le cytoplasme du gamète mâle participe à la constitution de l'embryon de *Taxodium distichum* (famille des Taxodiacées). C'est grâce à la présence d'amidon, facile à suivre en microscopie photonique, que les auteurs anciens purent suivre le devenir du cytoplasme mâle lors de la fécondation des plantes. Les années 60 et 70 sont marquées par l'essor des techniques d'étude cytologique, notamment la microscopie électronique, qui permettent de mettre en évidence la diversité des mécanismes assurant l'hérédité uniparentale. En France, ces travaux se sont particulièrement développés, et l'étude des gymnospermes notamment a permis de prédire sans aucune étude génétique l'hérédité *paternelle* : (a) des chloroplastes pour toutes les Conifères sauf *Ephedra* et (b) des mitochondries chez les Cupressacées et les Taxodiacées uniquement (Camefort 1966 1968 ; Chesnoy 1969 1987 ; Chesnoy et Thomas 1971).

Le début des années 70 : les bases de la biologie moléculaire des génomes cytoplasmiques

Les propriétés physico-chimiques de l'ADN chloroplastique des plantes furent étudiées à partir du début des années 1970. Le pourcentage des bases G et C est très semblable d'une espèce à l'autre (Kirk 1971). La structure circulaire de ce génome apparut ensuite au microscope électronique (45 µM de long, taille très conservée également, Kolodner et Tewari 1972). Une technique plus précise pour étudier la variabilité génétique du génome chloroplastique est basée sur l'étude des protéines codées par ce génome. Après isolement des chloroplastes, apparaît lors de la purification une protéine particulièrement abondante, appelée initialement *Fraction I protein*, catalysant la fixation du CO_2 par les plantes et constituée d'une petite et d'une grande sous-unité (Wildman et coll. 1971) (il s'agit de la ribulose bisphosphate carboxylase ou rubisco). Les premières

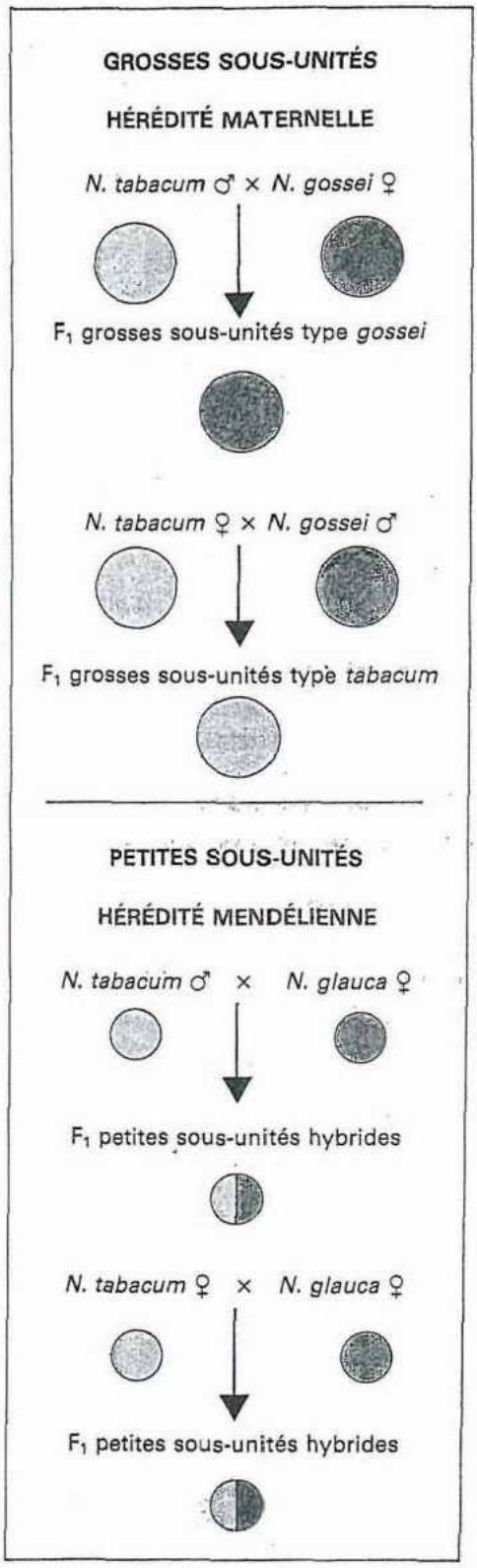


Figure 4 : Hérédité maternelle et hérédité mendélienne des sous-unités de la ribulose 1,5-diphosphate carboxylase chez le tabac. Des croisements interspécifiques réciproques ont été réalisés entre espèces de tabac américains (*Nicotiana tabacum* et *N. glauca*) et australien (*N. gossei*). Les grosses sous-unités sont transmises maternellement (gène chloroplastique), les petites ont une hérédité mendélienne (gène nucléaire).

analyses de cette protéine par hydrolyse à la trypsine et chromatographie sur papier permirent de mettre en évidence des différences entre espèces de tabac au niveau de la petite sous-unité ; ces différences étaient transmises selon les lois de Mendel, démontrant que cette sous unité était codée par des gènes nucléaires (Kawashima et coll. 1971). Par contre, une différence au niveau de la grande sous-unité, distinguant des espèces du genre *Nicotiana* originaires de l'hémisphère nord et de l'hémisphère sud, était transmise maternellement (Chan et Wildman 1972) (figure 4). Chan et Wildman concluent leur article ainsi : « *Thus it is now clear that information from both nuclear and chloroplast DNA is required for ordering the sequence of amino acids in Fraction I protein. The present investigation also reveals the probability that only one mutation of the cistron in chloroplast DNA coding for the more than 200 amino acids in the large subunit of Fraction I protein has survived during 150 million years of evolution separating the Australian from the Western Hemisphere* Nicotiana *species* ». Dès cette époque donc, la coopération des deux génomes (plus de 90% des protéines du chloroplaste sont en fait codées par le génome nucléaire, une petite partie seulement étant codée par le génome du chloroplaste ; Sugiura 1992) et la très faible vitesse d'évolution du génome chloroplastique apparaissaient clairement.

1975-1980 : l'essor de la biologie moléculaire des génomes cytoplasmiques

La première endonucléase de restriction fut caractérisée en 1972 ; il s'agit de l'enzyme *Eco*RI, issue du colibacille *Escherichia coli*, qui reconnaît et coupe une séquence spécifique de six bases nucléotidiques (Hedgpeth et coll. 1972). D'autres endonucléases seront découvertes rapidement. Elles seront appliquées à l'étude d'ADN viral (Landy et coll. 1974) et plasmidial (Thompson et coll. 1974), puis à l'analyse de l'ADN mitochondrial de mammifères (Potter et coll. 1975).

En 1976, vient le tour de l'ADN chloroplastique (Atchison et coll. 1976 ; Vedel et coll. 1976) et de l'ADN mitochondrial végétal (Levings and Pring 1976). Bedbrook et Bogorad réalisent, également en 1976, la première carte physique de l'ADN chloroplastique chez le maïs. D'emblée, l'intérêt de la comparaison des patrons de restriction pour mesurer la similarité des espèces apparaît ; là où les mesures physico-chimiques indiquaient une grande homogénéité, de nombreuses différences apparaissent. A noter aussi dès ces premières études un premier cas possible d'introgression de génome chloroplastique : « Nicotiana *species, especially those of similar geographical distribution, can be readily crossed and appear to possess identical chloroplast DNAs* » (Atchison et coll. 1976). Les premiers cas de polymorphismes intraspécifiques d'un génome cytoplasmique chez une plante concernent l'ADN mitochondrial, lors de la comparaison de lignées normales et mâles-stériles de maïs (Levings et Pring 1976) ou de blé (Quetier et Vedel 1977). Cela signifie-t-il que les mutations à l'origine de la stérilité cytoplasmique mâle (CMS) -si importante pour les programmes d'amélioration génétique- ont été identifiées ? Des cas similaires d'association avec la CMS, mais concernant cette fois l'ADN chloroplastique, jettent rapidement le doute (Frankel et coll. 1979). Dans la première véritable étude de variabilité cytoplasmique chez des populations naturelles végétales, est décrit un cas de polymorphisme intraspécifique « persistant » de l'ADN chloroplastique, chez *Nicotiana debneyi* (Scowcroft 1979). Il s'agit d'un site

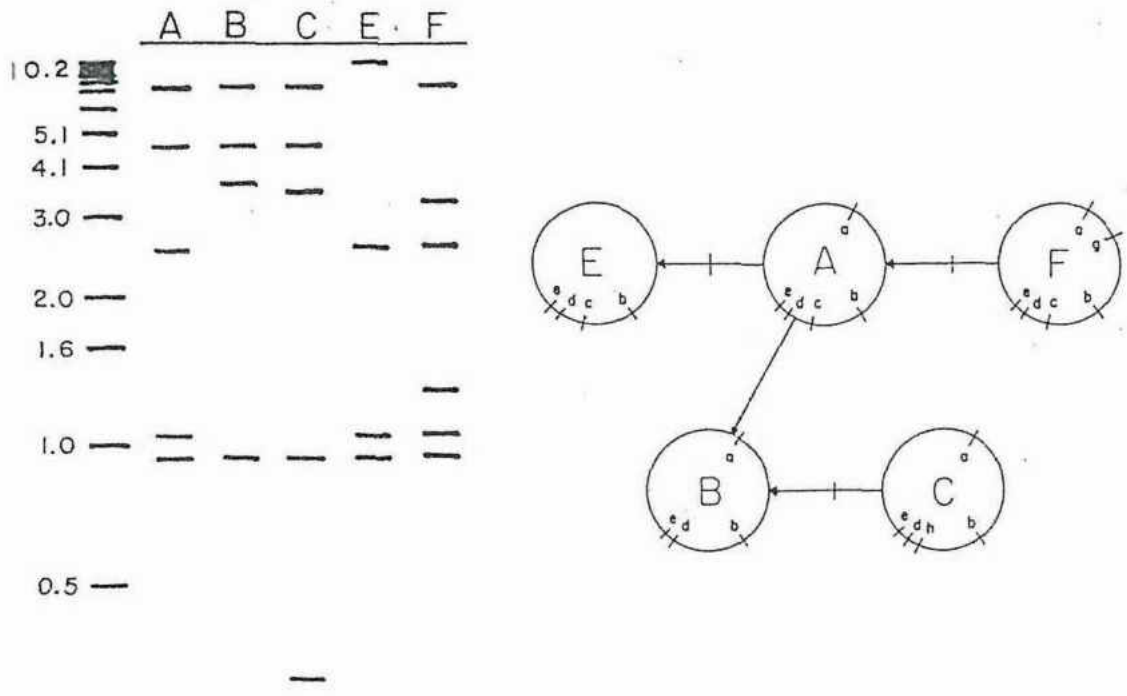


Figure 5 : Diagramme de restriction par l'enzyme *Bst*EII de l'ADN mitochondrial de différents individus du rongeur *Peromyscus maniculatus*, et cartes de restriction obtenues par la procédure de double digestion, correspondant aux profils de gauche. Un réseau obtenu par parcimonie indique les relations les plus probables entre ces variants d'ADN mitochondrial. Les flèches correspondent à la perte d'un site (et non à la direction de l'évolution). D'après Avise et coll. (1979, 1983).

de restriction *Eco*RI présent parmi tous les individus de sept des neuf populations étudiées, et totalement absent des deux autres. Scowcroft souligne l'homogénéité intrapopulation pour ce polymorphisme dont il démontre l'hérédité maternelle. Il fait également le raisonnement suivant : « *These 53 RI sites represent the potential for change at only 318 nucleotide pairs, which represents only 0.2% of the entire cp-DNA genome. In the* N. debneyi *polymorphism therefore, the deletion of a pre-existing EcoRI site could reflect an average of 500 nucleotide substitutions which distinguish the chloroplast of D2 from that of D1 types* ». Autrement dit, il y aurait une riche réserve de variabilité intraspécifique disponible chez cette espèce. Une remarque particulièrement en avance sur son temps, puisque de nombreux chercheurs considèreront, jusqu'à la fin des années 1980, voire au delà, que l'ADN chloroplastique ne contient pas suffisamment de variabilité utile pour les études de microévolution (Clegg 1987 ; Palmer 1987; Sandbrink et coll. 1990).

Scowcroft juge utile de préciser que les populations de tabac caractérisées par les deux variants chloroplastiques sont interfertiles, ainsi que leurs hybrides. Donc des différences de séquence nucléotidiques existent dans le génome des organelles, sans qu'y soit associé de stérilité mâle : « *Consequently, previous studies which imply a causal relationship between organelle DNA differences and male sterility must be treated with circumspection* ». Toutefois, la tentation d'attribuer une fonction à ces variations de séquence reste forte : « *No function can be ascribed to the polymorphism but several morphological characters distinguished the D1 and D2 populations. D2 populations had darker green leaves* ...». En 1979 également, Timothy et coll. rapportent de nombreux polymorphismes intraspécifiques concernant à la fois le génome mitochondrial et le génome chloroplastique chez le teosinte. Ils montrent que l'ADN chloroplastique peut varier indépendamment de l'ADN mitochondrial, et *vice versa*. Néanmoins les haplotypes identifiés correspondent assez bien aux groupes identifiés par des méthodes plus traditionnelles, sur la base de caractères codés par le génome nucléaire. Les auteurs concluent, ici aussi, qu'une évolution parallèle –et harmonieuse– des trois génomes a eu lieu, ce dont ils ne s'étonnent pas, sachant que ces génomes coopèrent nécessairement à l'élaboration des organelles et des cellules...

Une conséquence de l'hérédité uniparentale des deux génomes cytoplasmiques est un mode d'évolution clonal, propice aux reconstitutions phylogénétiques. La première étude phylogénétique basée sur le polymorphisme de restriction de l'ADN chloroplastique et mitochondrial fut réalisée en France et concernait les blés et les espèces apparentées (Vedel et coll. 1978). En 1979, Avise et coll. démontraient, en étudiant l'ADN mitochondrial d'un petit rongeur d'Amérique du Nord, que l'utilité phylogénétique des marqueurs cytoplasmiques était transposable également au niveau intraspécifique (figure 5). On verra plus loin qu'il faudra attendre longtemps avant que de tels travaux voient le jour chez les plantes.

Les années 1980

Retard des études de diversité

Malgré ces perspectives et les premières mises en évidence de polymorphisme intraspécifique, l'ADN chloroplastique a été néanmoins perçu comme presque invariant au sein des

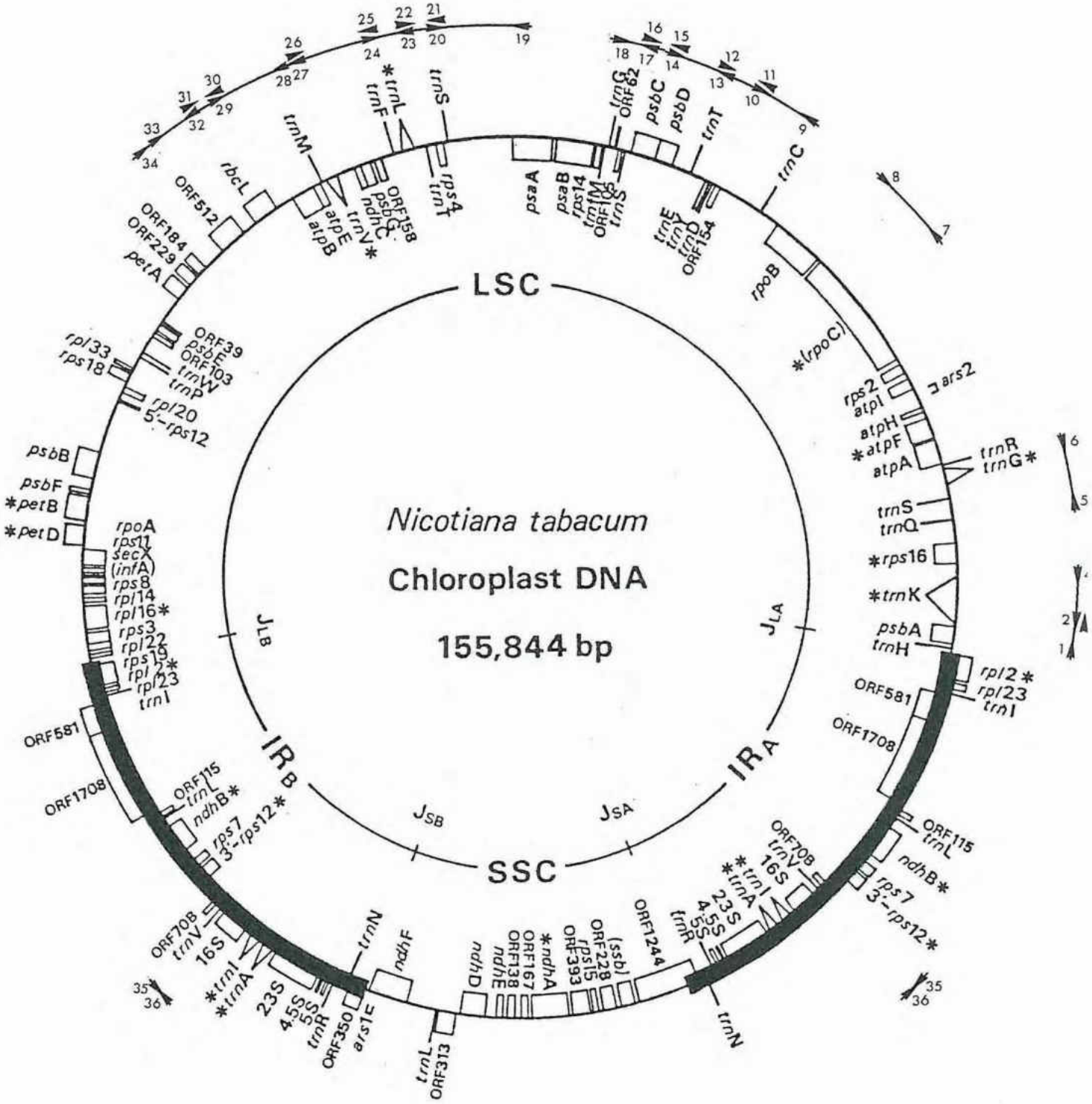


Figure 6 : Séquence complète du tabac (*Nicotiana tabacum*) obtenue en 1986 par Shinozali et coll. IR : *Inverted Repeat* (répétition inversée). LSC, SSC : L*arge & Small Single Copy region*. Les positions de quelques amorces universelles mises au point par Taberlet et coll. (1991), Demesure et coll. (1995) et Dumolin-Lapègue et coll. (1997) sont indiquées.

espèces. Comme le font remarquer Soltis et coll. en 1992, cette vision arrangeait pas mal de monde : « *Systematists have long sought methodologies that would facilitate phylogenetic reconstruction based on at most a few representative collections per species* » (pour ne pas dire un seul exemplaire !).

Chez les plantes, ce sont donc les études de phylogénie interspécifique qui vont se développer au cours des années 1980, aux dépens des études de diversité intraspécifique. Celles-ci se multiplièrent par contre chez les animaux, grâce au taux de substitution beaucoup plus élevé de l'ADN mitochondrial chez ces organismes. Dans une revue sur l'évolution de l'ADN chloroplastique et ses applications en biosystématique publiée en 1987, Palmer ne consacre qu'un court paragraphe sur la variabilité intraspécifique de l'ADN chloroplastique. La seule étude détaillée de génétique des populations qui est citée est celle de Banks et Birky (1985). Ces auteurs ont étudié la variabilité de l'ADN chloroplastique de 100 lupins répartis dans 21 populations, à l'aide de sept enzymes de restriction. Ils n'ont détecté que trois variants, dont un très fréquent (présent dans 88% des individus). Or cette espèce est caractérisée par une diversité nucléaire (étudiée à l'aide des allozymes) particulièrement élevée. C'est ce contraste qui frappera beaucoup de chercheurs et les découragera de se lancer dans des études de diversité intraspécifique approfondies. Pourtant, une des populations étudiée par Banks et Birky était totalement fixée pour un variant rare ; malgré la faible diversité, la structuration était donc beaucoup plus forte pour l'ADN chloroplastique que pour les allozymes. Toutefois, les données étaient jugées insuffisantes pour estimer la subdivision de la diversité (un rapide calcul me donne : $G_{ST} = 0,86$, $N_{ST} = 0,96$, mais la précision est effectivement faible). Comme dans l'étude de Scowcroft (1979), les potentialités de ces marqueurs maternels pour étudier la variabilité géographique auraient dû déjà frapper les chercheurs, mais le message était masqué par la faible variation.

Les potentialités de l'ADN chloroplastique pour étudier les changements démographiques furent mises en évidence par Clegg et coll. (1984). Ces auteurs ont comparé la diversité d'orges cultivés et de leur ancêtre sauvage, mettant en évidence l'absence de diversité chloroplastique des cultivars. Ce goulot d'étranglement (*bottleneck*) n'était pas apparu lors d'études utilisant les isozymes. En Amérique du sud, des études de variabilité de l'ADN chloroplastique chez des ancêtres de la tomate cultivée mettent en évidence des polymorphismes communs à plusieurs espèces, dans le complexe *Lycopersicon peruvianum* (Palmer et Zamir 1982). Ces échanges n'avaient pas non plus été repérés avec d'autres outils.

Des percées dans la description moléculaire de ces génomes et dans l'étude de leur évolution

Le gros des recherches porte toutefois sur l'organisation moléculaire des génomes chloroplastique et mitochondrial chez diverses plantes (revue dans Palmer 1985) et sur la phylogénie interspécifique (uniquement ou presque à l'aide de l'ADN chloroplastique, de structure moins complexe que le génome mitochondrial). Les premières études de séquençage confirment le faible taux de substitution de ces deux génomes, notamment par comparaison au génome mitochondrial des animaux (dans un rapport de 1 à 100 selon l'estimation de Zurawski et coll. 1985). Les travaux de Wolfe et coll. (1987) basés sur les séquences

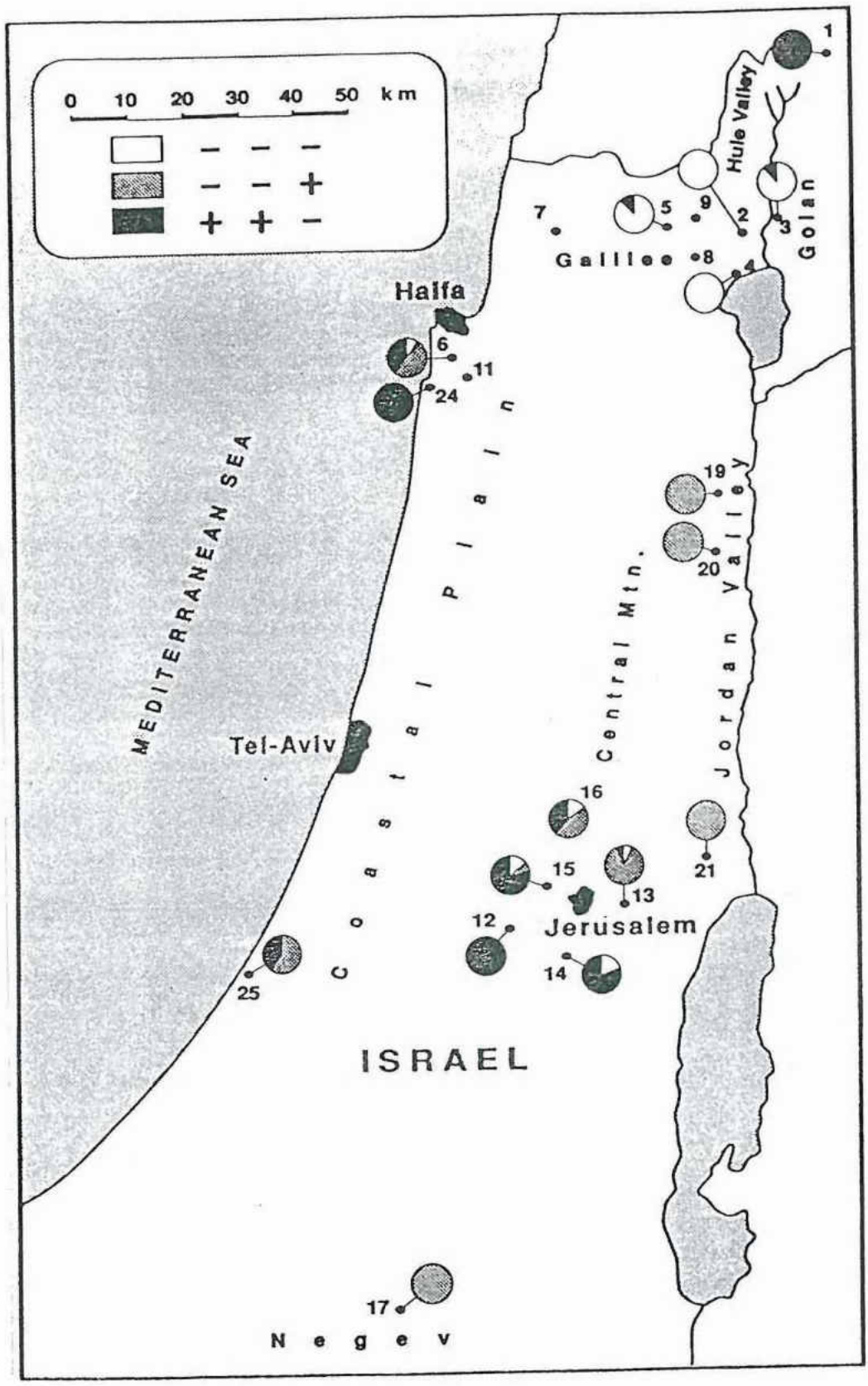


Figure 7 : Variabilité de l'ADN chloroplastique de l'orge *Hordeum vulgare* ssp. *spontaneum* en Israël. Trois variants ont été détectés. Leur fréquence est donnée pour chaque population étudiée. La méthode utilisée est la RFLP avec sondes chloroplastiques marquées radioactivement (Neale et coll. 1988).

qui se sont accumulées pendant la décennie 80, permettent de chiffrer précisément, chez les plantes, les taux de substitution chloroplastique et mitochondrial relativement à celui de l'ADN nucléaire. Le séquençage complet de génomes chloroplastiques par des équipes japonaises est un fait marquant de la fin des années 80 dans le domaine de la biologie moléculaire (Hiratsuka et coll. 1989, Ohyama et coll. 1986, Shinozaki et coll. 1986) (figure 6).

La génétique des populations des lignées cytoplasmiques

Un certain nombre de travaux théoriques commencèrent à apparaître sur la génétique des populations des génomes cytoplasmiques, allant depuis le niveau cellulaire (dérive liée au nombre fini d'organelles par cellule) jusqu'au niveau des populations (Birky 1978, Birky et coll. 1983 1989, Takahata et Maruyama 1981, Takahata et Palumbi 1985). En ce qui concerne la distribution de la diversité cytoplasmique, le message, même s'il ne concernait pas spécifiquement les plantes, était clair : « *a population can be effectively subdivided for organelle genes at migration rates which result in panmixis for nuclear genes, especially if males migrate more than females* » (Birky et coll. 1983). On remarquera, dans tous ces travaux, qu'il n'est jamais fait question des études sur la distribution des noms de famille chez l'homme (caractère transmis paternellement) (Lotka 1931 ; Crow et Mange 1965 ; voir Legay et Vernay 1999 pour une description récente des travaux dans ce domaine). Quelques travaux précoces avaient également considéré les conséquences de la stérilité mâle sur les flux de gènes nucléaires et cytoplasmiques entre populations (Caspari et coll. 1966, Watson et Caspari 1960). L'étude de la généalogie des variants n'était bien sûr pas prise en compte dans toutes ces études, bien que les génomes considérés évoluent de manière clonale.

Premières études de diversité basées sur la technique RFLP avec sondes marquées

Avec la parution de l'ouvrage de Maniatis et coll. (1982), la technique d'hybridation avec des sondes d'ADN préalablement clonées et marquées radioactivement (Southern 1975) devient plus largement disponible. Palmer et coll. clonent ainsi l'ADN chloroplastique du pétunia, qui servira de sonde dans de nombreuses études par la suite. En 1986 et 1987 paraissent deux études basées sur cette technique, toutes deux issues du laboratoire d'Allard en Californie. Ces études mettaient en évidence une variabilité chloroplastique intraspécifique importante sur de grands échantillons (245 individus dans l'étude de Neale et coll. (1986) sur l'orge (figure 7) ; 371 individus dans celle de Wagner et coll. (1987) portant sur deux espèces de pins. Toutefois, la diversité et sa répartition intra- et interpopulation ne furent pas directement quantifiées. L'étude sur l'orge aurait pu se prêter à une telle comparaison. Toutfois, l'orge est une espèce autogame, dont la diversité nucléaire est déjà très fortement structurée du fait des faibles flux par pollen ($G_{ST} = 0,49$). Le contraste entre la différentiation nucléaire et chloroplastique est donc moins fort que pour la plupart des autres angiospermes (voir plus bas). L'étude sur les conifères n'était pas conçue pour mesurer la subdivision de la diversité (présence d'un ou deux individus par population seulement).

Des redécouvertes 'surprenantes'

Désormais, il n'était plus nécessaire de purifier les plastes ou les mitochondries

Table 1. Examples of chloroplast capture due to hybridization/introgression

Example	Evidence for capture[1]	Reference
Argyroxiphium grayanum	C	Baldwin *et al.* (1990)
Brassica napus	C	Palmer *et al.* (1983)
Clarkia sect. *Fibula*	S	Sytsma *et al.* (1990)
Dubautia scabra	C	Baldwin *et al.* (1990)
Fuchsia perscendens	S	Sytsma *et al.* (1991)
F. excorticata	S	Sytsma *et al.* (1991)
Gossypium bickii	C	Wendel *et al.* (1991)
G. aridum	C	Wendel and Albert (1991)
G. cunninghamii	C	Wendel and Albert (1991)
Helianthus annuus ssp. *texanus*	C	Rieseberg *et al.* (1990a)
H. anomalus	C	Rieseberg (1991)
H. debilis ssp. *cucumerifolius*	C	Rieseberg *et al.* (1990a)
H. debilis ssp. *silvestris*	S	Rieseberg *et al.* (1991)
H. deserticola	C	Rieseberg (1991)
H. neglectus	C	Rieseberg *et al.* (1990b)
H. paradoxus	C	Rieseberg *et al.* (1990b)
H. petiolaris ssp. *fallax*	C	Rieseberg *et al.* (1991)
H. petiolaris ssp. *petiolaris*	C	Rieseberg *et al.* (1991)
Heuchera hallii	C	Soltis *et al.* (1991a)
H. micrantha	C	Soltis *et al.* (1991a)
H. nivalis	C	Soltis *et al.* (1991a)
H. parviflora	C	Soltis *et al.* (1991a)
Lycopersicon chilense	S	Palmer and Zamir (1982)
L. chmielewskii	S	Palmer and Zamir (1982)
Persea steyermarkii	C	Furnier *et al.* (1990)
Pisum sativum	C	Palmer *et al.* (1985)
Populus nigra	C	Smith and Sytsma (1991)
Quercus alba	C	Whittemore and Schaal (1991)
Q. macrocarpa	C	Whittemore and Schaal (1991)
Q. michauxii	S	Whittemore and Schaal (1991)
Q. stellata	C	Whittemore and Schaal (1991)
Salix melanopsis	C	Brunsfeld (1990)
S. taxifolia	C	Brunsfeld (1990)
Tellima grandiflora	C	Soltis *et al.* (1991a)
Triticum turgidum	S	Gill and Chen (1987)
Zea perennis	C	Doebley (1989)
Z. mays	S	Doebley and Sisco (1989)

[1]C = convincing evidence for chloroplast capture; S = suggestive evidence for chloroplast capture.

Figure 8 : Table tirée de l'article de Rieseberg et Soltis (1991) : *Phylogenetic consequences of cytoplasmic gene flow in plants*. Exemples de 'captures chloroplastiques' dues à l'hybridation / introgression.

pour étudier leur variabilité ; une extraction d'ADN génomique total (comprenant l'ADN *nucléaire, chloroplastique et mitochondrial*), suffisait. Les travaux sur l'hérédité des plastes et des mitochondries furent aussi relancés. L'hérédité paternelle des chloroplastes de conifères, prédite par les cytologistes, démontrée par Ohba et coll. chez le *Cryptomeria* en 1971, fut retrouvée (Neale et coll. 1986 ; Szmidt et coll. 1987 ; Wagner et coll. 1987). Par contre, l'hérédité des mitochondries était bien maternelle chez des pins (Neale et Sederoff 1988 1989) mais paternelle chez le séquoia, une Taxodiacée, confirmant de nouveau les prédictions des cytologistes, mais surprenant les généticiens américains peu au fait des travaux européens dans ce domaine.

Démonstration de la prévalence de la diversité de l'ADN chloroplastique

Dans une revue sur la diversité intraspécifique de l'ADN chloroplastique parue en 1992, mais couvrant tous les travaux antérieurs à 1991, Soltis et coll. considèrent surtout les implications de cette diversité pour la systématique et la phylogénie. Toutefois, ils démontrèrent que la diversité intraspécifique avait été trouvée pratiquement chaque fois qu'on l'avait recherchée. Ils reviennent sur le titre qu'ils jugent –rétrospectivement– mal inspiré du papier de Banks et Birky cité plus haut : « *Chloroplast DNA diversity is low in a wild plant, L. texensis* ». Soltis et coll. constatent le faible nombre d'études de diversité intraspécifique basés sur l'ADN chloroplastique, et répertorient 25 espèces seulement où au moins 100 individus (ou au moins 10 populations) ont été analysés. Dans un seul de ces 25 cas (*Pennisetum glaucum*), aucune variation n'a pu être détectée. En fait, dans seulement huit de ces 25 études, l'échantillonnage m'a permis –*a posteriori*– de chiffrer la manière dont la diversité est structurée (plusieurs individus par *population* et plusieurs populations représentées). Dans aucun cas, cette structuration (coefficient de différentiation G_{ST}) n'avait été mesurée par les auteurs de l'étude, soit par manque d'intérêt (cas de la plupart des taxonomistes intéressés par la phylogénie, y compris intraspécifique, mais peu par la biologie des populations), soit par manque de matière première (trop faible variation) ou par prudence, l'échantillonnage *étant jugé insuffisant pour de telles* estimations (voir la discussion sur l'échantillonnage au chapitre 1). Notons que les études de Soltis et coll. par exemple combinaient généralement plusieurs individus d'une population lors des analyses (en '*bulk*'), afin d'en augmenter la rapidité. La rareté supposée des cas de diversité intrapopulation (une idée reçue en chassant une autre !) justifiait selon eux cette pratique. Force leur est néanmoins de constater, dans cette revue, que la diversité intrapopulation n'est pas rare ; ils citent également les premiers cas de diversité intra-individu, chez un pin (Govindaraju et coll. 1988) et chez la luzerne (Johnson et Palmer 1989). En outre, ils dressent un premier bilan du niveau de divergence intraspécifique des séquences chloroplastiques, qui varie de 0 à 0,08% selon les espèces étudiées. Cela signifie qu'il n'y a en moyenne pas plus d'une base qui diffère sur 1000, entre deux individus conspécifiques.

Le problème de la capture cytoplasmique : exemple de confusion des genres

Une autre revue sur les travaux des années 80 fut écrite par Rieseberg et Soltis et publiée en 1991. Elle s'intitulait « *Phylogenetic consequences of cytoplasmic gene flow in plants* » (figure 8). Bien que publiée dans une revue d'audience assez

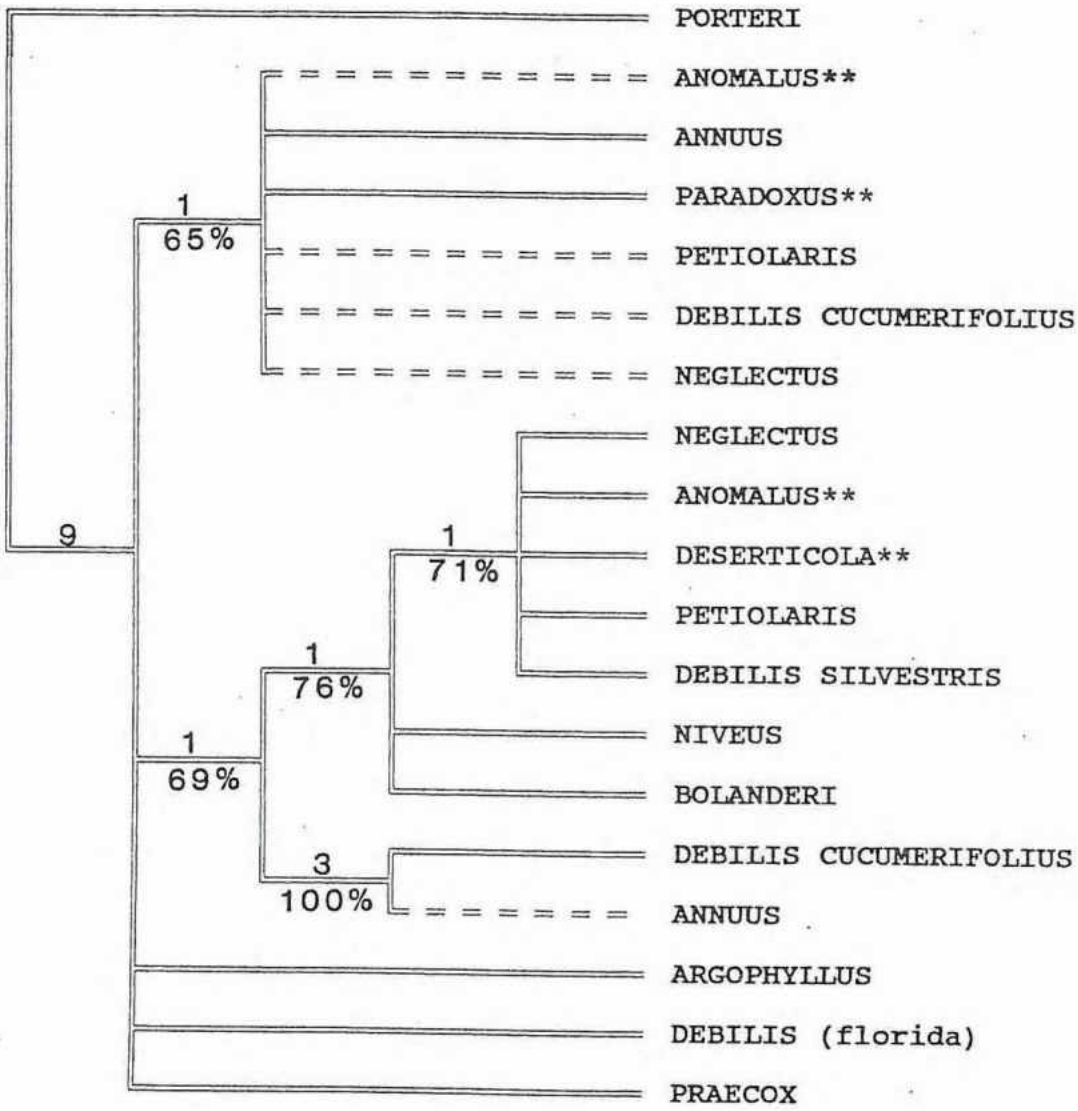


Figure 9 : Arbre phylogénétique basé sur l'ADN chloroplastique dans le genre *Helianthus*. Les valeurs bootstrap sont données en pourcentage. Les lignes pointillées indiquent les discordances entre classification morphologique et chloroplastique, résultant de l'introgression cytoplasmique. ** = espèce hybride stable. D'après Rieseberg & Soltis (1991).

limitée (*Evolutionary Trends in Plants*), elle a été fréquemment citée par la suite, tout comme la revue de Soltis et coll. de 1992. Le titre laissait supposer un intérêt réel pour la biologie des populations (les flux de gènes), malgré l'accent, de nouveau, sur les conséquences pour la phylogénie (interspécifique). Les auteurs concluent que dans de nombreux cas, qu'ils énumèrent, « *cytoplasmic exchange appears to have occurred in the absence of significant nuclear gene flow and sometimes the native cytoplasm has been completely displaced by an alien one* ». A l'exception d'un cas où l'introgression de marqueurs nucléaires avait été détectée en absence d'échanges chloroplastiques (chez des pins dont l'ADN chloroplastique est d'ailleurs transmis paternellement, Wagner et coll. 1987), tous les autres cas répertoriés indiquaient des « transferts » plus rapides pour le génome chloroplastique que pour le génome nucléaire.

Une première explication avancée par Rieseberg et Soltis est la suivante. Les génomes cytoplasmiques -présents en de multiples copies toutes identiques par cellule, donc effectivement haploïdes, sauf cas d'hétéroplasmie-, sont plus sujets à la dérive que le génome nucléaire diploïde (non dans un rapport de 1 à 4 comme l'écrivent ces auteurs, mais dans un rapport de 1 à 2, pour des espèces hermaphrodites). Après rencontre et hybridation de deux espèces, un des génomes cytoplasmiques serait rapidement éliminé par dérive, alors qu'un polymorphisme nucléaire persisterait plus longtemps.

Une deuxième explication serait liée à l'effet de la sélection disruptive, agissant sur le génome nucléaire mais non sur le génome cytoplasmique. Cette sélection, agissant de manière concertée, pourrait maintenir des linkats coadaptés et donc l'intégrité du génome nucléaire, tout en permettant les échanges cytoplasmiques.

Enfin, Rieseberg et Soltis proposent que des hybrides, puis leurs descendants, souvent mâle-stériles, se soient installés dans une population de l'une des deux espèces parentes, et transmettent ainsi leur génome cytoplasmique *sans échanges nucléaires* significatifs.

Bien que certaines de ces interprétations soient pertinentes, le vocabulaire même qui est utilisé pose un sérieux problème. L'association de la phylogénie et de la génétique des populations constitue évidemment un progrès considérable. Dans le cas présent, toutefois, la transposition dans le domaine de la génétique des populations de termes servant à décrire des phénomènes visualisés à l'aide d'arbres phylogénétiques engendre la confusion : une branche de l'arbre phylogénétique semble 'capter' un haplotype d'une autre lignée (figure 9). En estimant la fréquence de tels événements (on parle de réticulation), on estimerait alors les 'flux de gènes' cytoplasmiques. Cela peut laisser entendre que les mouvements (physiques cette fois) des graines et du pollen ont été évalués, alors qu'il n'en est rien. Le cytoplasme est finalement considéré comme '*alien*', et le génome nucléaire '*native*', sous-entendant que les déplacements des graines sont plus importants que ceux des pollens, ce qui paraît difficile à concevoir. Cette confusion est encore apparente dans bon nombre de travaux plus récents.

Début des années 90

Au début des années 90, le séquençage complet de plusieurs génomes chloroplastiques, et l'identification d'une grande majorité des quelques 100 gènes pour lesquels code l'ADN chloroplastique, en faisait un des fleurons de la biologie

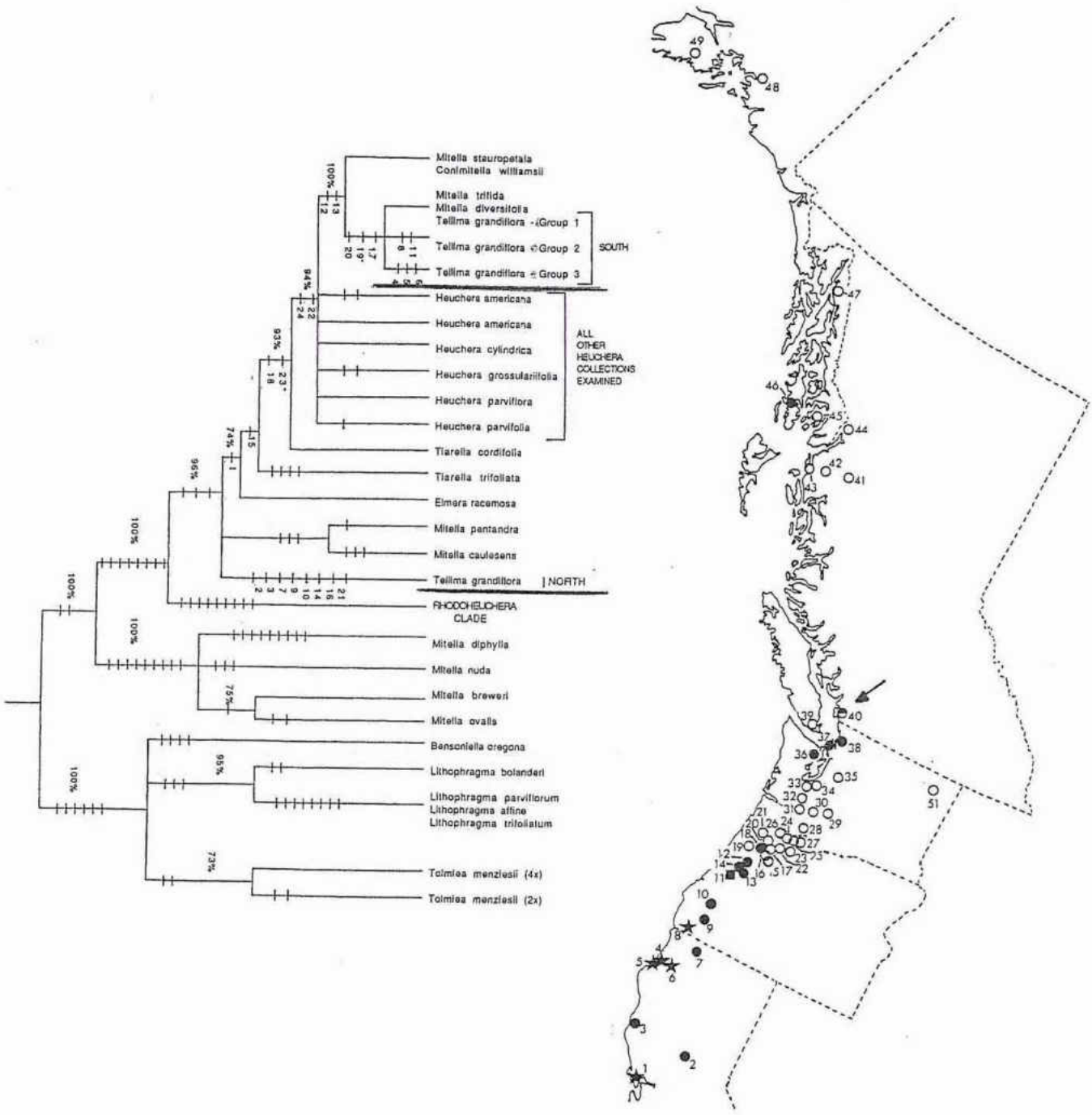


Figure 10 : Localisation géographique des populations de *Tellima grandiflora*, distribution des 4 types chloroplastiques trouvés par la technique RFLP et relations de parenté des 4 haplotypes dans l'arbre phylogénétique des 9 genres du groupe *Heuchera* de la famille des Saxifragacées, auquel appartient *Tellima*. L'haplotype 'Nord' est représenté par des cercles vides, les 3 autres haplotypes par des cercles pleins, des carrés et des étoiles. Une seule population sur 40 était polymorphe.

moléculaire végétale (Sugiura 1992). Celle-ci avait d'ailleurs démarré en 1976 par l'analyse des fragments de restriction des deux génomes cytoplasmiques. Cette avance et le caractère quasi exhaustif de l'information disponible laissait pressentir que les biologistes moléculaires allaient désormais se tourner vers de nouveaux modèles : le génome mitochondrial, plus complexe, responsable de la stérilité mâle cytoplasmique, où a été décrit le phénomène d'*editing*[16] à la fin des années 1980 (Lamattina et coll. 1989) ; et surtout le génome nucléaire, où presque tout restait à découvrir.

Si l'étude moléculaire des génomes chloroplastiques semblait donc devenue moins prioritaire, la nouvelle décade coïncide en revanche avec la multiplication des études de populations, grâce au développement de la phylogéographie et à celui de la génétique des populations des génomes cytoplasmiques des plantes.

Apparition tardive de la phylogéographie pour l'étude des plantes

La première étude de phylogéographie intraspécifique chez les plantes, faisant explicitement référence aux méthodes développées par Avise, est celle de Soltis et coll. parue en 1991 sur *Tellima grandiflora*, une Saxifragacée présente le long de la côte ouest de l'Amérique du Nord. Chez *T. grandiflora*, une discontinuité génétique majeure apparaît entre le nord et le sud de l'aire lors de l'étude de l'ADN chloroplastique, alors que la distribution de l'espèce est parfaitement continue, et que des marqueurs nucléaires n'avaient rien révélé de particulier (figure 10). Une explication en termes de processus historiques (refuges) était également avancée dans ce travail.

Le terme de phylogéographie a été forgé par Avise en 1987, mais son papier sur les rongeurs publié en 1979 marque réellement le début de cette discipline : il aura donc fallu 12 ans pour adopter ces concepts et les méthodes d'analyse à l'étude des plantes. Par comparaison, la technique des isozymes et les méthodes permettant l'analyse des données correspondantes s'étaient répandus en moins de cinq ans (1966-1971). Le faible niveau de variabilité intraspécifique des génomes cytoplasmiques végétaux ne semble pas suffire à expliquer à lui seul un aussi long délai. La difficulté d'associer des disciplines différentes pourrait expliquer ce retard. A l'époque, la phylogénie constitue le domaine des systématiciens ; la biogéographie est dominée par des écologistes, et la génétique des populations, par des généticiens très influencés par les théories de Fisher et Wright et par des débats très fondamentaux sur l'évolution. Le chapitre 2 (*History of molecular phylogenetics*) du livre d'Avise (1994) illustre précisément les difficultés rencontrées pour arriver à sortir des 'diversions' et autres préoccupations et controverses très fondamentales qui monopolisaient l'attention des généticiens des populations.

Apparition tardive de la génétique des populations des génomes cytoplasmiques chez les plantes

Les travaux utilisant les isozymes avaient permis d'emblée, dès l'émergence de la technique en 1966, l'analyse de grands effectifs, nécessaires pour les études de génétique des populations. Les techniques de biologie moléculaire sont restées longtemps plus laborieuses. En outre, la faible diversité

[16] *Editing* : processus de correction de mutations au stade de l'ARN messager, mis en évidence en séquençant un gène avant transcription (ADN génomique) et après (à partir de l'ADN copie, obtenu par transcription réverse des ARN messagers).

intrapopulation des génomes cytoplasmiques chez les plantes n'incitait pas à échantillonner au sein des populations. Or, par analogie avec les études correspondantes basées sur les isozymes, l'estimation des paramètres habituels de diversité semblait nécessiter au minimum des effectifs de plusieurs dizaines d'individus par population. Comme je l'ai démontré dans le chapitre précédent, il s'agissait là d'une idée reçue car des estimations très précises du coefficient de différentiation (G_{ST}) et de la diversité totale peuvent être obtenues avec des petits effectifs par population, à condition qu'un grand nombre de populations soit analysé.

Autre diversion, l'idée déjà à l'esprit des précurseurs des années 70 d'étudier les associations nucléo-cytoplasmiques revenait, avec la publication d'une série de papiers sur les déséquilibres nucléo-cytoplasmiques (Arnold et coll. 1988, Asmussen et coll. 1987, Schnabel et Asmussen 1989). Théorie complexe et difficilement applicable (car nécessitant des effectifs considérables pour pouvoir être testée), qui voyait le jour alors que les modèles plus simples de génétique des populations n'avaient pas encore été appliqués ni adaptés réellement au cas des plantes. Les applications pratiques de l'étude de ces déséquilibres furent peu nombreuses chez les plantes. On peut citer celles de l'équipe de D. Wagner dans le Kentucky sur les pins contorta et banksiana (Dong et Wagner 1994, Wagner et coll. 1991).

Subdivision de la diversité chloroplastique et nucléaire : premières données empiriques

Les premiers travaux où la subdivision de la diversité cytoplasmique a été estimée et contrastée avec celle du génome nucléaire chez une plante sont ceux issus de ma thèse sur les chênes blancs européens (Kremer et coll. 1991[17], Petit 1992[18], Petit et coll. 1993a[19]).

Ils étaient basés sur l'étude de 834 arbres issus de 107 populations, analysés avec la technique RFLP avec sonde d'ADN chloroplastique marquée radioactivement, puis directement sur gel d'agarose.

La diversité nucléotidique, a été estimée à : $\pi = 0{,}0006$, soit une différence d'un nucléotide toutes les 1400 bases entre deux individus pris au hasard. La distribution inter- et intrapopulation de quatre haplotypes a ensuite été étudiée. Une différenciation très élevée pour le génome chloroplastique ($G_{ST} = 0{,}90$) contraste avec une différenciation remarquablement faible pour les allozymes ($G_{ST} = 0{,}03$). Les potentialités des marqueurs à hérédité maternelle pour différencier les populations apparaissaient désormais pleinement.

Subdivision de la diversité cytoplasmique et nucléaire : application de la théorie aux plantes

En outre, les modèles de génétique des populations de Birky et coll. (1989) furent appliqués aux plantes hermaphrodites. Celles-ci sont caractérisées par des flux de pollen (haploïde) et de graines (diploïdes).

---

[17] Kremer A, R J Petit, A Zanetto, V Fougère, A Ducousso, D B Wagner, C Chauvin (1991) Nuclear and organelle gene diversity in *Quercus robur* and *Q. petraea*. In: Genetic variation in European populations of forest trees. Ed: G. Müller-Starck, M. Ziehe. Sauerländer's Verlag, Frankfurt am Main, 141-166.

[18] Petit R J. 1992. Polymorphisme de l'ADN chloroplastique dans un complexe d'espèces: les chênes blancs européens. Thèse, Paris XI-Orsay.

[19] Petit R J, A Kremer et D B Wagner. 1993a. Geographic structure of chloroplast DNA polymorphisms in European oaks. *Theoretical and Applied Genetics*, **87**, 122-128.

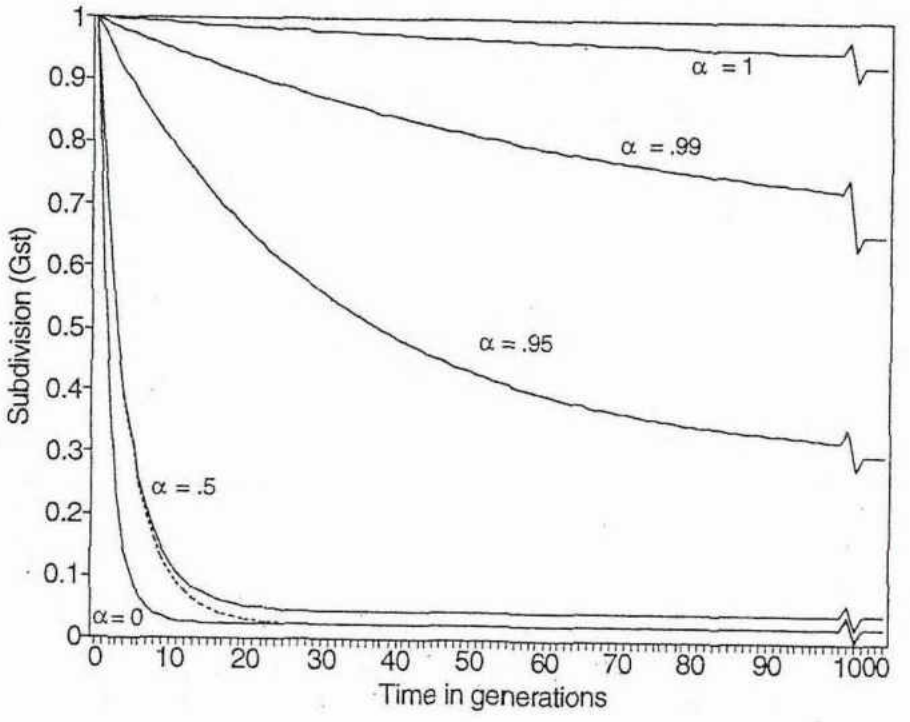


Figure 11 : Approche de l'équilibre pour le coefficient de différentiation G_{ST} dans un modèle en île, selon le mode d'hérédité du gène cytoplasmique ($\alpha = 1$, hérédité maternelle, $\alpha = 0$, hérédité paternelle), en cas de forte asymétrie des flux de gènes ($m_P >> m_S$). Dans ce cas, le mode de transmission a un effet prépondérant. D'après Petit et coll. (1993).

Cette plus grande différenciation cytoplasmique était parfaitement prévisible, du fait de la dérive plus grande des génomes haploïdes et surtout de l'asymétrie assez générale des flux par pollen et par graines (Petit et coll. 1993b[20]). Par contre, en cas de transmission paternelle, la différenciation attendue à l'équilibre est proche (légèrement inférieure) de celle de gènes à transmission mendélienne (figure 11). L'estimation conjointe de la différenciation nucléaire et cytoplasmique permettait en outre, moyennant certaines hypothèses assez lourdes (modèle en île, équilibre mutation dérive, taille constante des populations, tailles efficaces mâle et femelle identiques...), d'estimer l'importance relative des flux par pollen et par graine (Petit 1992), un résultat retrouvé indépendamment par Ennos (1994). Ces travaux, parce qu'ils démontraient qu'on pouvait étudier certains processus biologiques non accessibles avec les seuls marqueurs nucléaires, furent fréquemment cités par la suite et suscitèrent de nouvelles recherches sur la biologie des populations de plantes (revue dans McCauley 1995).

Capture cytoplasmique ou déplacement d'espèce ?

Trois travaux étaient parus presque simultanément sur la variabilité de l'ADN chloroplastique des chênes (Whittemore et Schaal 1991, Kremer et coll. 1991 puis Petit et coll. 1993a, Ferris et coll. 1993). Dans chacune de ces études, plusieurs espèces étaient représentées. Le genre *Quercus* est très important (450 espèces environ) et pose en effet des problèmes de délimitation des espèces particulièrement marqués, déjà mis en évidence par M. A. de Candolle au début du XIXème siècle et repris par Darwin dans son chapitre 2 de l'Origine des espèces (1858) (à la rubrique *Espèces douteuses*).

Les trois travaux cités avaient tous démarré avant la publication de l'article de Rieseberg et Soltis (1991), signalant la fréquence des cas de 'capture' cytoplasmique. L'objectif consistait donc encore à tenter, avec ces marqueurs utilisés intensivement par les taxonomistes, de distinguer enfin ces espèces de chênes. Bien au contraire, dans les trois cas, une quasi absence de différenciation entre espèces, associée à une très forte structuration géographique, a été mise en évidence.

Les travaux théoriques précédents et l'asymétrie des coefficients de différenciation nucléaires et cytoplasmiques indiquent que les flux de gènes par pollen sont sans doute considérablement plus élevés que ceux par pollen (dans un rapport de 100 à 1 au moins). En outre, des études sur le système de reproduction de ces chênes, publiées ultérieurement (Bacilieri et coll. 1996[21]), montrent que les croisements interspécifiques sont fréquents mais asymétriques.

Pour rendre compte de la similarité de la structuration géographique de l'ADN chloroplastique de différentes espèces, le modèle suivant a été proposé, qui est exactement l'antithèse de celui avancé par Rieseberg et Soltis en 1991 (Petit 1992[22], Petit et coll. 1997).

Des flux polliniques interspécifiques à longue distance permettent à une espèce de s'installer là où une espèce apparentée s'était établie initialement, par hybridations et croisements en retour répétés, en intégrant le

[20] Petit R J, A Kremer et D B Wagner. 1993b. Finite island model for organelle and nuclear genes in plants. *Heredity*, **71**, 630-641.

[21] Bacilieri R, A Ducousso, R J Petit et A Kremer. 1996. Mating system and asymmetric hybridization in a mixed stand of European oaks. *Evolution* **50**, 900-908.

[22] Petit R J. 1992. Polymorphisme de l'ADN chloroplastique dans un complexe d'espèces: les chênes blancs européens. Thèse, Paris XI-Orsay.

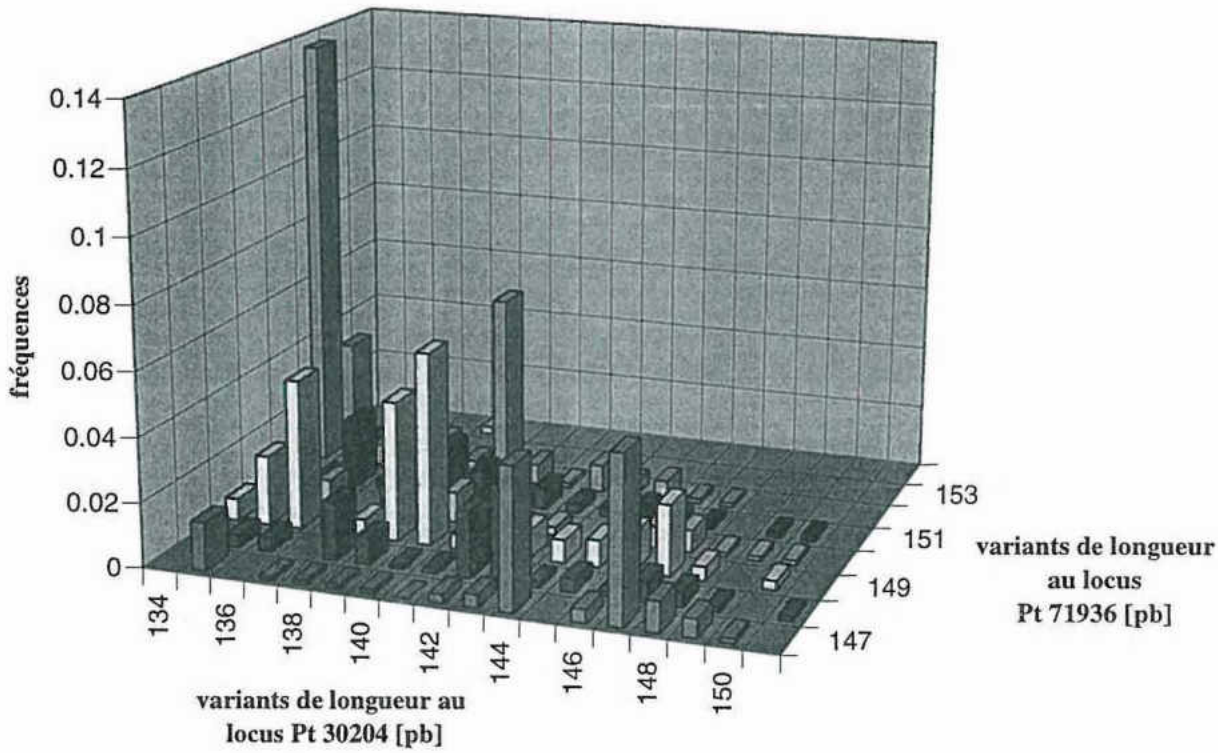


Figure 12 : Variants de longueur trouvés à deux locus microsatellites chez le sapin (*Abies alba*). Le nombre de combinaisons observées parmi 714 individus de 17 populations est 90 (le nombre maximum de combinaisons possibles est 144), soulignant le taux de réversion très important à ces deux locus. D'après Vendramin et coll. (1999).

génome maternel de l'espèce arrivée en premier. Ce serait donc le génome nucléaire qui se déplacerait sur une base maternelle parfaitement statique : des flux de gènes nucléaires auraient donc lieu en l'absence de flux cytoplasmiques, aboutissant à l'introgression des génomes cytoplasmiques maternels dans un nouveau fonds nucléaire. Je reviendrai sur ce mécanisme dans le chapitre suivant, car il fait beaucoup plus place à l'histoire et à la dynamique écologique des populations que le modèle de 'capture cytoplasmique' de Rieseberg décrit précédemment.

Introduction de la technique PCR

A chaque décennie sa nouvelle technique d'analyse. En 1986, la technique de la PCR (*Polymerase Chain Reaction*) est décrite par Mullis et coll. mais c'est dans les années 1990 que les thermocycleurs et les polymérases thermostables se répandront massivement dans les laboratoires.

Pierre Taberlet et Ludovic Gielly, à Grenoble, proposent les premières amorces consensus pour les chloroplastes, permettant l'amplification de séquences non-codantes particulièrement variables (Taberlet et coll. 1991). Quelques autres amorces seront décrites ensuite, plus ou moins conservées (Arnold et coll. 1992, Liston 1992, Rieseberg et coll. 1992). C'est ensuite la publication par notre équipe d'un jeu de 12 paires d'amorces chloroplastiques et mitochondriales très conservées qui va faciliter désormais les études de diversité (Demesure et coll. 1995[23]).

[23] Demesure B, N Sodzi et R J Petit. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Mol. Ecol.* **4**, 129-131.

Chez les conifères, et en particulier les pinacées, l'étude de microsatellites chloroplastiques a permis la détection d'une diversité plus grande encore (Powell et coll. 1995, Vendramin et coll. 1996, Vendramin et coll. 1999[24]) (figure 12).

TRAVAUX REALISES

Mise au point d'amorces universelles

J'ai mentionné ci-dessus la première liste d'amorces universelles publiée par notre équipe (Demesure et coll. 1995). A ce jour, j'ai eu connaissance de 31 études utilisant soit ces amorces, soit celles de Taberlet et coll. (1991), soit le plus souvent une combinaison des deux, pour des études de structuration des populations (*cf.* la revue bibliographique présentée plus loin). Nos amorces ont également servi à d'autres applications (phylogénie, hérédité des génomes cytoplasmiques, etc.).

Ce jeu d'amorces fut ensuite complété. La dernière publication recense 28 paires d'amorces, permettant d'amplifier respectivement 47 et 24 kilobases des génomes chloroplastique et mitochondrial (Dumolin-Lapègue et coll. 1997[25]).

Ce projet continue, et Delphine Grivet, actuellement en thèse, complète cette étude pour aboutir à l'amplification intégrale de la grande région simple copie du génome chloroplastique (la région généralement décrite comme la plus variable), soit 86 kilobases chez le tabac. De plus, les amorces

[24] Vendramin G G, B. Degen, R J Petit, M Anzidei, A Madaghiele et B Ziegenhagen. 1999. High level of variation at *Abies alba* microsatellite loci in Europe. *Mol. Ecol.*, sous presse.

[25] Dumolin-Lapègue S, M H Pemonge et R J Petit. 1997. An enlarged set of consensus primers for the study of organelle DNA in plants. *Mol. Ecol.* **6**, 393-397.

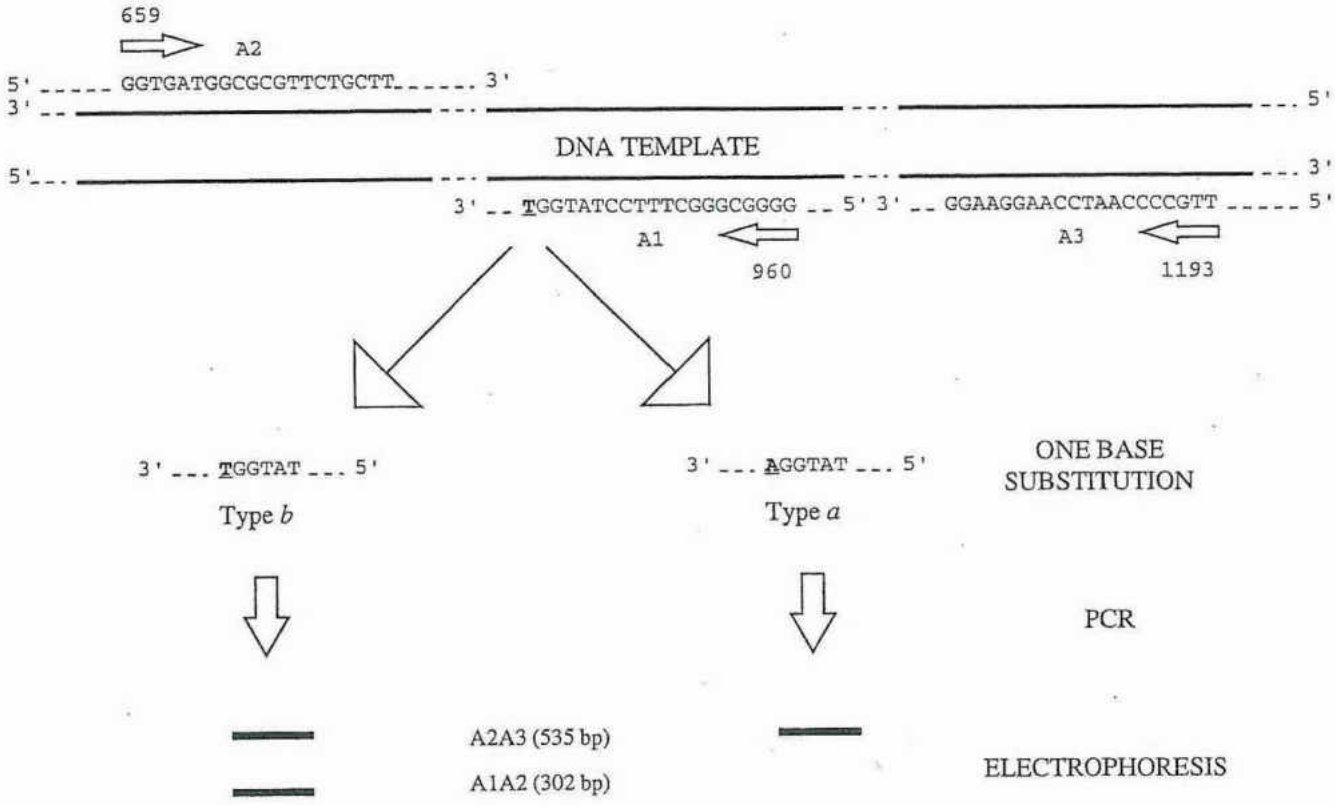


Figure 13 : Méthode d'amplification spécifique d'allèle, destinée à la caractérisation rapide de mutations de séquence connue chez de nombreux individus. Trois amorces sont nécessaires, dont une (A1) se termine au site nucléotidique polymorphe, A3 servant à obtenir un contrôle de PCR. D'après Dumolin Lapègue et coll. 1998.

amplifiant les plus grands fragments (4-5 kb), souvent délicats à amplifier, sont combinées avec de nouvelles amorces internes afin d'amplifier des fragments de taille plus faible (1,5-3 kb), plus faciles à utiliser pour les études de diversité. A ce jour, nous disposons d'amorces pour l'amplification d'environ 80% de cette région (près de 70 kb, donc plus de la moitié de l'ADN chloroplastique).

Méthodes de détection de mutation et d'analyse rapide de la variabilité

A partir de ces fragments amplifiés en PCR, des méthodes de détection de mutation peuvent être mises en œuvre. La technique standard, utilisée dans toutes les études effectuées au laboratoire, consiste à digérer ces fragments par des enzymes de restriction (PCR-RFLP[26]). L'électrophorèse qui fait suite peut être plus ou moins résolutive, selon la dimension des gels d'acrylamide utilisés.

Sylvie Dumolin-Lapègue, au cours de sa thèse, a utilisé la technique SSCP[27] pour la recherche de mutations très rares et isolées, notamment dans le génome mitochondrial (Dumolin-Lapègue et coll. 1997[28]).

Elle a ensuite utilisé une méthode très rapide d'amplification spécifique d'allèles pour étudier une substitution particulière du génome mitochondrial déterminée préalablement par SSCP puis caractérisée par séquençage.

Des amorces ont été mises au point de façon à n'amplifier qu'un des deux variants, avec un contrôle interne de PCR (Dumolin-Lapègue et coll. 1998[29]) (figure 13). Le séquençage de fragments chloroplastiques correspondant à des haplotypes différents repérés par PCR-RFLP a permis de détecter des motifs microsatellites chez le chêne, et de mettre au point des amorces internes pour l'amplification de petits fragments (<300 pb) susceptibles d'amplifier de l'ADN très dégradé (Dumolin-Lapègue et coll. 1999[30]).

Hérédité des génomes cytoplasmiques

Comme l'a montré la partie précédente, les deux génomes cytoplasmiques ne sont pas toujours transmis maternellement chez les plantes, et il n'est pas rare que leur mode de transmission soit discordant. Il convient donc de vérifier dès que possible cette hérédité. Plus une technique moléculaire est efficace pour détecter du polymorphisme, plus grande sera la probabilité qu'un croisement intraspécifique soit informatif pour le marqueur considéré (père différent de la mère). Fréquemment, les premières études d'hérédité utilisant des marqueurs moléculaires se basaient sur des croisements interspécifiques ; dans ce cas, des incompatibilités nucléo-cytoplasmiques peuvent engendrer des résultats non représentatifs.

---

[26] RFLP : *Restriction Fragment Length Polymorphism* : technique consistant à fragmenter l'ADN en des sites spécifiques et à séparer les fragments obtenus sur gels d'électrophorèse.

[27] SSCP : *Single-Stranded Conformation Polymorphism* : les fragments d'ADN rendus simple brin par chauffage sont séparés en électrophorèse en conditions dénaturantes. La migration de ces brins peut être affectée par de simples changements de conformation dus à des substitutions.

[28] Dumolin-Lapègue S, C Bodénès et R J Petit. 1997. Detection of rare polymorphisms in mitochondrial DNA of oaks with PCR-RFLp combined to SSCP analysis. *Forest Genetics* **3**, 227-230.

[29] Dumolin-Lapègue S, M-H Pemonge et R J Petit. 1998. Association between chloroplast and mitochondrial lineages in oaks. *Mol. Biol. Evol.* **15**, 1321-1331.

[30] Dumolin-Lapègue S, M-H Pemonge, L. Gielly, P Taberlet et R J Petit. Amplification of oak DNA from ancient and modern wood. Soumis à *Plant. Mol. Biol.*

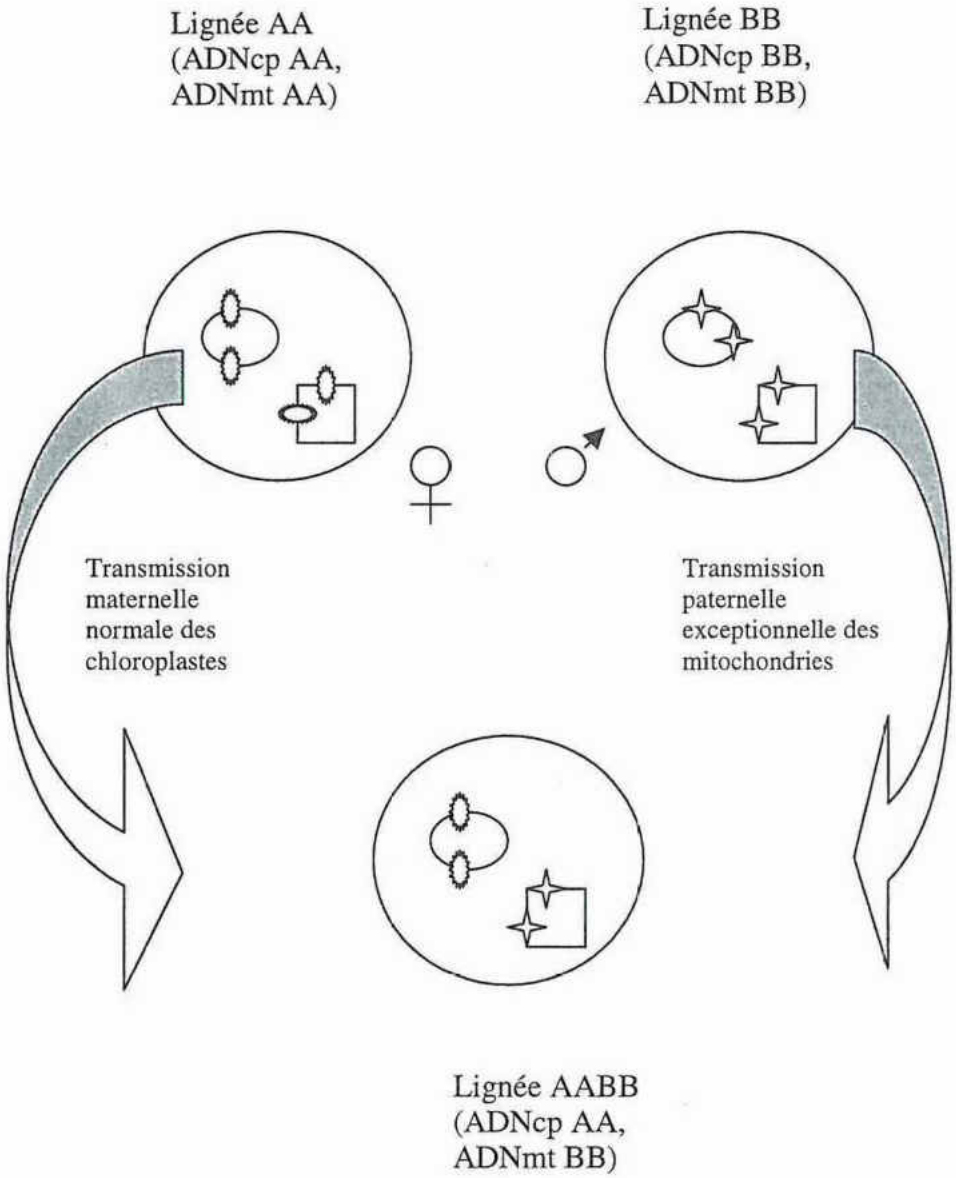


Figure 14 : Conséquence d'une transmission différentielle de l'un des deux génomes cytoplasmiques (ADNcp ou ADNmt), sur les haplotypes combinés chloroplastiques et mitochondriaux, chez une espèce où le mode de transmission normal est la transmission maternelle. Le nombre d'associations est 'multiplié', comme s'il y avait 'recombinaison'.

Deux études destinées à étudier simultanément le mode d'hérédité des deux génomes cytoplasmiques ont été réalisées. L'hérédité maternelle stricte des deux génomes cytoplasmiques a été démontrée chez le chêne pédonculé (*Quercus robur*), à l'aide de la technique PCR-RFLP ; il y avait plus d'une centaine de descendants répartis dans 3 ou 4 croisements (Dumolin et coll. 1995[31]). Joëlle Chat, du laboratoire d'arboriculture fruitière de l'INRA à Bordeaux, a étudié l'hérédité cytoplasmique chez le kiwi (*Actinidia deliciosa*). Dans ce genre, une étude précédente avait démontré l'hérédité paternelle des chloroplastes dans des croisements interspécifiques (Cipriani et coll. 1995).

En utilisant des croisements intraspécifiques, l'hérédité paternelle stricte a été confirmée ; par ailleurs, il a pu être démontré que les mitochondries étaient simultanément transmises maternellement (Chat et coll. 1999[32]). On retrouve donc la même configuration que pour les conifères. Il s'agit du seul cas d'hérédité paternelle stricte des chloroplastes décrit à ce jour chez les angiospermes (des cas de transmission paternelle des mitochondries ont été décrits chez la banane et chez des cucurbitacées).

Démonstration de l'association stricte des deux lignées cytoplasmiques

Les plantes disposent de deux génomes cytoplasmiques. Quand ceux-ci sont transmis de manière opposée (cas des conifères ou du kiwi), l'intérêt est évident pour le généticien : il peut notamment comparer la différenciation pour deux marqueurs transmis uniparentalement, donc effectivement haploïdes, mais qui migrent par graine seulement (hérédité maternelle), ou par pollen et graine (hérédité paternelle). Qu'en est-il cependant de deux génomes transmis maternellement ? Leur étude combinée va-t-elle nous apprendre quelque chose de plus que l'étude d'un seul des deux génomes ? Si ces génomes sont rigoureusement co-transmis, ils se comporteront effectivement comme une seule lignée (avec toutefois des taux de substitution et des modes d'évolution contrastés). Par contre, si des 'fuites' paternelles peuvent se produire pour l'un ou l'autre des génomes, même à très faible fréquence, il en résultera de nouvelles combinaisons entre haplotypes chloroplastiques et mitochondriaux (figure 14). L'étude d'un nombre même limité de mutations situées sur les deux génomes devrait donc pouvoir nous renseigner sur le caractère strict ou non du mode de transmission cytoplasmique. Une telle étude a été effectuée chez les chênes. L'analyse de 378 populations dans une zone de forte diversité cytoplasmique (chloroplastique et mitochondriale) n'a pas permis de détecter de combinaisons impliquant une contamination paternelle. Nous en avons conclu que les deux lignées ont été transmises maternellement, sans exception, sur de très longues périodes (Dumolin-Lapègue et coll. 1998[33]).

Etude de la variabilité cytoplasmique des chênes blancs européens

Ce travail, en prolongement direct de ma thèse, représente certainement l'étude de

[31] Dumolin S, B Demesure et R J Petit. 1995. Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theor. Appl. Genet.* **91**, 1253-1256.

[32] Chat J, L Chalak L et R J Petit. Strict paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in intraspecific crosses of kiwifruit. *Theor. Appl. Genet.*, en révision.

[33] Dumolin-Lapègue S, M-H Pemonge et R J Petit. 1998. Association between chloroplast and mitochondrial lineages in oaks. *Mol. Biol. Evol.* **15**, 1321-1331.

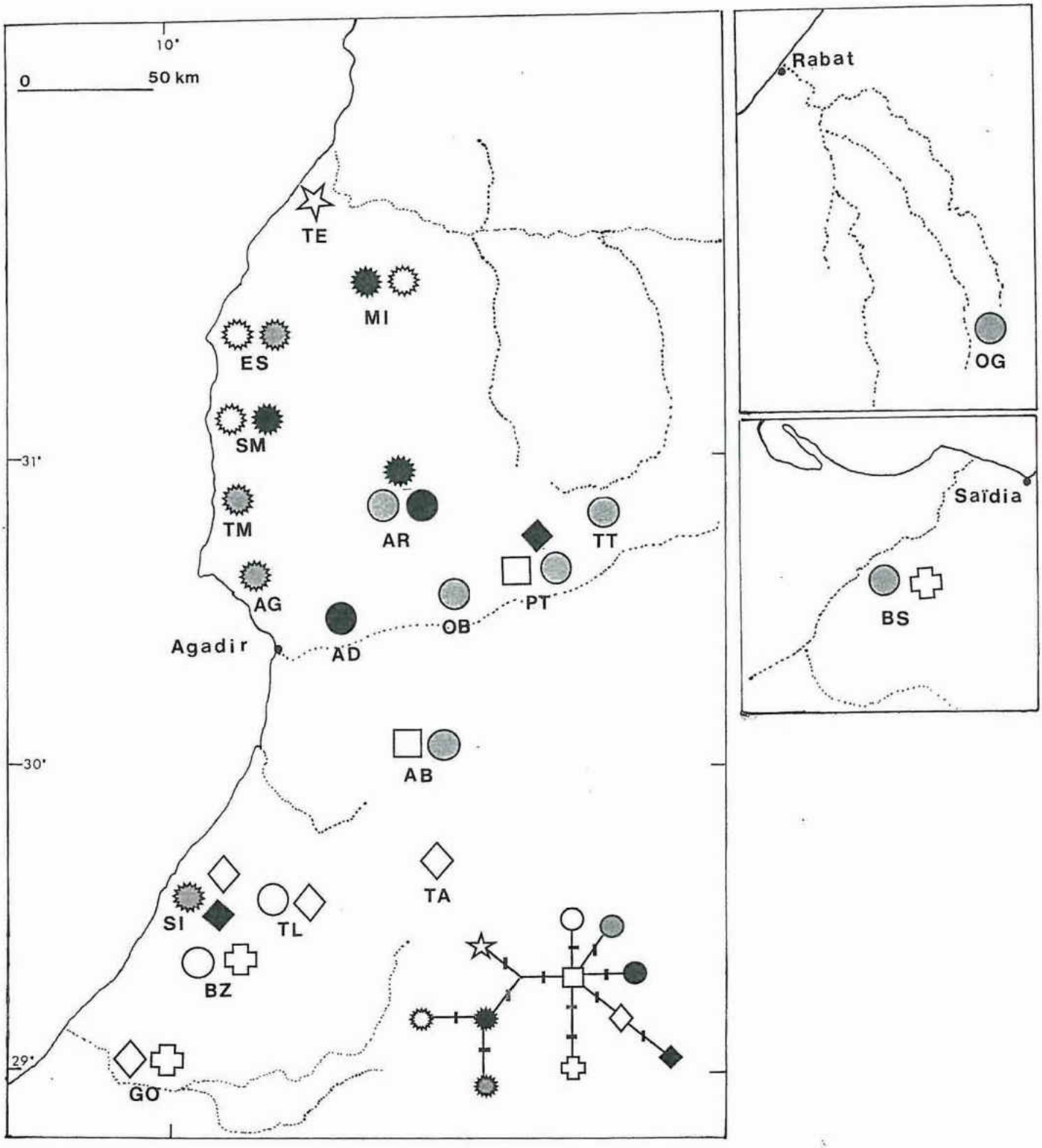


Figure 15 : Variabilité de l'ADN chloroplastique de l'arganier au Maroc. Les deux isolats du nord du Maroc (OG et BS) n'ont pas de variants originaux ; ils représenteraient donc plutôt des transferts durant l'époque historique, plutôt que des « relictes » témoignant de l'extension ancienne de l'espèce. D'après El Mousadik et Petit (1996).

variabilité cytoplasmique la plus importante réalisée à ce jour chez les végétaux (6000 individus typés environ au laboratoire, par PCR-RFLP, à l'aide de deux à quatre combinaisons amorces/enzyme de restriction). Les différentes études qui se succédé au laboratoire sur cette espèce seront décrites en détail dans le chapitre suivant. En ce qui concerne les résultats relatifs à la biologie de la dispersion des graines, les dernières études effectuées fournissent des éléments comparatifs entre espèces de chênes (Dumolin-Lapègue et coll. 1999[34] ; Latouche-Hallé 1998[35]).

Les coefficients de différentiation, quelque soit l'échelle de l'étude, sont élevés, mais diffèrent entre les espèces. En France, le coefficient de differentiation est maximum chez le chêne pubescent, intermédiaire chez le chêne sessile, et plus faible chez le chêne pédonculé (G_{ST} = 0,87, 0,83 et 0,75 respectivement). En outre, pour le chêne pubescent, le niveau de différenciation est hétérogène géographiquement : très élevé dans le sud-est de la France, où l'espèce est présente à l'état pur (chênaies pubescentes climaciques), il diminue ailleurs, là où l'espèce est en mélange avec d'autres espèces. Dans ces milieux, le chêne pubescent est caractérisé par un comportement plus pionnier.

La plus grande différenciation du chêne sessile comparé au chêne pédonculé va parfaitement dans le sens prédit par l'écologie comparée des deux espèces. Le chêne pédonculé est en effet considéré comme plus pionnier, le chêne sessile étant plus forestier. En outre, dans le passé, le chêne pédonculé a sans doute été préférentiellement planté par l'homme, vu son importante fructification. Ces résultats ont été confirmés en Allemagne, où le chêne sessile est également plus différencié que le chêne pédonculé (König et coll. 1998).

Etudes de variabilité cytoplasmique chez d'autres espèces que les chênes blancs

L'étude de la variabilité de l'ADN chloroplastique de l'arganier du Maroc (*Argania spinosa*, Sapotacée), réalisée par A. El Mousadik (Doctorat d'Etat marocain, université d'Agadir) au laboratoire, a permis de mettre en évidence l'efficacité de la technique dans des délais très courts (3 mois).

En effet, 12 fragments d'ADN chloroplastique ont été amplifiés par PCR et digérés par un seul enzyme. Onze haplotypes a ainsi pu être détectés sur 19 populations et 95 individus (El Mousadik et Petit 1996[36]) (figure 15).

La différenciation génétique est moyenne chez cette espèce dont les fruits sont disséminés par des ruminants ou par l'homme (G_{ST} = 0,61), mais elle augmente quand on prend en compte les similarités des haplotypes (N_{ST} = 0,73, significativement plus grand que le G_{ST}), indiquant l'existence d'une structuration phylogéographique.

La même méthode a été utilisée par Nicole Muloko-Ntoutoume, doctorante gabonaise, lors d'un court séjour au laboratoire, pour étudier la variabilité de l'okoumé (*Aucoumea klaineana*, Burseraceae). Cette fois, 14 fragments chloroplastiques ont été amplifiés et digérés par deux enzymes : une seule substitution a pu être détectée parmi 80 individus couvrant une large partie de l'aire de l'espèce, au Gabon.

[34] Dumolin-Lapègue S, A Kremer et R J Petit Are chloroplast and mitochondrial DNA variation species-independent in oaks? *Evolution*, sous presse.

[35] Latouche-Hallé C. 1998. Variabilité de l'ADN chloroplastique des chênes blancs en France. DEA de Biologie, Diversité et Adaptation des Plantes cultivées, Paris XI Orsay.

[36] El Mousadik A et R J Petit. 1996. Chloroplast DNA phylogeography of the argan tree of Morocco. *Mol. Ecol.* **5**, 547-555.

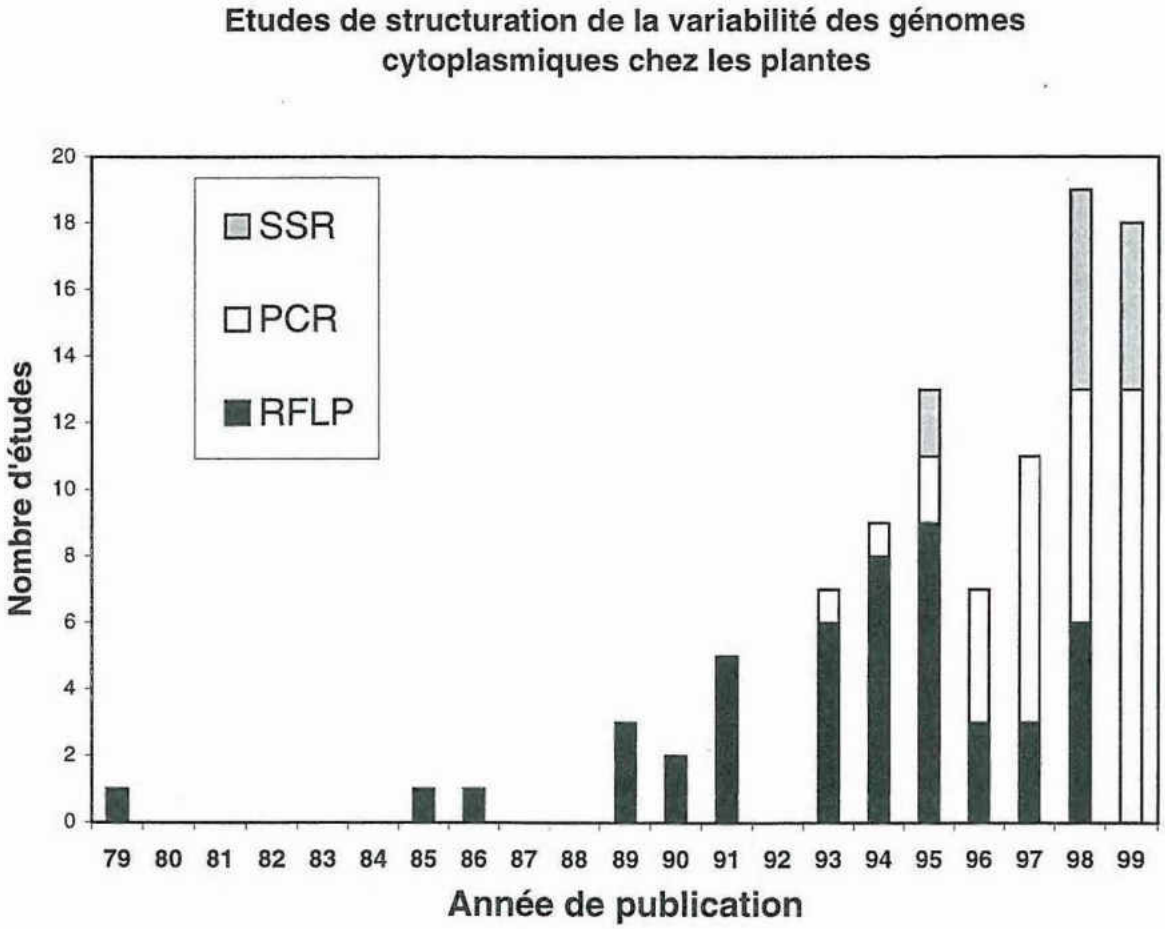


Figure 16 : Evolution du nombre d'études sur la structuration de la diversité cytoplasmique chez les plantes, par technique d'étude: (1) la RFLP avec sondes marquées ou la RFLP directe à partir de chloroplastes purifiés (une seule étude, celle de Scowcroft en 1979) (en noir) ; les techniques dérivées de la PCR (principalement PCR-RFLP), à l'exclusion de l'étude des microsatellites ; (3) l'étude des microsatellites chloroplastiques.

La différenciation est relativement faible (G_{ST} = 0,54) chez cette espèce colonisatrice aux graines disséminées par le vent (Muloko-Ntoutoume et coll. 1999[37]).

Une étude plus complète a été réalisée sur la variabilité chloroplastique du hêtre (*Fagus sylvatica*) dans le cadre de la thèse de Brigitte Demesure (Demesure et coll. 1996). Elle portait sur 83 populations réparties sur l'ensemble de l'aire naturelle, et se poursuit dans le cadre d'un projet européen, en collaboration avec deux autres laboratoires italien et autrichien. La différenciation, mesurée à l'aide de 11 haplotypes, est très forte (G_{ST}>80%). Les autres résultats seront discutés dans le chapitre suivant.

Deux études successives ont été réalisées sur la variabilité de l'ADN chloroplastique d'un arbre de Guyane française, l'angélique *Dicorynia guyanensis* (Césalpiniacée) (DEA de Sabine Dumas et de Gilles Marque), permettant de détecter 9 haplotypes. La diversité intrapopulation est plus élevée chez cette espèce que chez les autres espèces étudiées jusqu'ici (G_{ST} = 0,41).

Une étude est en cours sur le chêne liège (*Quercus suber*) et le chêne vert (*Q. ilex*), au Maroc (thèse d'Etat de Nadia Belhabib) et en Espagne (thèse de Pilar Jimenez). Les résultats illustrent de nouveau l'importance des échanges cytoplasmiques interspécifiques, cette fois entre des espèces considérées comme beaucoup plus distinctes (les chênes vert et liège appartiennent à des sections différents du genre *Quercus*). La différenciation entre populations est très forte, comme chez les chênes du groupe *robur*.

[37] Muloko-Ntoutoume N, R J Petit, L White et K Abernathy. Chloroplast DNA variation of a rainforest tree (*Aucoumea klaineana*, Burseraceae) in Gabon. Soumis à *Mol. Ecol.*

STRUCTURATION DES GENOMES CYTOPLASMIQUES CHEZ LES PLANTES : REVUE BIBLIOGRAPHIQUE[38]

Présentation de l'étude

L'étude de la variabilité de l'ADN chloroplastique et mitochondrial a 10 ans de retard sur l'étude de la variabilité des protéines enzymatiques (technique des isozymes) (1966-1976). Le premier bilan important des études de diversité utilisant les isozymes –qui incluait 100 espèces de plantes différentes– est paru 13 ans après l'émergence de la technique (Hamrick et coll. 1979). En 1999, 23 ans après les premières études moléculaires sur les génomes cytoplasmiques des plantes, on est loin de pouvoir effectuer une synthèse aussi complète que celle d'Hamrick et coll., même si les données sont en train de s'accumuler à un rythme beaucoup plus soutenu qu'auparavant. Les données dont j'ai pu disposer sont fréquemment présentées de manière incomplète, l'échantillonnage est souvent loin d'être optimal, même quand le principal objectif poursuivi était la description de la diversité cytoplasmique. Des informations simples sur la biologie des espèces concernées font le plus souvent défaut dans ces publications, et les mesures de diversité ou de différenciation utilisées sont loin d'être standardisées.

L'objectif de cette revue est de démontrer l'intérêt de ces comparaisons entre espèces et de faciliter les synthèses futures en indiquant comment recueillir et présenter les données pour qu'elles puissent

[38] Petit, R J, S Fineschi et G G Vendramin. Genetic structure of chloroplast and mitochondrial DNA polymorphisms in plants. En préparation.

	Espèce	Famille	Form	Txn	Sex	Poll	Graine	Rep	Référence	yr	gén.	Distribution	Méthode	Pop	Ind	Hapl	Ht	Gst	Nst	Rst
				G/A	M. D. G	A, E. Z	disp	S. O										%	%	
1	*Alnus glutinosa*	Betulaceae	T	A	M	A	W	O	Kling & Ferris	1998	cp	rangewide	PCR-RFLP	56	2.8	10	0.78	86.8	91.9	_
2	*Aquilegia chrysanta*	Renonculaceae	V	A	M	E	G		Strand et al.	1996	cp	SW-USA & Mex	PCR-DGGE	9	13.3	5	0.52	87.2	_	_
3	*Aquilegia longissima*	Renonculaceae	V	A	M	E	G		Strand et al.	1996	cp	SW-USA & Mex	PCR-RFLP	9	9.9	5	0.69	88.6	_	_
4	*Argania spinosa*	Sapotaceae	T	A	M	E	AI	O	El Moussadik & Petit	1996	cp	rangewide	PCR-RFLP	19	4.7	11	0.90	60.5	73.2	_
5	*Aucoumea klaineana*	Burseracea	T	A	M		W		Mukoko-Ntoutoume et al.	1999	cp	rangewide	PCR-RFLP	15	4.4	2	0.39	53.8	_	_
6	*Beta vulgaris*	Chenopodiaceae	VI	A	G	A	W		Forcioli et al.	1998	cp	F-UK	RFLP	39	4.7	8	0.82	39.4	41.1	_
7	*Castanea sativa*	Fagaceae	T	A	M	AE	AD	O	Fineschi et al. unpubl.	1999	cp	Medit.	PCR-RFLP	38	4.7	11	0.68	43.3	52.2	_
8	*Corallorhiza maculata*	Orchidaceae	V	A	M	E	W		Freudenstein & Doyle	1994	cp	rangewide	RFLP	6	2.7	7	0.90	26.7	37.7	_
9	*Coreopsis grandiflora*	Asteraceae	A	A	M	E	W	O	Mason-Gamer et al.	1995	cp	Geor. & Arkans.	RFLP	14	15.7	13	0.77	63.6	83.4	_
10	*Dicorynia guianensis*	Cesalpinaceae	T	A	M		W	OS	Pierroton unpubl	1999	cp	F Guyana	PCR-RFLP	12	11.5	9	0.69	41.0	25.0	_
11	*Dryas integrifolia*	Rosaceae	V	A	M	E	W		Tremblay MS Thesis	1999	cp	rangewide	PCR-RFLP	42	2.0	20	0.90	44.4	47.1	_
12	*Eucalyptus nitens*	Myrtaceae	T	A	M	Z	W	O	Byrne & Moran	1994	cp	rangewide	RFLP	8	10.0	13	0.95	61.9	77.7	_
13	*Fagus sylvatica*	Fagaceae	T	A	M	A	AI	O	Demesure et al.	1996	cp	rangewide	PCR-RFLP	74	4.8	9	0.24	83.1	90.2	_
14	*Gliricidia sepium*	Leguminosae	T	A	M	E	Expl	O	Lavin et al.	1991	cp	rangewide	RFLP	11	4.7	5	0.76	69.1	84.5	_
15	*Glycine soja*	Leguminosae	A	A	M	E			Powell et al.	1995	cp	Russia/E Asia	SSR	9	3.0	7	_	62.6	_	_
16	*Helianthus petiolaris*	Asteraceae	A	A	M	E	AT		Beckstrom-Sternberg et al.	1991	cp	SW-USA & Mex	RFLP	9	6.8	3	0.44	92.3	_	_
17	*Heuchera grossulariifolia*	Asteraceae	V	A	M	E			Wolf et al.	1990	cp	NW-USA	RFLP	15	6.0	5	0.74	100.0	100.0	_
18	*Heuchera micrantha*	Asteraceae	V	A	M	E			Soltis D. et al.	1989	cp	W-USA	RFLP	28	7.0	10	0.82	100.0	100.0	_
19	*Hordeum vulgare*	Poaceae	A	A	M	A	AD	S	Neale et al.	1986	cp	Israel-Iran	RFLP	21	8.7	3	0.69	74.7	_	_
20	*Ipomopsis aggregata cplx*	Polemoniaceae	V	A	M	EZ	N	O	Wolf et al.	1997	cp	W-USA	PCR-RFLP	18	10.0	11	0.95	100.0	100.0	_
21	*Liriodendron tulipifera*	Magnoliaceae	T	A	M	E	AD	O	Sewell et al.	1996	cp	rangewide	RFLP	10	2.4	2	0.36	100.0	100.0	_
22	*Lupinus texensis*	Leguminosae	A	A	M	E			Banks & Birky	1985	cp	rangewide	RFLP	15	5.4	3	0.15	85.6	95.9	_
23	*Melaleuca sp.*	Myrtaceae	T	A	M	E			Butcher et al.	1995	cp	rangewide	RFLP	12	8.0	7	0.86	90.6	99.0	_
24	*Nicotiana debneyi*	Solanaceae	A	A	M	E	AI		Scowcroft	1979	cp	rangewide?	RFLP	9	10.0	2	0.39	100.0	100.0	_
25	*Nothofagus nervosa*	Fagaceae	T	A	M	A		O	Marchelli et al.	1998	cp	Argentina	PCR-RFLP	11	8.8	2	0.44	100.0	100.0	_
26	*Pedicularis chamissonis*	Scrophulariaceae	V	A	M	E			Fujii et al.	1997	cp	rangewide	PCR-SSCP	19	2.2	10	0.87	90.9	98.7	_
27	*Penstemon centrantifolius*	Scrophulariaceae	V	A	M	EZ			Wolf & Elisens	1995	cp	California	RFLP	16	4.6	4	0.52	100.0	100.0	_
28	*Penstemon palmeri*	Scrophulariaceae	V	A	M	E			Wolf & Elisens	1995	cp	California	RFLP	7	5.4	3	0.76	100.0	100.0	_
29	*Penstemon spectabilis*	Scrophulariaceae	V	A	M	E			Wolf & Elisens	1995	cp	California	RFLP	10	5.2	2	0.36	100.0	100.0	_
30	*Phacelia dubia*	Hydrophyllaceae		A			N	O	Levy et al.	1996	cp	Virginia, S-Carol.	RFLP	12	7.6	13	0.87	61.1	68.7	_
31	*Quercus alba complex*	Fagaceae	T	A	M	A	AI	O	Whittemore & Schaal	1991	cp	E-USA	RFLP	11	8.7	7	0.73	87.0	92.2	_
32	*Quercus ilex*	Fagaceae	T	A	M	A	AI	O	Pierroton unpubl	1999	cp	Morocco	PCR-RFLP	82	5.7	2	0.31	82.5	_	_
33	*Quercus petraea*	Fagaceae	T	A	M	A	AI	O	Dumolin-Lapègue et al.	1999	cp	S-France	PCR-RFLP	51	3.8	6	0.64	82.9	_	_
34	*Quercus pubescens*	Fagaceae	T	A	M	A	AI	O	Dumolin-Lapègue et al.	1999	cp	S-France	PCR-RFLP	142	4.4	8	0.70	90.7	_	_
35	*Quercus robur*	Fagaceae	T	A	M	A	AI	O	Dumolin-Lapègue et al.	1999	cp	S-France	PCR-RFLP	142	4.0	9	0.70	78.2	_	_
36	*Quercus suber*	Fagaceae	T	A	M	A	AI	O	Pierroton unpubl	1999	cp	Morocco	PCR-RFLP	82	5.7	2	0.31	82.5	_	_
37	*Saxifraga oppositifolia*	Caryophillaceae		A					Abbot et al.	1995	cp	North EU	RFLP	_	_	5	low	_	_	_
38	*Silene acaulis*	Caryophillaceae	V	A					Abbot et al.	1995	cp	North EU	RFLP	_	_	2	low	_	_	_
39	*Silene alba*	Caryophillaceae	A	A	D	E	N	O	McCauley	1997	cp	local	PCR-RFLP	11	11.6	4	0.68	71.8	_	_
40	*Silene vulgaris*	Caryophillaceae	A	A	G	E	N	O	McCauley	1998	cp	local	PCR-RFLP	16	15.2	3	_	62.4	_	_
41	*Sorbus aria*	Rosaceae	T	A	M	E	AI	O	Demesure unpublished	1999	cp	France	PCR-RFLP	5	4.7	18	0.96	43.7	21.3	_
42	*Sorbus aucuparia*	Rosaceae	T	A	M	E	AI	O	RaspéPhD Thesis	1998	cp	rangewide	PCR-RFLP	6	24.0	12	0.87	30.7	29.9	_
43	*Sorbus torminalis*	Rosaceae	T	A	M	E	AI	O	Demesure unpublished	1999	cp	France	PCR-RFLP	44	11.7	23	0.57	44.1	45.3	_
44	*Tellima grandiflora*	Saxifragaceae	V	A	M	E			Soltis et al.	1991	cp	rangewide	RFLP	51	4.1	4	0.55	98.1	97.9	_
45	*Thymus vulgaris*	Labiatae	VL	A	G	E	N		Tarayre et al	1997	cp	local (F)	RFLP	17	5.1	8	0.63	43.6	36.1	_
46	*Tolmeia menziesii*	Saxifragaceae	V	A	M	E			Soltis D. et al.	1989	cp	rangewide	RFLP	21	6.0	4	0.59	100.0	100.0	_
47	*Tragopogon dubius*	Compositae	BA	A	M	E			Soltis & Soltis	1989	cp	centr. USA?	RFLP	11	6.0	2	0.55	100.0	100.0	_
48	*Vouacapoua americana*	Caesalpiniaceae	T	A	M		AD	OS	Dutech pers. comm.	1999	cp	rangewide	PCR-RFLP	10	5.7	5	0.86	88.4	88.5	_
	ADNcp angiospermes													**27.7**	**7.0**	**7.1**	**0.66**	**75.9**	**78.1**	
1	*Beta vulgaris*	Chenopodiaceae	VI	A	G	A	W		Cuguen et al. & Forcioli PhD	1994	mt	France	RFLP	35	4.9	11	0.80	43.3	41.2	_
2	*Daucus carota*	Umbelliferae	BA	A	G	E			Rontort et al.	1995	mt	local (Montpellier)	RFLP	7	7.1	13	_	7.8	_	_
3	*Fagus crenata*	Fagaceae	T	A	M	A	AI	O	Tomaru et al.	1998	mt	rangewide	PCR-RFLP	17	23.5	8	0.88	96.5	96.8	_
4	*Gliricidia sepium*	Leguminosae	T	A	M	E	Expl	O	Dawson et al.	1996	mt	Guatemala	RFLP	7	5	2	0.57	90.0	_	_
5	*Ipomopsis aggregata cplx*	Polemoniaceae	V	A	M	EZ	N	O	Wolf et al.	1997	mt	W-USA	PCR-RFLP	18	10	1	0.00	_	_	_
6	*Raphanus sativus*	Cruciferaceae	BA	A	G	E			Yamagishi & Terachi	1994	mt	Japan rangewide	PCR	24	4	2	0.49	54.8	_	_
7	*Thymus vulgaris*	Labiatae	VL	A	G	E	N		Tarayre PhD	1997	mt	F	RFLP	13	4.8	53	1.00	23.4	_	_
	ADNmt angiospermes													***17.3***	***8.5***	***12.9***	***0.82***	***52.8***		
1	*Abies alba*	Pinaceae	T	G	M	A	W	O	Vendramin et al. sub.	1999	cp	rangewide	SSR	17	23.8	90	0.98	13.3	15.9	18.1
2	*Abies mariesii*	Pinaceae	T	G	M	A	W	O	Tsumura et al.	1994	cp	rangewide	RFLP	7	22.9	2	0.50	7.3	_	_
3	*Picea abies*	Pinaceae	T	G	M	A	W	O	Vendramin's group	1999	cp	rangewide	SSR	106	10.6	47	0.68	8.9	9.4	6.3
4	*Picea glauca*	Pinaceae	T	G	M	A	W	O	Furnier & Stine	1995	cp	rangewide	RFLP	22	16.9	2	0.48	14.0	_	_
5	*Pinus banksiana*	Pinaceae	T	G	M	A	W	O	Dong & Wagner	1994	cp	rangewide	RFLP	8	47.5	8	0.43	1.6	_	_
6	*Pinus brutia*	Pinaceae	T	G	M	A	W	O	Bucci et al.	1998	cp	rangewide	SSR	8	8.8	40	1.00	12.4	27.1	34.8
7	*Pinus contorta*	Pinaceae	T	G	M	A	W	O	Dong & Wagner	1994	cp	rangewide	RFLP	8	45.6	6	0.44	3.3	_	_
8	*Pinus densata*	Pinaceae	T	G	M	A	W	O	Wang X-R & Szmidt	1994	cp	rangewide	RFLP	7	20.3	3	0.62	16.5	27.6	_
9	*Pinus flexilis*	Pinaceae	T	G	M	A	W	O	Latta & Mitton	1997	cp	local	PCR-RFLP	7	6.0	41	0.71	1.6	_	_
10	*Pinus halepensis*	Pinaceae	T	G	M	A	W	O	Bucci et al	1998	cp	rangewide	SSR	10	12.1	28	0.64	28.7	35.2	27.9
11	*Pinus lambertiana*	Pinaceae	T	G	M	A	W	O	Vendramin's group	1999	cp		SSR	8	21.0	48	0.94	14.3	59.2	84.0
12	*Pinus leucodermis*	Pinaceae	T	G	M	A	W	O	Vendramin's group	1999	cp	rangewide	SSR	7	23.5	48	0.94	5.4	10.0	13.0
13	*Pinus monticola*	Pinaceae	T	G	M	A	W	O	White	1990	cp	Br. Columbia	RFLP	16	8.9	2	0.23	9.8	_	_
14	*Pinus mugo*	Pinaceae	T	G	M	A	W	O	Vendramin's group	1998	cp	rangewide	SSR	19	25.3	80	0.96	5.2	9.6	12.2
15	*Pinus muricata*	Pinaceae	T	G	M	A	W	O	Hong et al	1993	cp	rangewide	RFLP	10	10.0	3	0.69	100.0	100.0	_
16	*Pinus pinaster*	Pinaceae	T	G	M	A	W	O	Vendramin et al.	1998	cp	rangewide	SSR	10	24.0	34	0.82	10.1	14.6	19.1
17	*Pinus radiata*	Pinaceae	T	G	M	A	W	O	Hong et al.	1993	cp	rangewide	RFLP	5	23.9	3	0.61	7.3	_	_
18	*Pinus resinosa*	Pinaceae	T	G	M	A	W	O	Echt et al.	1998	cp	rangewide	SSR	7	22.6	25	0.64	9.5	9.5	11.0
19	*Pinus sylvestris*	Pinaceae	T	G	M	A	W	O	Vendramin's group	1999	cp	rangewide	SSR	23	31.0	156	0.98	1.7	1.8	1.8
20	*Pseudotsuga menziesii*	Pinaceae	T	G	M	A	W	O	Ponoy et al.	1994	cp	Br. Columbia	RFLP	18	4.0	16	0.79	8.5	9.0	_
21	*Tsuga canadensis*	Cupressaceae	T	G	M	A	W	O	Wang C. et al.	1997	cp	rangewide	RFLP	11	5.4	6	0.56	19.8	20.0	_
	ADNcp Conifères													**15.9**	**19.7**	**32.8**	**0.71**	**14.2**	**24.9**	**22.8**
1	*Abies firma*	Pinaceae	T	G	M	A	W	O	Tsumura et al.	1998	mt	rangewide	RFLP	7	13.9	7	0.77	86.4	88.3	_
2	*Abies homolepis*	Pinaceae	T	G	M	A	W	O	Tsumura et al.	1998	mt	rangewide	RFLP	8	15.9	7	0.59	47.1	48.8	_
3	*Abies sachalinensis*	Pinaceae	T	G	M	A	W	O	Tsumura et al.	1998	mt	rangewide	RFLP	5	28.3	8	0.30	22.8	_	_
4	*Abies veitchi*	Pinaceae	T	G	M	A	W	O	Tsumura et al.	1998	mt	rangewide	RFLP	12	19.3	4	0.03	_	_	_
5	*Picea abies*	Pinaceae	T	G	M	A	W	O	Sperisen's group	1999	mt	rangewide	PCR-RFLP							
6	*Pinus contorta*	Pinaceae	T	G	M	A	W	O	Dong & Wagner	1993	mt	rangewide	RFLP	8	45.4	5	0.74	72.0	_	_
7	*Pinus flexilis*	Pinaceae	T	G	M	A	W	O	Latta & Mitton	1997	mt	local	PCR-RFLP	7	39.5	3	0.32	68.8	_	_
8	*Pinus muricata*	Pinaceae	T	G	M	A	W	O	Strauss, Hong & Hipkins	1993	mt	rangewide	RFLP	9	7.69	7	0.80	95.8	_	_
9	*Pinus radiata*	Pinaceae	T	G	M	A	W	O	Strauss et al.	1993	mt	California	RFLP	5	8.8	3	0.45	74.5	_	_
10	*Pseudotsuga menziesii*	Pinaceae	T	G	M	A	W	O	Aagaard et al.	1998	mt	rangewide	mt RAPD	6	30.2	14	0.86	74.0	_	_
	ADNmt Conifères													**7.4**	**23.2**	**8.2**	**0.54**	**67.7**		
	Toutes études													**21.7**	**12.1**	**13.8**	**0.65**	**57.4**	**62.8**	**22.8**

Table 1. Etudes de diversité des génomes cytoplasmiques chez les plantes.

être intégrées dans ces compilations. L'intérêt des méthodes d'analyse décrites dans le premier chapitre est également mis en évidence.

Evolution des techniques

Une recherche de références sur la structuration géographique des génomes cytoplasmiques chez les plantes a été effectuée, depuis l'avènement de ces recherches (1979) jusqu'à fin 1998 ; des études non publiées qui m'ont été communiquées sont également inclues. Elles sont alors regroupées dans l'année 1999. Un total de 97 études ('entrées') a été identifié. Le nombre d'espèces est de 76, car certaines espèces ont été étudiées à plusieurs reprises. Le nombre de publications réel est également inférieur au nombre d''entrées', car certaines publications incluent plusieurs espèces ou portent sur les deux génomes cytoplasmiques. La figure 16 illustre l'évolution du nombre d'études sur la structuration de la diversité cytoplasmique chez les plantes, par technique moléculaire ; trois techniques ont été distinguées : (1) la RFLP avec sondes marquées (la technique de RFLP directe à partir de chloroplastes purifiés n'a été utilisée qu'une seule fois (Scowcroft 1979) et n'a donc pas été distinguée) ; (2) les techniques dérivées de la PCR (principalement PCR-RFLP), à l'exclusion de l'étude des microsatellites ; (3) l'étude des microsatellites chloroplastiques.

On constate que les techniques dérivées de la PCR sont en train de se substituer à celles basées sur les sondes moléculaires. Dans certains cas toutefois, ces techniques restent encore difficilement remplaçables. Le développement des microsatellites a été limité jusqu'à présent à l'étude du chloroplaste, notamment chez les conifères, du fait de la disponibilité de la séquence complète de l'ADN chloroplastique d'un pin japonais (Wakasugi et coll. 1994). La diversité révélée est souvent très importante dans ces études. Les mutations identifiées au niveau de ces microsatellites (variants de longueur) obéissant plutôt à un modèle de mutation de type SMM[39]. Les mutations détectées par RFLP obéissant plutôt à un modèle IAM[40], les deux techniques semblent assez complémentaires (études de diversité *versus* études phylogéographiques par exemple). Le développement du séquençage systématique des individus dans les études de diversité, outre son coût encore élevé, pose le problème de la rareté des mutations très dispersées pour ces génomes cytoplasmiques, chez les plantes. Des techniques plus ciblées de détection de mutation semblent plus efficaces.

Diversité et différenciation

Un premier bilan des diversités nucléotidiques déduites des études RFLP avait été réalisé par Soltis et coll. (1992). Vu la diversité des techniques et du mode de présentation des données, je n'ai pas tenté de comparer les niveaux de diversité entre espèces. En effet, on peut constater immédiatement, dans notre compilation, que la technique utilisée a un effet majeur sur le nombre d'haplotypes révélés (table 1). Les études basées sur les microsatellites chloroplastiques (cpSSR) ont permis d'identifier souvent plusieurs dizaines d'haplotypes, et jusqu'à 156 chez le pin sylvestre, contre une dizaine environ pour les études par PCR-RFLP.

[39] SMM : *Stepwise Mutation Model.* Modèle de mutation de séquences répétées où les nouveaux allèles ne diffèrent que d'une répétition (en plus ou en moins) des allèles dont ils sont issus.

[40] IAM : *Infinite Allele Model.* Modèle de mutation où les nouveaux allèles diffèrent de tous les allèles préexistants.

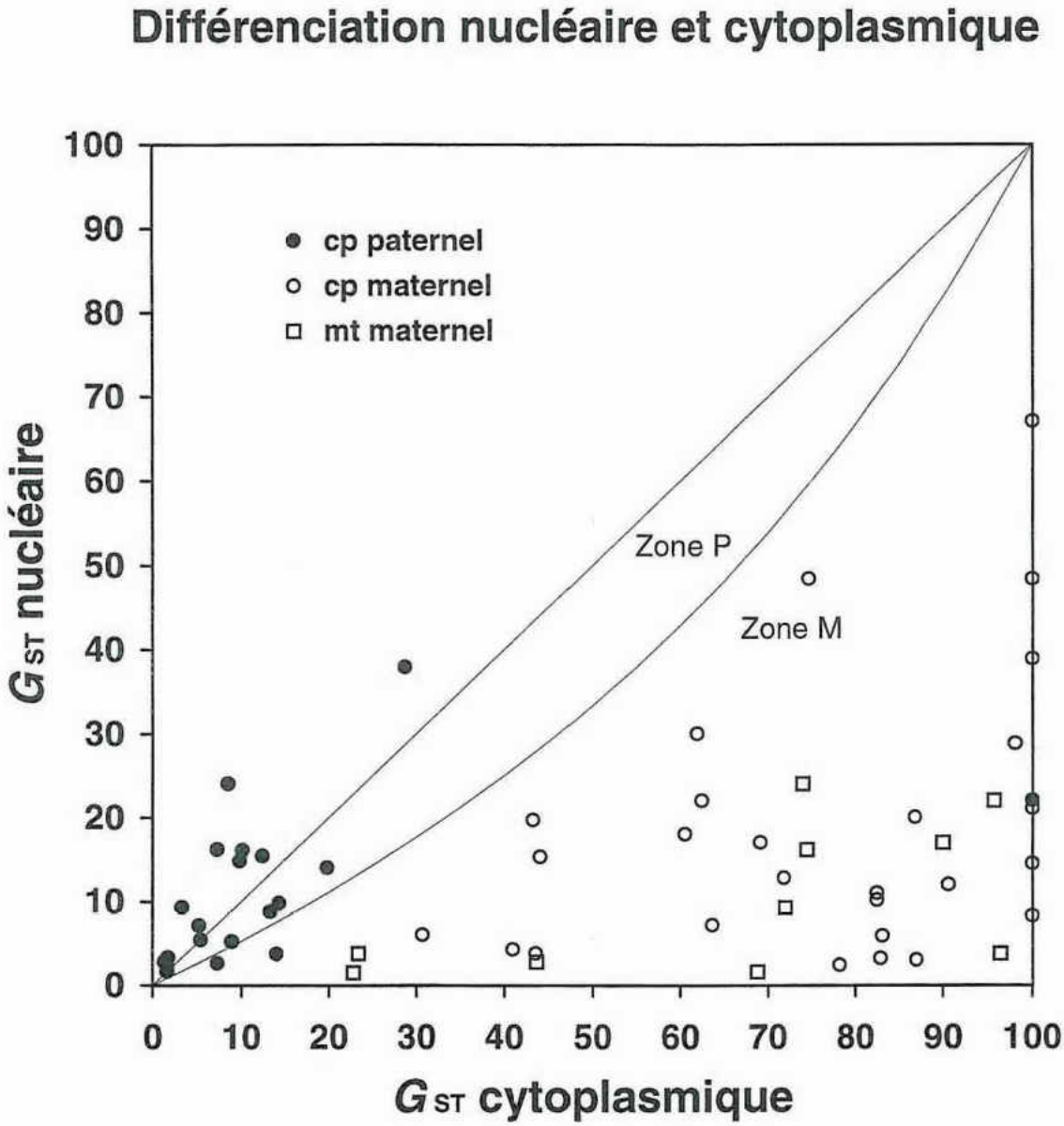


Figure 17 : Relation G_{ST} nucléaire en fonction du G_{ST} cytoplasmique chez les plantes, selon le type de génome (chloroplaste ou mitochondrie) et le mode d'hérédité (paternel ou maternel). Les domaines à l'équilibre sont situés entre les courbes (zone P pour les cas d'hérédité paternelle, zone M en cas d'hérédité maternelle). On note la plus forte corrélation entre le G_{ST} nucléaire et le G_{ST} cytoplasmique en cas d'hérédité paternelle, et le grand nombre d'espèces situées en dehors du domaine d'équilibre.

La comparaison des coefficients de différentiation est plus informative et constituait le principal objectif de cette synthèse. Sur 46 études concernant la variabilité de l'ADNcp chez les angiospermes (hérédité présumée maternelle), on trouve un G_{ST} moyen de 76%. Pour l'ADNmt (6 études seulement dont trois espèces gynodioïques), on trouve 53%. Chez les gymnospermes (il s'agit presque exclusivement de pinacées), la valeur moyenne pour l'ADNcp est de 14% (sur 21 études), contre 68% pour l'ADNmt (8 études). Les chloroplastes étant généralement transmis paternellement chez les conifères, comme on l'a vu plus haut, c'est bien le mode de transmission qui affecte le plus le coefficient de différentiation.

Ces valeurs peuvent être comparées à celles trouvées dans la synthèse bibliographique sur la diversité des allozymes chez les arbres (marqueurs nucléaires) (chapitre 1). Pour les génomes cytoplasmiques, les données ne concernent pas exclusivement les arbres : 29 des 86 études portaient sur d'autres plantes (annuelles ou vivaces). Toutefois, le type biologique (arbre, plante vivace ou annuelle), ou l'appartenance taxonomique (angiosperme ou gymnosperme) joue peu pour les génomes maternels : le G_{ST} est dans tous les cas proche de 0,70 (0,68 pour les gymnospermes, 0,70 pour les annuelles, 0,75 pour les plantes vivaces (herbacées ou ligneuses, sauf arbres), 0,73 pour les arbres angiospermes). Pour les isozymes, j'avais trouvé un G_{ST} moyen de 0,10 (voir table 1 au chapitre 1), beaucoup plus proche de la valeur trouvée pour l'ADNcp des conifères (0,14).

Pour affiner cette comparaison des niveaux de différentiation entre génomes, nous avons recherché, pour toutes les espèces dont la diversité chloroplastique ou mitochondriale avait été étudiée, des références sur la diversité nucléaire (isozymes). La corrélation entre G_{ST} cytoplasmique et nucléaire (G_{STc} et G_{STn}) a été calculée séparément pour les génomes cytoplasmiques transmis maternellement et paternellement. Les résultats sont présentés sous forme de diagramme ($G_{STn} = f(G_{STc})$) (figure 17). Pour l'ADNcp des gymnospermes (transmission paternelle), la corrélation entre G_{STc} et G_{STn} est de 0,71 ($R^2 = 0,53$). Pour les génomes transmis maternellement, on a une corrélation de 0,41 seulement entre G_{STc} et G_{STn} ($R^2 = 0,17$). Un tel résultat peut s'expliquer par le fait qu'un gène transmis paternellement migrera à la fois par pollen et par graine, alors qu'un gène transmis maternellement ne migrera que par graine. Or un gène nucléaire est transmis à la fois par graines et par pollen, comme un gène transmis par voie mâle (mais pas dans les mêmes proportions toutefois).

La diagonale permet de comparer immédiatement les coefficients de différentiation ($G_{STn} = G_{STc}$). La zone située entre les deux courbes (zone P) précise l'espace des valeurs possibles en cas d'hérédité strictement paternelle du génome cytoplasmique, selon les valeurs des flux par pollen et par graine, à **l'équilibre dans un modèle en île** (Petit 1992, Petit et coll. 1993). Si les hypothèses du modèle sont vérifiées, il ne devrait pas y avoir de points au dessus de la diagonale. On constate que ce n'est pas le cas : il y a même légèrement plus de valeurs au dessus de la diagonale qu'en dessous. Comment rendre compte de cette observation ? Notons que pour un gène transmis paternellement, comparé à un gène situé à un locus diploïde nucléaire, la dérive sera plus forte mais le taux de migration aussi sera plus fort. En effet, ces gènes sont systématiquement véhiculés par pollen, alors qu'un gène nucléaire ne le sera qu'une fois sur deux. J'avais montré au cours de ma

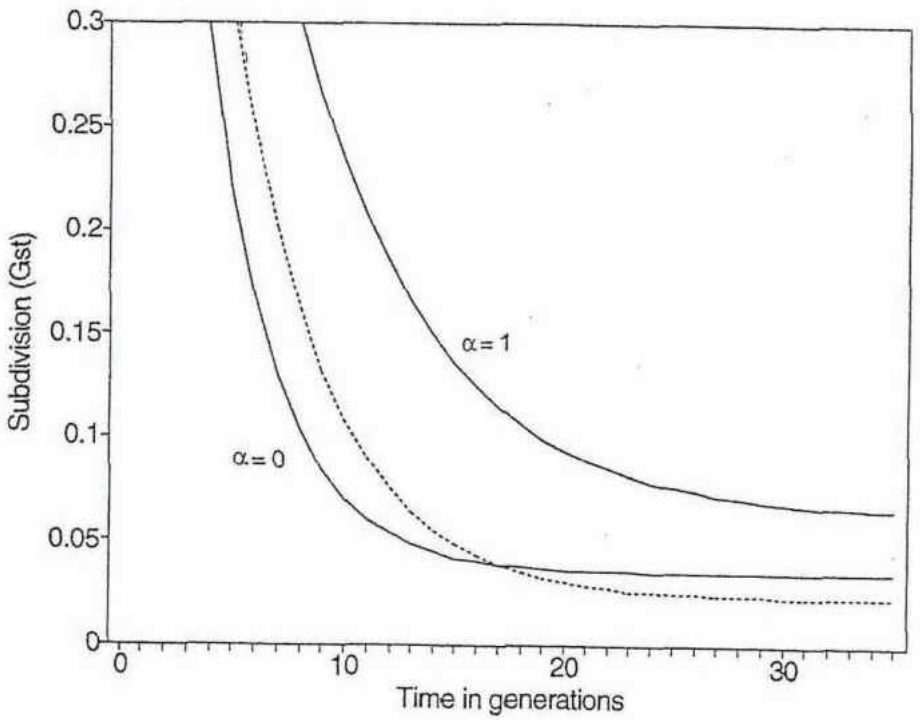


Figure 18 : Effet du type d'hérédité cytoplasmique sur l'approche de l'équilibre du G_{ST}, et comparaison avec un gène nucléaire, dans le cas où le taux de migration par pollen est égal à celui par graine ($m_P = m_S$). Le taux de transmission maternelle est α ($\alpha = 1$: transmission strictement maternelle, $\alpha = 0$: transmission strictement paternelle). En pointillé : gène nucléaire. Quand $\alpha = 1$, on a toujours $G_{STc} > G_{STn}$. Par contre, quand $\alpha = 0$, on a initialement $G_{STn} > G_{STc}$, mais à l'équilibre c'est le contraire, la différenciation cytoplasmique est plus élevée. L'approche de l'équilibre est en effet plus rapide pour le gène cytoplasmique (effet prédominant des flux de gènes par pollen), tandis qu'à l'équilibre l'effet qui prédomine est celui de la dérive.

thèse que la résultante de ces deux forces antagonistes (la dérive qui augmente la différenciation, les flux qui la diminuent) était, à l'équilibre, dominée par la dérive : le G_{STc} est supérieur ou égal au G_{STn} à l'équilibre. Par contre, la dynamique d'approche de l'équilibre est plutôt contrôlée par les effets des flux de gènes (plus forts pour l'ADNcp transmis paternellement). Si un goulot d'étranglement récent a augmenté la différenciation, le G_{STc} (paternel) redescendra moins rapidement que le G_{STn}. Si l'équilibre n'est pas encore atteint, on aura alors $G_{STc} < G_{STn}$ et non l'inverse (figure 18).

Pour un génome transmis maternellement, l'espace correspondant est situé sous la courbe du bas (zone M). Ici, toutes les valeurs sont situées dans la zone attendue, ce qui ne signifie pas toutefois qu'un équilibre existe effectivement. Le rapport des flux par pollen sur les flux par graine est donné, toujours dans un modèle à l'équilibre, par la formule :

$$r = \frac{2(1/G_{STc} - 1) - (1/G_{STn} - 1)}{\beta(1/G_{STn} - 1) - (1/G_{STc} - 1)},$$

où β est le taux de transmission du génome cytoplasmique par la voie paternelle (Petit 1992). Je n'applique ici la formule, à titre d'illustration, qu'aux génomes cytoplasmiques transmis maternellement ($\beta = 0$), car l'absence d'équilibre patente en cas de transmission paternelle invalide cette procédure. Sur un total de 31 espèces, on trouve en moyenne $r = m_p/m_s = 65$ (amplitude : 1,1 à 696). Les deux espèces avec les plus faibles valeurs sont l'orge et le chataigner. L'orge est la seule espèce autogame de l'étude ; tous les flux ou presque devraient donc être le fait des mouvements de graines. Or on trouve des flux par graine et par pollen équivalents ($r = 1,1$), indiquant que les rares cas d'allofécondation jouent peut être un rôle disproportionné dans les flux. Le cas du chataigner, une fagacée, comme le chêne ou le hêtre, mais dont le pollen est disséminé surtout par les insectes et dont les fruits ont été largement disséminés par l'homme, et qui de plus a été abondamment propagée de manière asexuée (par greffage), est particulier. Les autres espèces pour lesquelles les flux par graine sont relativement importants ($r < 5$) sont principalement des arbres entomophiles caractérisés par des mécanismes de dispersion efficaces des graines (*Eucalyptus nitens*, *Sorbus aucuparia*, *Argania spinosa*).

A l'autre extrême ($r > 100$), on trouve notamment des fagacées (hêtres et chênes), dont le pollen est disséminé par le vent sur de grandes distances.

Structuration phylogéographique : prise en compte des distances entre haplotypes

Dans le chapitre 1, le coefficient de différentiation N_{ST} avait été défini et comparé au G_{ST} dont il dérive. Dans toutes les études de variabilité cytoplasmique où c'était possible (nombre suffisant d'haplotypes, et caractérisation assez précise de ces haplotypes disponible dans la publication), le N_{ST} a été calculé et comparé par un test (Pons et Petit 1996) au G_{ST}. Les résultats sont présentés sur la figure 19 ($N_{ST} = f(G_{ST})$). On constate que, quel que soit le mode d'hérédité (paternel ou maternel), le N_{ST} est presque toujours supérieur au G_{ST} (24 cas contre 3 où l'inverse est vrai). En outre, si on ne tient compte que des différences significatives, on a 16 cas où N_{ST} est significativement supérieur au G_{ST}, et aucun cas où l'inverse est vrai, sur un total de 39 cas où les deux paramètres avaient été estimés (en excluant les cas où $G_{ST} = N_{ST} = 1$).

Que peut-on en conclure sur la structuration génétique ? La mutation crée de nouveaux variants, initialement localisés

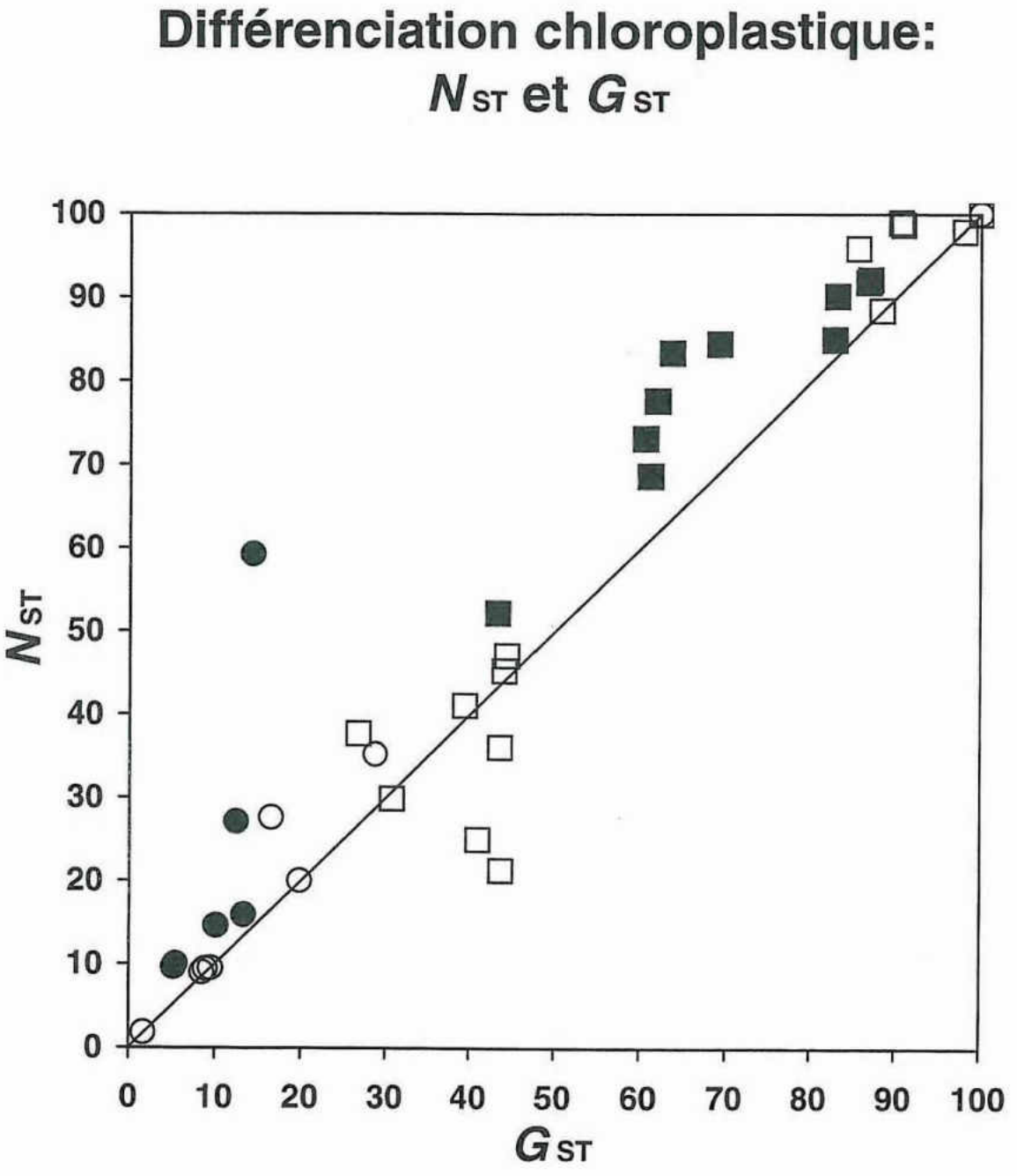


Figure 19 : Relation entre N_{ST} et G_{ST} cytoplasmique chez les plantes, selon le mode d'hérédité (paternel (ronds) ou maternel (carrés)). On note que le N_{ST} est généralement plus grand que le G_{ST} (valeurs significatives en noir).

dans les mêmes régions que les variants dont ils dérivent, entraînant l'apparition d'une structure phylogéographique. Aux échelles considérées (fréquemment l'aire entière de l'espèce), la redistribution des nouveaux variants par flux de gènes (par graines où pollen) n'a généralement pas le temps de faire disparaître cette structuration. L'histoire des espèces joue donc certainement un rôle prédominant, et la phylogéographie intraspécifique, en prenant en compte l'histoire des variants, devrait permettre de mieux retracer cette histoire.

Le chapitre suivant sera donc entièrement consacré à l'influence de l'histoire des populations sur leur structuration génétique.

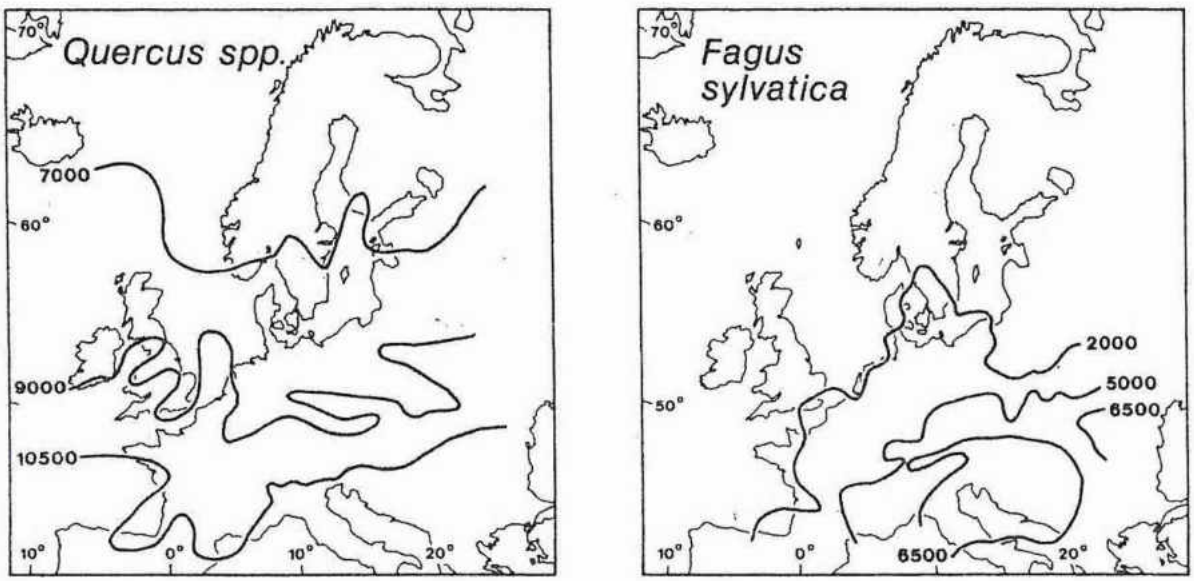


Figure 1. Colonisation postglaciaire du chêne et du hêtre en Europe, d'après Huntley et Birks (1983).

Chapitre 3

Contribution des marqueurs génétiques à la connaissance de l'histoire des populations d'arbres

« Now that it has been accepted that the ice-ages are expressions of perpetual variations in the Earth's orbit, it is clear that we can use the data of Quaternary research to illuminate processes in the history of life on time-scales between those of ecology and palaeontology. » Bennett, 1997.

INTRODUCTION

Les chapitres précédents ont démontré que les données actuellement recueillies par les généticiens sur la diversité et la structuration des populations étaient devenues non seulement plus abondantes mais surtout qualitativement plus informatives. En effet, les séquences ou les haplotypes, contrairement aux allèles détectés avec la technique des isozymes, contiennent une information sur leur propre histoire. La phylogéographie intraspécifique (Avise et coll. 1979) a pu se développer, chez les animaux puis beaucoup plus tard chez les plantes (chapitre 2), remettant la biogéographie historique au cœur des études de variabilité génétique des populations.

En outre, de nouveaux outils mathématiques adaptés à l'analyse de telles données, ainsi qu'une nouvelle génération de modèles théoriques (notamment ceux issus de la théorie de la coalescence) ont contribué à renforcer le lien entre génétique des populations et histoire, qui s'était relâché quand les généticiens se focalisaient sur des populations immortelles, à l'équilibre.

Une prise de conscience assez générale de l'importance de l'histoire des populations pour interpréter les niveaux de diversité et la structuration des espèces est palpable. C'est particulièrement vrai pour les études basées sur les marqueurs neutres, y compris ceux qui ne se prêtent pas à une analyse phylogénétique comme les isozymes. Par contre, l'importance relative de processus historiques sur le niveau et la structuration des caractères adaptatifs reste difficile à apprécier à ce jour.

Dans ce chapitre, je vais présenter quelques résultats, obtenus grâce à l'étude de marqueurs moléculaires, qui ont permis d'améliorer les connaissances sur l'histoire des populations d'arbres forestiers. A l'échelle de l'ensemble de l'aire, les marqueurs ont parfois permis de confirmer, voire d'identifier, l'emplacement des zones refuges et des voies de recolonisation postglaciaires (figure 1). A l'échelle régionale, les données moléculaires ont permis de préciser certains des mécanismes de recolonisation, et de quantifier leurs conséquences sur la diversité génétique des populations. Enfin, dans les forêts européennes, l'histoire locale est largement dominée par les effets anthropiques. Toutefois, très peu de résultats ont été obtenus à ce jour à cette dernière échelle, et je complèterai donc cette présentation en

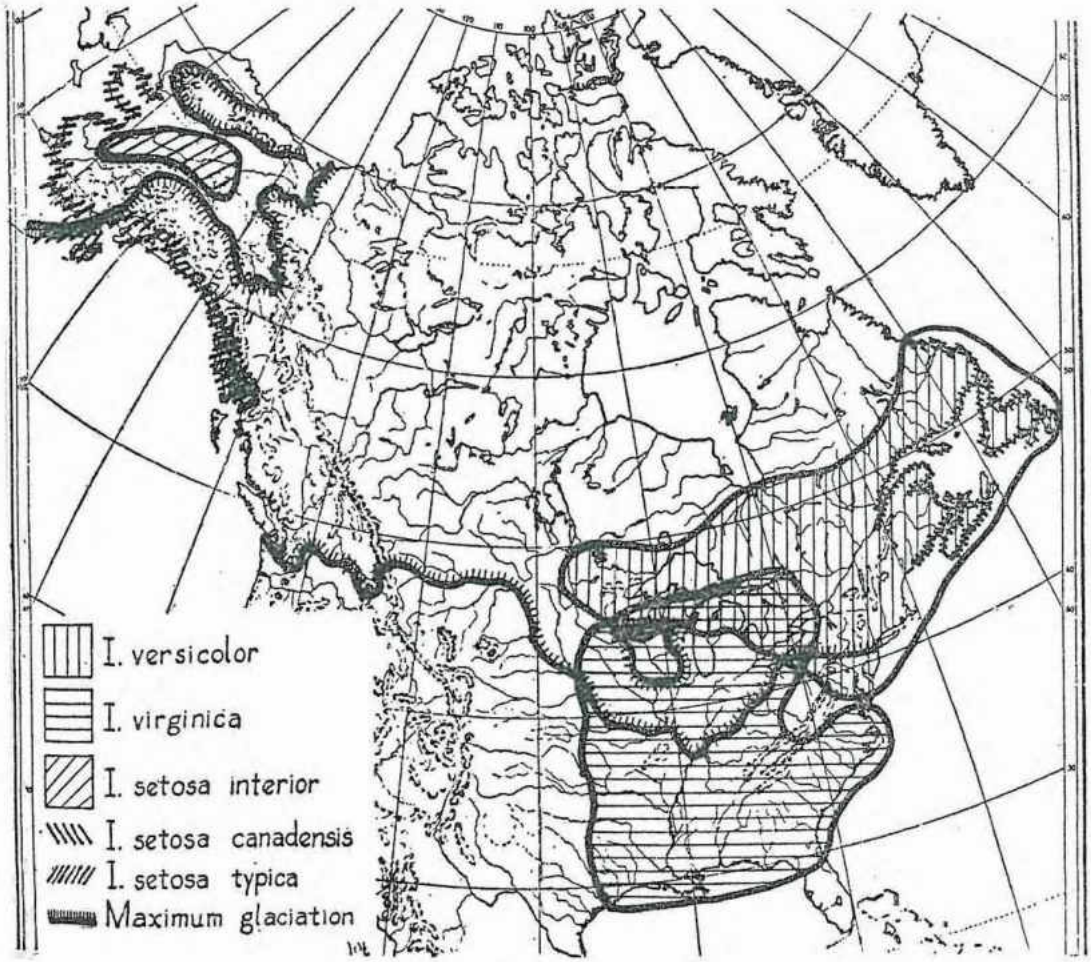


Figure 2. Distribution géographique de *Iris versicolor*, espèce polyploïde issue de l'hybridation de *I. virginica* et *I. setosa*, et relations avec la dernière glaciation. L'hybride polyploïde occupe des territoires qui étaient entièrement recouverts par la glace au maximum glaciaire. Il s'agirait donc d'une espèce récente. Une des espèces parentes, maintenant localisée en Alaska, aurait donc eu une aire de distribution plus étendue vers le nord-est, en sympatrie avec l'autre parent. D'après Anderson 1936, in Stebbins 1950.

décrivant des projets de recherche qui permettraient d'évaluer l'importance de l'utilisation et de la gestion des forêts sur la structuration génétique des espèces qui les composent. A toutes ces échelles, la paléogénétique (l'étude de l'ADN ancien) pourra également contribuer à tester certaines hypothèses et à élucider la dynamique spatio-temporelle des populations d'arbres forestiers.

LA RECOLONISATION POSTGLACIAIRE : LES REFUGES ET LES VOIES DE COLONISATION

« Nous supposerons l'apparition d'une nouvelle époque glaciaire commençant lentement, puis disparaissant, comme cela a eu lieu autrefois. A mesure que le froid augmentera, les zones les plus méridionales deviendront plus propres à recevoir les habitants du Nord, ceux-ci s'y porteront et remplaceront les formes des régions tempérées qui s'y trouvaient auparavant. Ces dernières, à leur tour, et pour la même raison, descendront de plus en plus vers le Sud, à moins qu'elles ne soient arrêtées par quelque obstacle, auquel cas elles périront. » Charles Darwin, L'origine des espèces, 6ème édition, traduction d'E. Barbier, 1878.

Les arbres comme modèles pour l'étude des conséquences génétiques de l'histoire postglaciaire

En Europe, la dernière glaciation a commencé il y a 115 000 ans et a culminé vers 18 000 BP (*Before Present* = 1950), pour s'achever vers 13 000 BP. D'une manière générale, dans la zone tempérée, l'histoire récente des arbres forestiers anémophiles, qui produisent donc un abondant pollen disséminé par le vent, est remarquablement bien connue, car la paléopalynologie, combinée à l'étude des macrorestes fossiles, permet d'identifier la présence des espèces au cours du temps dans les sédiments des lacs ou des tourbières.

Peu d'organismes bénéficient d'études aussi détaillées (on peut toutefois citer certains insectes (coléoptères par exemple) ou des rongeurs, mais il n'en reste pas moins que, selon Bennett (1997, p. 193) : « *...the only situation where there is any data on complete distribution movements over complete glacial-interglacial oscillations [is represented by] trees in Europe* ». Cela donne donc une certaine responsabilité aux généticiens des arbres qui disposent finalement d'espèces avec lesquelles ils peuvent tenter de comprendre de manière assez exceptionnelle les relations entre variabilité génétique et histoire.

Génétique et histoire postglaciaire

Cwynar et MacDonald (1987) ont ouvert la voie dans ce domaine en soulignant pour la première fois l'intérêt de combiner des études de variabilité génétique à l'aide de marqueurs à des travaux de paléobotanique, chez le pin contorta en Amérique du Nord.

En remontant plus loin, on peut même considérer que l'étude comparée de la distribution d'espèces allopolyploïdes et de leurs ancêtres diploïdes représentait déjà un effort pour intégrer génétique, évolution et histoire climatique récente (le cas d'Iris d'Amérique du Nord, étudiés par Stebbins (1950), est représenté dans la figure 2).

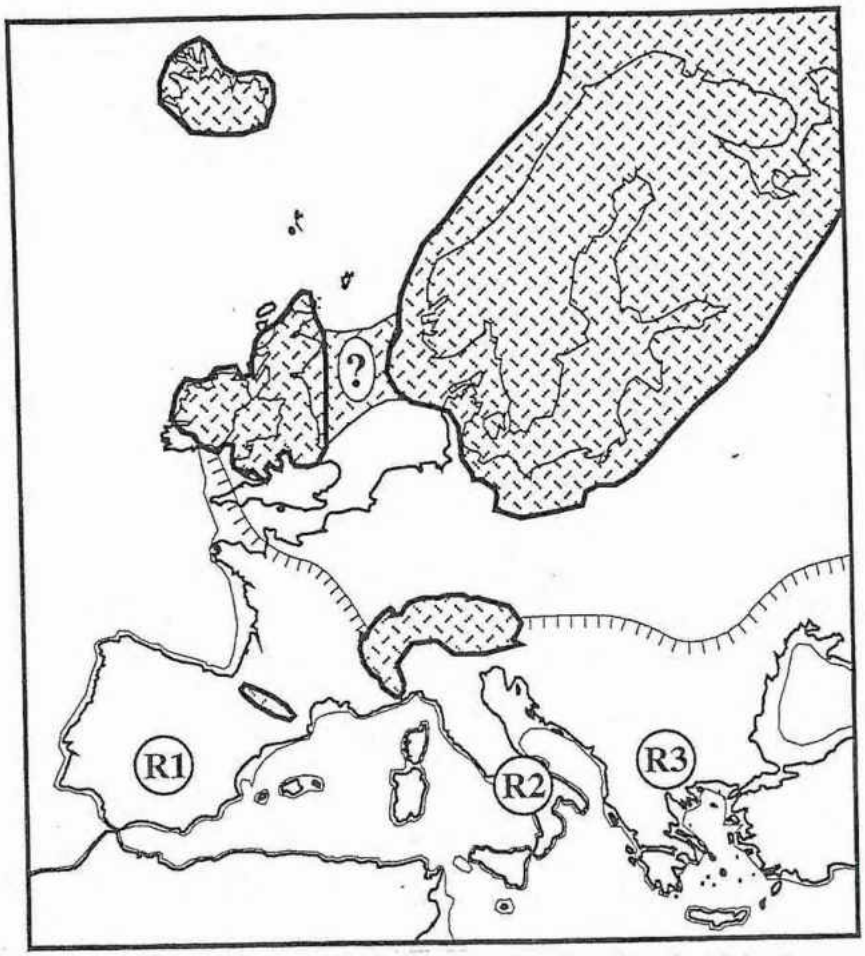


Figure 3. L'Europe lors du maximum glaciaire (18000 BP), d'après Taberlet et coll. (1998). La limite sud de la zone de permafrost est indiquée par une ligne en pointillé. Le niveau de la mer, plus bas que l'actuel, est précisé par une ligne plus fine.

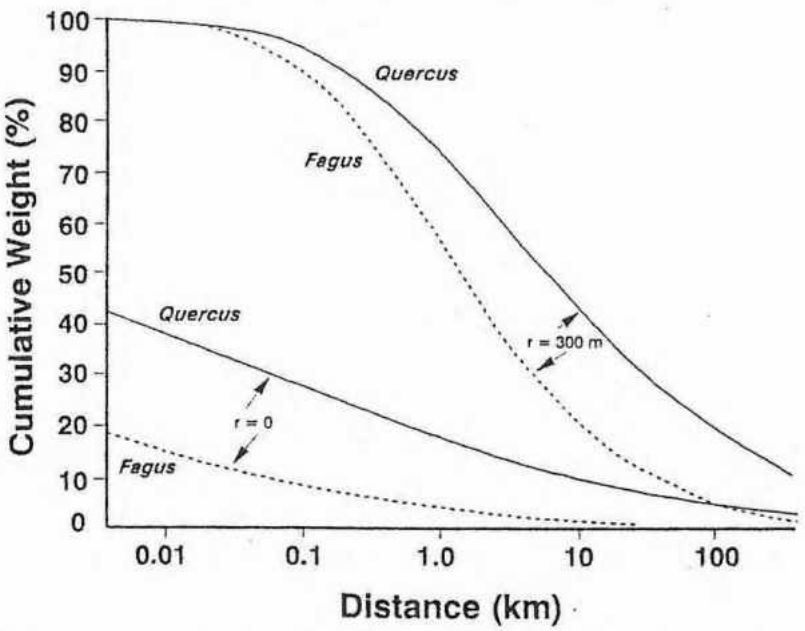


Figure 4. Proportion du pollen de chêne (*Quercus*) ou de hêtre (*Fagus*) s'accumulant au centre d'un lac de 300 mètres de rayon ($r = 300$), ou dans une tourbière ($r = 0$), en fonction de son origine (distance en kilomètres). D'après MacDonald (1993).

Variabilité de l'ADN chloroplastique des chênes en Europe et recolonisation postglaciaire

Les pollens des différentes espèces de chênes étudiés (notamment les chênes pédonculé, sessile et pubescent), sont pratiquement indistinguables par les palynologistes. En outre, ces espèces partagent presque systématiquement les mêmes haplotypes chloroplastiques en un lieu donné. Dans cette partie, on discutera donc des chênes blancs comme s'il s'agissait d'une seule espèce, pour revenir ultérieurement sur cet aspect de la question.

Données paléobotaniques

On a montré que l'hérédité des chloroplastes (et des mitochondries) était maternelle chez les chênes (chapitre 2). Or la recolonisation postglaciaire de territoires initialement recouverts par des glaciers ou des steppes froides (figure 3) fut nécessairement effectuée par l'intermédiaire de graines (en l'absence d'adaptations pour la dispersion à l'aide d'organes végétatifs chez les chênes). L'utilisation du pollen par les paléobotanistes pour reconstituer cette même colonisation est plus problématique, celui-ci, selon les taxons considérés, pouvant être transporté sur d'assez longues distances, donc éventuellement dans des régions où l'espèce ne pousse pas (figure 4). Pour chaque espèce, toutefois, des seuils à partir desquels la présence locale de l'espèce est vraisemblable ont été calculés (MacDonald 1993). La présence de macrorestes (débris végétaux divers dont l'identification taxonomique est parfois possible) démontre de manière plus nette la présence d'une espèce en un lieu et à une époque donnés. Une première synthèse des données palynologiques fossiles en Europe, destinée à préciser la répartition des espèces d'arbres (à l'aide cartes ***isochrones***, résumant l'abondance de pollen fossiles dans les sédiments) a été publié en 1983 (Huntley et Birks). Une nouvelle synthèse beaucoup mieux datée et plus complète est en cours, grâce à la base de données polliniques européenne (*European Pollen Database*, EPD : S. Brewer, R. Cheddadi et J-L de Beaulieu, comm. pers.).

L'emplacement des zones refuges pour les chênes se situe au niveau des péninsules ibérique, italienne, et balkanique, pour ce qui concerne l'Europe. La recolonisation fut précoce, des mouvements vers le nord sont sensibles dans les trois péninsules dès 12 000 à 11 000 BP. En conséquence, seuls les sondages correspondant aux séquences sédimentaires les plus longues peuvent être utilisés, or ceux-ci sont moins fréquents que ceux démarrant plus récemment. Il est important de noter que la résolution spatiale de ces reconstructions ne peut être augmentée indéfiniment, elle dépend des caractéristiques géomorphologiques des zones concernées (distribution des lacs ou des tourbières notamment). Jusqu'ici, la faible densité des points disponibles et l'hétérogénéité de leur distribution n'a pas permis d'utiliser des méthodes de cartographie quantitative : les cartes ont été tracées à la main, et des interprétations subjectives des données ne sont pas exclues (MacDonald 1993). Ceci est en train de changer, et des cartes plus fiables devraient voir le jour prochainement.

Histoire postglaciaire des chênes et ADN chloroplastique

Le premier travail portait sur 107 populations, principalement françaises et allemandes (Petit et coll. 1993[41]).

[41] Petit R J, A Kremer et D B Wagner. 1993. Geographic structure of chloroplast DNA

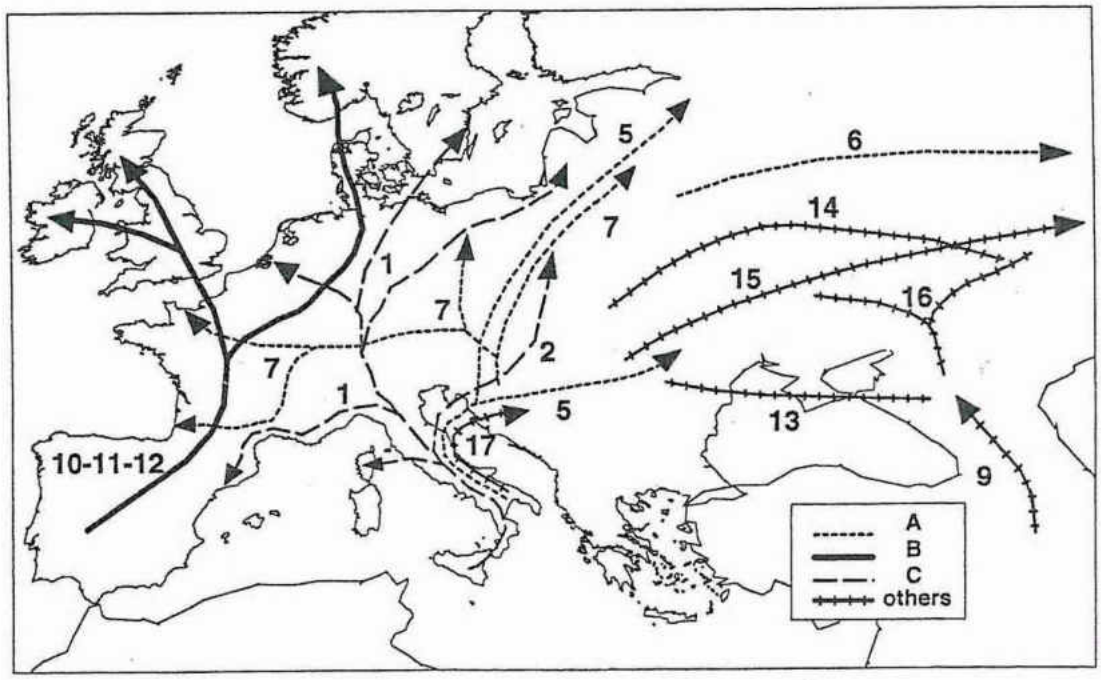


Figure 5. Voies de recolonisation des chênes en Europe, déduites de l'étude de l'ADN chloroplastique (Dumolin-Lapègue et coll. 1997).

Quatre haplotypes ont été détectés, dont trois fréquents (correspondant aux trois lignées chloroplastiques retrouvées par la suite à l'aide d'une technique basée sur la PCR). Une étude d'autocorrélation spatiale pour chacun de ces trois haplotypes démontra que les populations situées à moins de 300 kilomètres les unes des autres se ressemblaient particulièrement ; en fait, la prédominance d'un seul haplotype sur de grandes surfaces était évidente. J'ai proposé alors l'argument suivant, pour expliquer cette forte structuration : « *...the colonizing abilities of oaks, which are relatively important, must be distinguished from the level of migration which may be much lower in dense forests. A scenario which would involve the rapid postglacial spread of the species from glacial refugia, followed by the 'freezing' of the cytoplasmic genetic structure set during the colonization period, can be imagined* ». De plus, l'origine ibérique d'un des haplotypes était suggérée, cette région correspondant à l'un des refuges connus pour les chênes.

Ce travail s'est poursuivi, principalement dans le cadre des thèses de Brigitte Demesure[42] et de Sylvie Dumolin-Lapègue[43]. Une deuxième publication a fait le bilan de ces travaux (Dumolin-Lapègue et coll. 1997[44]).

Cette fois, l'échantillonnage avait été considérablement augmenté (345 populations réparties dans toute l'Europe), le nombre d'haplotypes était beaucoup plus important (23) et les références à la recolonisation étaient devenues centrales à l'analyse et à l'interprétation des données. Le nombre d'haplotypes présent dans le sud de l'aire (à proximité des zones refuges) était plus élevé qu'ailleurs. De plus, les haplotypes trouvés dans le sud constituaient un sous-ensemble de ceux déjà détectés dans le nord, indiquant que la plupart des haplotypes préexistaient dans les refuges, et que parmi les mutations observées très peu étaient apparues au cours de la migration. Ceci indique qu'une étude phylogénétique de l'expansion des chênes, basés sur des caractères qui ont évolué pendant cette migration, selon le modèle proposé par Thorpe (1984), n'est pas possible avec les outils dont nous disposons actuellement.

Une carte des voies de recolonisation fut proposée pour résumer les mouvements de recolonisation, sur la base de la distribution géographique des haplotypes les plus fréquents (figure 5). Une lignée maternelle (B) comprenant trois haplotypes (10-11-12) est visiblement originaire de la péninsule ibérique. Deux lignées et de nombreux haplotypes sont présents en Italie. D'autres haplotypes moins bien regroupés par l'analyse phylogénétique sont présents plus à l'est, jusque dans le Caucase, qui pourrait donc également avoir servi de refuge. Souvent, des haplotypes d'une même lignée semblent suivre des voies de recolonisation assez semblables (cas des haplotypes de la lignée B). Toutefois, ce n'est pas toujours le cas ; les haplotypes 2 et 5, par exemple, appartenant à des lignées différentes (C et A) mais issus d'un même refuge (italien) ont suivi des routes semblables. Par contre, les haplotypes 1 et 2 (lignée C) se sont séparés au niveau des Alpes. L'haplotype 1 est même redescendu vers le nord-est de l'Espagne, démontrant

---

polymorphisms in European oaks. *Theor. Appl. Genet.* **87**, 122-128.

[42] Demesure B. 1996. Analyse de la diversité chloroplastique en utilisant des fragments PCR chez des Fagacées : *Fagus sylvatica* L. et *Quercus* ssp. Thèse de Doctorat, Bordeaux I.

[43] Dumolin-Lapègue S. 1998. Variabilité chloroplastique et mitochondriale chez les chênes blancs européens à deux échelles spatiales. Thèse de Doctorat, Paris XI Orsay.

[44] Dumolin-Lapègue S, B Demesure, V Le Corre, S Fineschi et R J Petit. 1997. Phylogeographic structure of white oaks throughout the European continent. *Genetics* **146**, 1475-1487.

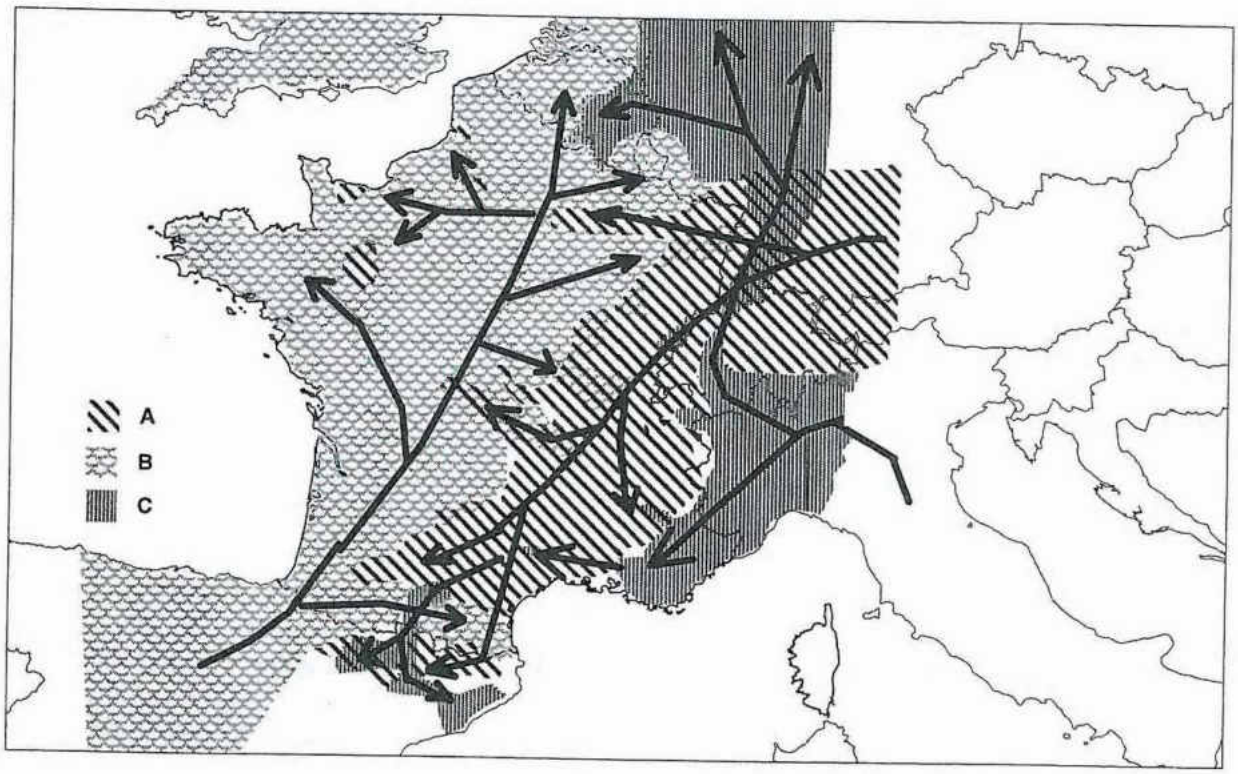


Figure 6. Voies de recolonisation détaillée des chênes en France. Trois principales voies de recolonisation sont visibles (originaires d'Espagne, d'Italie et des Balkans via le nord des Alpes).

que les mouvements postglaciaires ne sont pas uniformément orientés plein nord, et que des communications entre zones refuges ont eu lieu. Cette observation est importante car elle démonte (en partie au moins) l'argument de Bennett (1997) selon qui « *the occupation of the new range [by trees] is followed by local extinction, because environmental changes in the cooling phase of a glacial-interglacial oscillation are not the reverse of changes during a warming phase, particularly with respect to soil development. In this situation, the representativity of the spreading part of the species (and its subsequent genetic history) is irrelevant for the long-term genetic history of the species, because it is all lost.* ».

Il semble difficile de caler une horloge moléculaire pour de telles données, car les mutations détectées sont surtout de type insertion ou délétion. Toutefois, si on suppose que la divergence au sein de la lignée B (haplotypes 10-12) est apparue durant la dernière glaciation (100 000 ans environ), la divergence entre lignées (10-20 fois plus importante) est certainement beaucoup plus ancienne. Taberlet et coll. (1998) concluent que la divergence des lignées d'ADN mitochondrial des espèces animales en Europe date généralement de la fin du Tertiaire. Il leur semble donc difficile d'utiliser les données moléculaires pour dater des événements qui se sont produits durant le Quaternaire.

Une certaine interpénétration des lignées durant la recolonisation est nécessaire pour rendre compte des résultats ; en effet, l'haplotype 7, par exemple, se retrouve assez loin dans des régions où dominent des haplotypes de la lignée ouest (B).

Une analyse plus détaillée à l'échelle de la France (Dumolin-Lapègue et coll. 1998[45] ; Latouche 1998[46]) a permis d'aller plus loin dans l'interprétation des voies de colonisation dans cette région.

Elle a été effectuée dans le cadre d'un projet européen coordonné par A. Kremer[47] et auquel participent 12 laboratoires de l'Union Européenne. Un total de 869 populations (France, Wallonie et Luxembourg) ont été échantillonnées (3578 individus analysés au total). Des informations sur les pays limitrophes, obtenues dans le cadre de ce projet par nos partenaires, ont aidé à interpréter ces résultats.

La figure 6 schématise nos interprétations effectuées à partir des résultats obtenus. L'haplotype 7 (pratiquement le seul représentant de la lignée A) semble avoir contourné les Alpes par le nord, puis être redescendu vers le sud pour atteindre la Catalogne. Des tâches correspondant à des populations fixées pour l'haplotype 7 sont présentes assez loin vers l'ouest, au sein de la lignée B. A sein de cette lignée, issue de la péninsule ibérique, chacun des trois principaux haplotypes (10-11-12) peut dominer localement, résultant également en une structure en tâches. Par ailleurs, les haplotypes de la lignée C ont d'une part suivi la côte méditerranéenne, pour atteindre la Catalogne, et d'autre part ont traversé les Alpes et se sont répandus en Allemagne et à l'est de la Belgique. Il est

[45] Dumolin-Lapègue S, M-H Pemonge et R J Petit. 1998. Association between chloroplast and mitochondrial lineages in oaks. *Mol. Biol. Evol.* **15**, 1321-1331.

[46] Latouche-Hallé C. 1998. Variabilité de l'ADN chloroplastique des chênes blancs en France. DEA de Biologie, Diversité et Adaptation des Plantes cultivées, Paris XI Orsay.

[47] Projet FAIR : Synthetic maps of gene diversity and provenance performance for conservation and utilisation of oak genetic resources. Coordonné par A. Kremer.



Chêne pédonculé (*Quercus robur*).



Chêne sessile (*Quercus petraea*).



Chêne pubescent (*Quercus pubescens*).



Chêne tauzin (*Quercus pyrenaica*).

Figure 7. Distribution des quatre espèces de chênes blancs en France.

donc possible d'arriver à une description fort détaillée des mouvements de populations de chênes à cette échelle. On étudiera dans la partie suivante la structuration à une échelle beaucoup plus locale, et ce qu'elle nous enseigne sur le processus de recolonisation.

Il est difficile de préciser quelles espèces étaient représentées dans les différents refuges, puisque les grains de pollen des chênes à feuilles caduques ne peuvent être distingués. Toutefois, en considérant l'abondance des quatre espèces françaises (figure 7), on peut avancer les hypothèses suivantes : la lignée B (ibérique) devait vraisemblablement comporter le chêne pédonculé et le chêne tauzin. La rareté du chêne sessile dans le sud-ouest de la France et en Espagne pourrait indiquer qu'il s'y est récemment installé, et que son arrivée, au moins initialement, était concomitante de celle de l'haplotype 7, dont le point d'entrée en France est le nord des Alpes. La lignée A aurait quant à elle été représentée initialement par le chêne pubescent, en provenance d'Italie. L'hybridation entre ces espèces (dont il sera question plus bas) aurait ensuite gommé les associations initiales entre haplotypes et espèces.

Comme on le verra ci-dessous, l'étude sur le hêtre, dont la distribution en Espagne est très proche de celle du chêne sessile, renforce l'hypothèse de l'arrivée récente de cette espèce de chêne dans la péninsule ibérique. En effet, nous aboutissons à la même conclusion de colonisation tardive (*via* un refuge balkanique, comme pour le chêne sessile), avec deux types de marqueurs génétiques différents, et en accord avec les données polliniques qui n'ont pas l'ambiguïté de celles des chênes (car une seule espèce de hêtre est présente en Europe).

Variabilité du hêtre en Europe et recolonisation postglaciaire

L' histoire postglaciaire du hêtre commence a être bien connue. Il était présent au cours de l'époque glaciaire en Italie et au sud-est des Balkans (Grèce, Bulgarie, Macédoine notamment). La recolonisation fut beaucoup plus tardive que pour les chênes (vers 5 000 BP en France), et les populations les plus occidentales ou les plus nordiques sont très récentes, et sans doute encore en expansion en Espagne. Deux études de variabilité à l'échelle de l'aire ont été réalisées. La première est basée sur l'étude de l'ADN chloroplastique (Demesure et coll. 1996[48]). Elle a été complétée depuis, dans le cadre d'un projet européen sur la génétique du hêtre[49], en collaboration avec deux autres équipes européennes. La deuxième est une étude réalisée par Bernard Comps à l'université de Bordeaux, basée sur l'étude des isozymes ; j'ai participé à l'analyse et à l'interprétation de ces résultats (Comps et coll. 1999[50]).

Déductions sur l'histoire postglaciaire du hêtre à partir de l'ADN chloroplastique

Un total de 85 populations a été analysé, et 11 haplotypes ont été détectés. Toutefois, l'un de ces haplotypes est très abondant, depuis l'Espagne à l'ouest jusqu'à la Roumanie et l'Ukraine à l'est. Les seules régions différenciées étaient la Crimée (deux

[48] Demesure B, B Comps et R J Petit. 1996. Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution* **50**, *2515-2520.*

[49] Petit R J et S Dumas, étude en cours dans le cadre d'un projet FAIR sur la variabilité génétique du hêtre.

[50] Comps B, D Gömöry, J Letouzey, B Thiébaut et R J Petit. Imprint of postglacial colonization on the distribution of heterozygosity and allelic richness in the European beech. Soumis à *Genetics.*

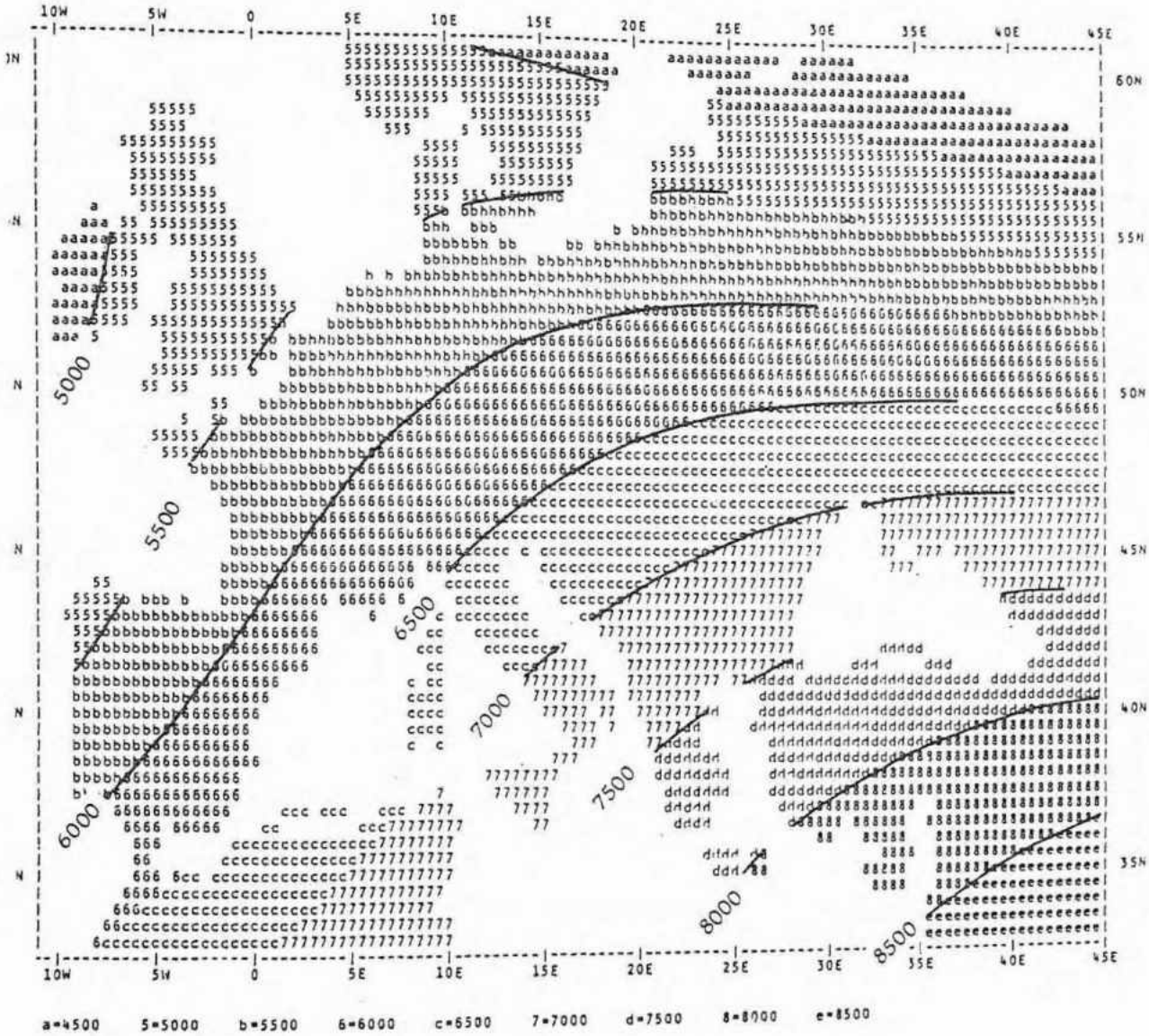


Figure 8. Progression de l'agriculture à l'Holocène à travers l'Europe occidentale, à partir du Moyen Orient, basée sur les datations C_{14} des premiers restes de l'agriculture dans chaque région. D'après Bodmer et Cavalli-Forza (1979).

haplotypes apparentés) et l'Italie et la Corse (trois haplotypes apparentés). Une étude plus complète (en cours) a confirmé ces résultats mais a montré que la Bulgarie et la Grèce étaient également distinctes (lignées très divergentes). L'haplotype le plus abondant en Europe se sépare désormais en trois haplotypes, dont l'un est distribué en Roumanie, les deux autres étant répartis en proportion variable depuis le nord des Balkans jusqu'en Espagne. Il semblerait donc qu'une première voie de recolonisation, dont l'origine se situe dans les Balkans, mais probablement pas parmi les populations les plus méridionales, a été à l'origine de la majorité des populations de hêtre apparues depuis la dernière glaciation. Les populations italiennes n'auraient donc pas dépassé les Alpes, contrairement à ce qui était observé pour les chênes.

Faut-il pour rendre compte des mécanismes de colonisation des arbres au Quaternaire faire abstraction de l'action de l'homme ? L'homme a pu agir en disséminant des graines : on sait que des graines produites en grande quantité, comme les glands, étaient fréquemment utilisés pour la consommation humaine (Lieutaghi 1998). L'homme peut aussi agir indirectement en favorisant une espèce. La progression de l'agriculture à l'Holocène (*cf.* figure 8) précède de manière assez remarquable la progression de certains arbres apparus tardivement comme le hêtre. Selon Sauer (1988), « *The puzzling acceleration during the Subboreal [5000 BP] of the advance of beech,* Fagus sylvatica, *has also been attributed to human agency. Unlike its North American congener, the European beech is not usually able to invade undisturbed forest. Its sudden expansion coincides in time with the elm fall and the beginning of shifting cultivation.* ».

Déductions sur l'histoire postglaciaire du hêtre à partir des isozymes

Une étude très détaillée de la variabilité isoenzymatique du hêtre en Europe a été réalisée (Comps et coll. 1999). L'aire est très largement couverte : 389 populations ont été échantillonnées de la Galice à la Crimée et de la Sicile à la Suède. Il y a environ 50 arbres par population, étudiés à l'aide de 15 locus enzymatiques (dont 12 polymorphes). Seuls, les niveaux de diversité (richesse allélique et hétérozygotie) ont été étudiés pour l'instant, afin d'examiner l'impact de la colonisation sur les niveaux de diversité à l'échelle de l'ensemble de l'aire. Par contre, il est trop tôt pour examiner si ces marqueurs permettent d'identifier ou de préciser les voies de colonisation de l'espèce : il faudrait pour cela étudier les distances génétiques entre ces populations, et pas seulement leur niveau de diversité. En outre, il faudrait également se baser sur les données multilocus, car les associations entre locus différents, surtout s'ils sont liés, peuvent être très utiles pour reconstituer l'histoire de ces populations (Kremer et Zanetto 1997).

Néanmoins, la cartographie des niveaux de diversité (richesse allélique et hétérozygotie) est déjà extrêmement informative. Les zones refuges sont caractérisées par une faible hétérozygotie et une forte richesse allélique (figure 9), ainsi que par une forte hétérogénéité dans la composition allélique des populations. Il est fréquent que la comparaison des zones refuges et des zones nouvellement recolonisées soit compliquée par d'autres facteurs. Ici, les populations méridionales correspondant aux zones refuge sont situées presque exclusivement en altitude. Toutefois, l'existence d'un massif

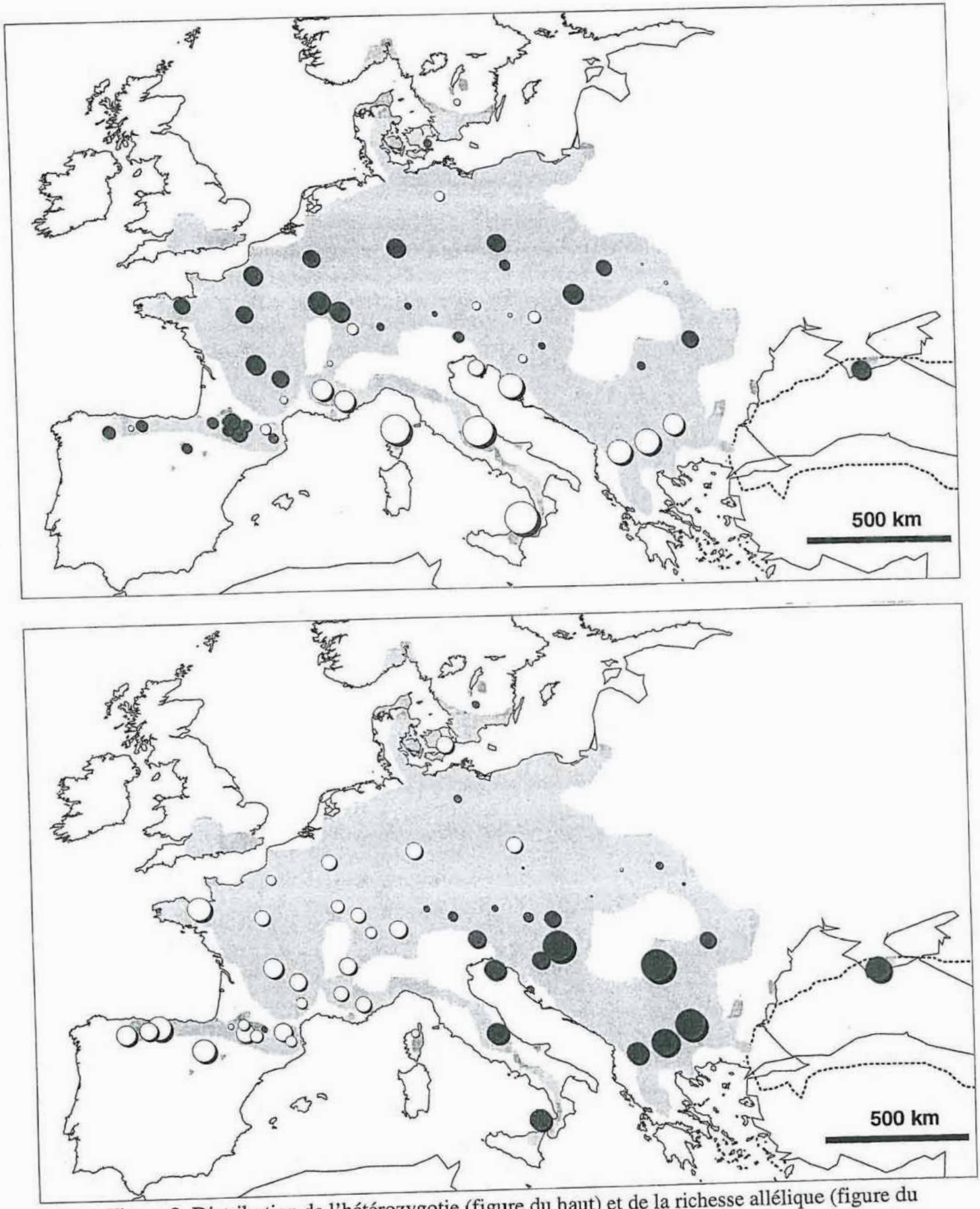


Figure 9. Distribution de l'hétérozygotie (figure du haut) et de la richesse allélique (figure du bas) chez le hêtre, d'après une étude de la variabilité des isozymes (Comps et coll. 1999). Les symboles en noir indiquent des valeurs supérieurs à la moyenne, ceux en blanc les valeurs inférieures à la moyenne. Noter l'opposition entre les deux mesures de diversité. Les zones de forte richesse et de faible hétérozygotie correspondent bien aux zones refuges.

montagneux (les Pyrénées) éloigné des zones refuges facilite les interprétations.

Autres espèces étudiées

Le sapin pectiné

L'aire actuelle du sapin pectiné s'est mise en place entre 11 000 BP et 6 000 BP, à partir de cinq refuges, dont trois ont servi de source à l'expansion de l'espèce (Gliemeroth 1997, Konnert et Bergmann 1995). Dans une étude européenne à laquelle j'ai participé, 714 arbres répartis dans 17 populations ont été étudiés à l'aide de deux microsatellites chloroplastiques (transmis paternellement chez cette espèce), et 90 haplotypes ont été identifiés (Vendramin et coll. 1999[51]).

Une très forte hétérogénéité des niveaux de diversité (héterozygotie et richesse allélique) a été mise en évidence. La population pyrénéenne est celle qui est la plus divergente des autres. Toutefois, c'est aussi celle qui possède le moins d'haplotypes. Ces résultats sont compatibles avec l'existence d'un refuge en cet endroit, mais qui serait resté isolé pendant longtemps. Les autres populations particulièrement divergentes sont celles situées en Calabre, en Bulgarie et en Croatie. Celle de Croatie avait en outre la plus forte richesse allélique. Ces trois régions font précisément partie des zones refuges déduites à l'aide d'autres marqueurs (Konnert et Bergmann 1995).

La cochenille du pin maritime (Burban et coll. 1999[52])

Cet insecte (*Matsucoccus feytaudi*) est inféodé au pin maritime, auquel il cause des dégâts, très spectaculaires (dépérissements allant jusqu'à la disparition des peuplements), du moins dans une partie de l'aire (sud-est de la France et nord de l'Italie, et désormais en Corse également). Il s'agit là de populations dites épidémiques. Par contre l'insecte est présent sans que l'arbre hôte semble en souffrir dans une large partie de l'aire (populations dites 'endémiques', dans la péninsule ibérique, le Maroc, le sud-ouest de la France). Une étude de la phylogéographie de l'insecte a été effectuée, à l'aide de l'ADN mitochondrial (Burban et coll. 1999) ; celle du pin est en cours de réalisation. Les résultats démontrent très clairement que l'étude d'espèces étroitement associées à des espèces d'arbres forestiers peut nous donner des informations très précieuses sur l'histoire de l'arbre lui-même. Dans le cas présent, l'ADN mitochondrial de l'insecte a permis d'identifier les zones refuges du pin et de mettre en évidence des transferts artificiels de certaines populations de cet arbre. Certaines précautions liées à la spécificité de la relation plante-organisme considérée sont toutefois indispensables. Par example, l'absence de variation de l'ADN mitochondrial de l'insecte dans les populations épidémiques du sud-est de la France et d'Italie correspond à un goulot d'étranglement des populations de l'insecte mais non des peuplements de pin maritime.

[51] Vendramin G G, B. Degen, R J Petit, M Anzidei, A Madaghiele et B Ziegenhagen. 1999. High level of variation at *Abies alba* microsatellite loci in Europe. *Mol. Ecol.*, sous presse.

[52] Burban C, R J Petit, E Carcreff et H Jactel. Rangewide variation of the maritime pine bast scale *Matsucoccus feytaudi* Duc. (Homoptera: Matsucoccidae) in relation to the genetic structure of its host. Soumis à *Mol. Ecol.*

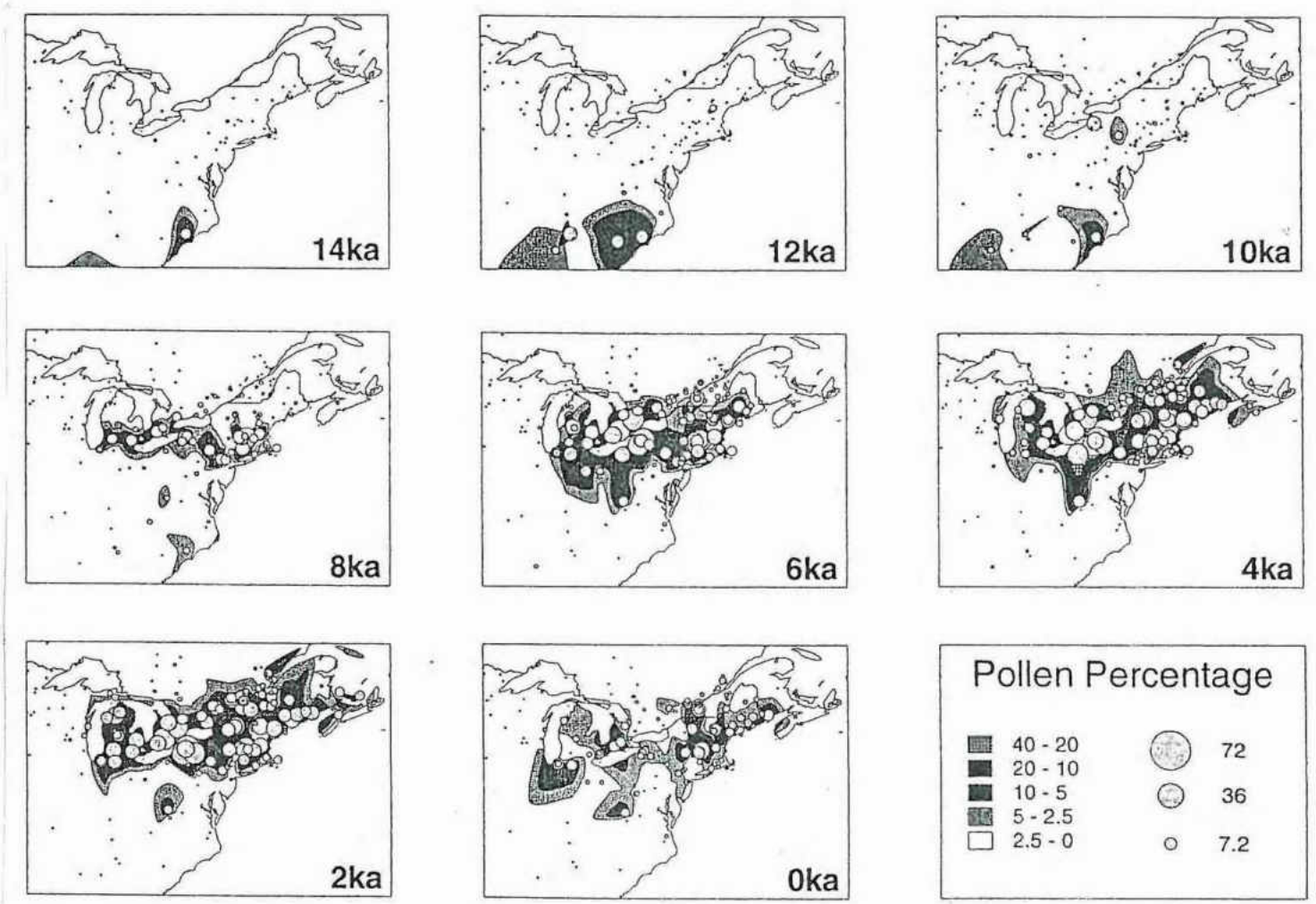


Figure 10 : Progression du hêtre (*Fagus grandifolia*) en Amérique du Nord, 14000 BP à l'époque actuelle. Les cercles représentent l'emplacement des sites palynologiques, et leur taille est liée au pourcentage de pollen de *F. grandifolia* dans les sédiments. Les contours donnent les pourcentages de pollen obtenus par interpolation, pour montrer les niveaux d'abondance. La migration fut d'abord orientée vers le nord, puis vers l'ouest. L'aspect de la progression est irrégulier, reflétant la distribution des sites ayant des conditions édaphiques et topographiques convenables, mais aussi des évènements de dispersion singuliers, à de grandes distances. D'après Clark et coll. (1998).

PROCESSUS DE COLONISATION ET DE CROISSANCE DES POPULATIONS

Evénements de fondation

« *Je dois dire maintenant quelques mots sur ce qu'on a appelé* les moyens accidentels de dispersion, *moyens qu'il vaudrait mieux appeler* occasionnels *[...] Les 14/100 de graines de plantes d'un pays pourraient donc être transportés à travers 924 milles (1487 kilomètres) de mer jusque dans un autre pays, et germer...* ».
Charles Darwin, 1858. L'origine des espèces, chapitre 12. Traduction d'Edmond Barbier, 1876.

« *The oak, to gain its present most northerly position in North Britain after being driven out by the cold probably had to travel fully six hundred miles, and this without external aid would take something like a million years* » (Reid 1899).

En écrivant ceci, Reid avait une idée bien précise ; il avait en effet publié quelques années auparavant une courte note dans la revue *Nature* intitulée « *The dispersal of acorns by rooks* » (Reid 1895), travaux peut être inspirés de ceux de Darwin, fasciné depuis son voyage autour du monde sur le *Beagle* par la colonisation des îles océaniques.

Chez les chênes, la vitesse de colonisation pendant l'époque postglaciaire a atteint des pointes de l'ordre de 500 mètres par an en Europe (MacDonald 1993). Ceci suppose des distances considérables de migration, si l'on tient compte de l'âge minimum requis pour qu'un chêne serve de source de graines pour un nouveau bond en avant (20 ans en milieu ouvert semble une valeur minimale, ce qui nous donne des distances de dispersion de 10 km par 'génération'). En fait, les taux de doublement des populations minimum, estimés d'après les données de pollen fossile, sont pour les chênes de l'ordre de 70 ans (MacDonald 1993) ; on obtient avec ce chiffre des valeurs de 35 km pour ces migrations à longue distance. On peut aussi raisonner que, si ces événements de colonisation à très longue distance sont rares, donc improbables, il faut que la population source produise suffisamment de graines pour que ces événements aient une chance non nulle de se produire. D'après Sauer (1988), « *During this free migration, location of the moving borders would have been largely controlled by how far seed was dispersed and how long it took for new colonies to mature and produce the next wave of seed* ». Si l'on rajoute deux ou trois générations supplémentaires pour que la population source atteigne une taille suffisante pour devenir un véritable front de colonisation susceptible d'engendrer un nouveau bond en avant, on arrive alors à des distances de dispersion de l'ordre de 100 km ; les corbeaux ou les geais doivent penser que cela fait beaucoup...

Les paléobotanistes ont accumulé des résultats sur des espèces telle le hêtre *Fagus grandifolia* aux USA, démontrant que l'invasion de cette espèce s'était en effet effectuée *via* l'établissement de petites populations très en avant du front de colonisation (Bennett 1985, 1988, Birks 1989, Clark et coll. 1998, Davis et coll. 1986, Woods et Davis 1989, MacDonald et Cwynar 1991, Webb 1987 ; voir figure 10). Les distances citées varient de 25 à 120 km, ce qui n'est pas si éloigné des estimations précédentes. En outre, de larges barrières physiques qui auraient pu bloquer la dispersion (par exemple la mer du Nord, la Baltique, le lac Michigan, les Alpes...)

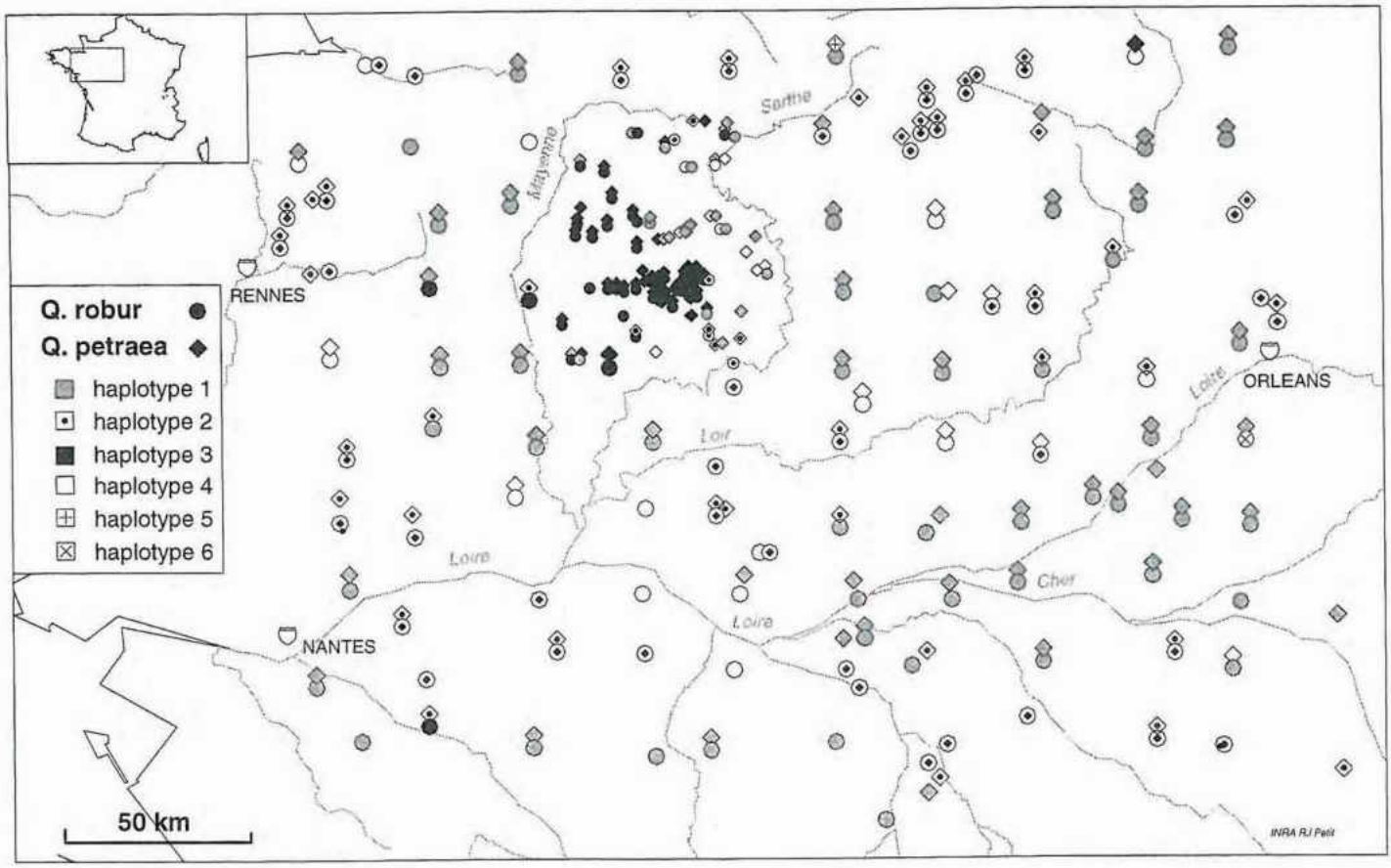


Figure 11 : Distribution des variants chloroplastiques de deux espèces de chênes (*Quercus robur* et *Q. petraea*) dans l'ouest de la France (Petit et coll. 1997). Noter la forte structuration spatiale, en « tâches » de quelques dizaines de kilomètres.

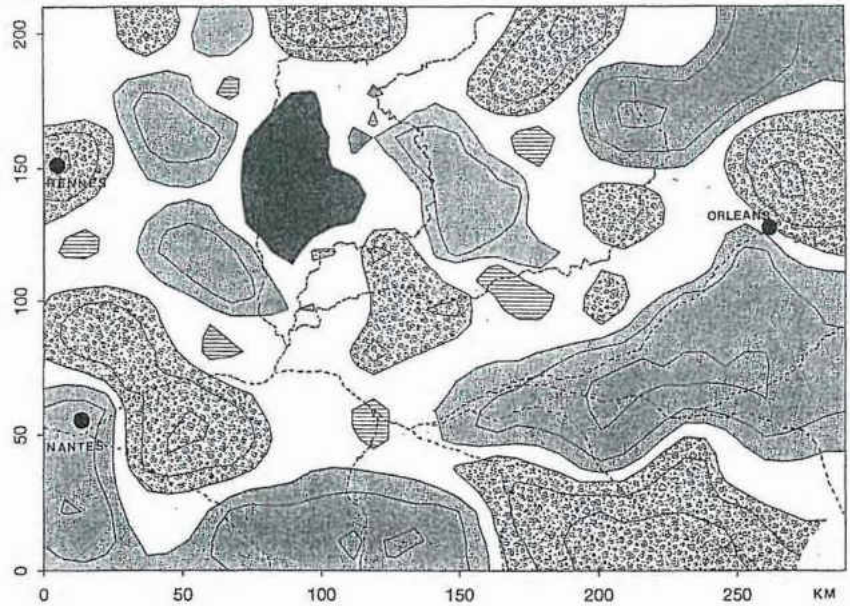


Figure 12 : Carte obtenue par interpolation des résultats précédents après modélisation (techniques de géostatistiques). Les courbes représentent les fréquences prédites pour chacun des quatre haplotypes les plus représentés.

semblent avoir posé peu de problèmes aux arbres (Clark et coll. 1998). Tout ceci contribua aux développements de modèles de dispersion mixtes (Mollison 1977, Moody et Mack 1988, Shigesada et coll. 1995, Clark 1998), qualitativement différents des modèles de diffusion développés auparavant (Fisher 1937, Skellam 1951). Ces modèles sont caractérisés par leur queue de dispersion particulière (« *fat tails* »). Une conséquence de ce mode de dispersion mixte est que le rythme de colonisation est variable (il ne s'agit pas d'une 'vague' de migration, mais d'un processus fait d'accélérations brusques et de pauses plus ou moins longues). Ce sont bien, dans ces conditions, les événements les plus rares de colonisation qui ont le plus d'importance, et non le comportement moyen de la population.

Ces événements exceptionnels ne sont guère observables (en compulsant la littérature ornithologique, j'ai toutefois trouvé une description de mouvements de geais transportant des glands sur près de huit kilomètres, poursuivis par un berlinois muni de jumelles et chevauchant une bicyclette (Schuster 1950) !). La dispersion moyenne, ou même la valeur du quartile ou du centile supérieur ne présentent aucun intérêt ; ce sont les événements extrêmes, dont l'étude est par définition en dehors du champ d'investigation classique, qui jouent le rôle prépondérant. Chaque bond en avant est un événement 'historique', et la structuration génétique qui en découlera ne sera qu'une réalisation d'un processus stochastique.

Dans des synthèses récentes sur la migration des arbres à l'Holocène, les écologistes, paléoécologistes, biogéographes et modélisateurs (Pitelka et coll. 1997, Clark et coll. 1998) n'imaginaient pas que les processus de migration qu'ils étudiaient indirectement mais qu'ils avaient du mal à visualiser, puisse apparaître aussi nettement aux yeux des généticiens des populations. En effet, nous avons mis en évidence chez les chênes, à l'échelle locale, une structuration chloroplastique très forte, organisée en tâches de quelques dizaines de kilomètres (Petit et coll. 1996[53] 1997[54] ; voir figures 11 et 12). Cette structuration paraît liée aux mouvements des graines lors de la colonisation, et à la fondation de petites populations isolées, constituées initialement d'un seul semencier qui fera office d'une véritable 'Eve' locale dont dérivent par la voie maternelle tous les arbres environnants.

L'hypothèse qui avait été formulée pour rendre compte de cette structuration à l'échelle locale (événements de colonisation à longue distance lors de la recolonisation) a été évaluée par simulation (Le Corre et coll. 1997[55]).

Dans ces simulations, la vitesse de recolonisation était très sensible aux paramètres contrôlant la dispersion à longue distance (probabilité de ces événements et distance maximum permise). Par exemple, une probabilité aussi faible que 5×10^{-6} combinée à une distance maximale de 50 km donnaient une vitesse de migration et une structuration maternelle compatibles avec nos observations.

---

[53] Petit R J, B Demesure et E Pineau. 1996. Chloroplast DNA differentiation at a local and continental scale in European oak species : the importance of historical factors. In A Kremer, H-H Muhs, eds. *Inter- and intraspecific variation in European oaks : Evolutionary implications and practical consequences.* Union Européenne, Luxembourg.

[54] Petit R J, E Pineau, B Demesure, R Bacilieri, A Ducousso et A Kremer. 1997. Chloroplast DNA footprints of postglacial recolonization by oaks. *Proc. Natl. Acad. Sci. USA* **94**, 9996-10001.

[55] Le Corre V, N Machon, R J Petit et A Kremer. 1997. Colonization with long-distance seed dispersal and genetic structure of maternally inherited genes in forest trees: a simulation study. *Genet. Res., Camb.* **69**, 117-125.

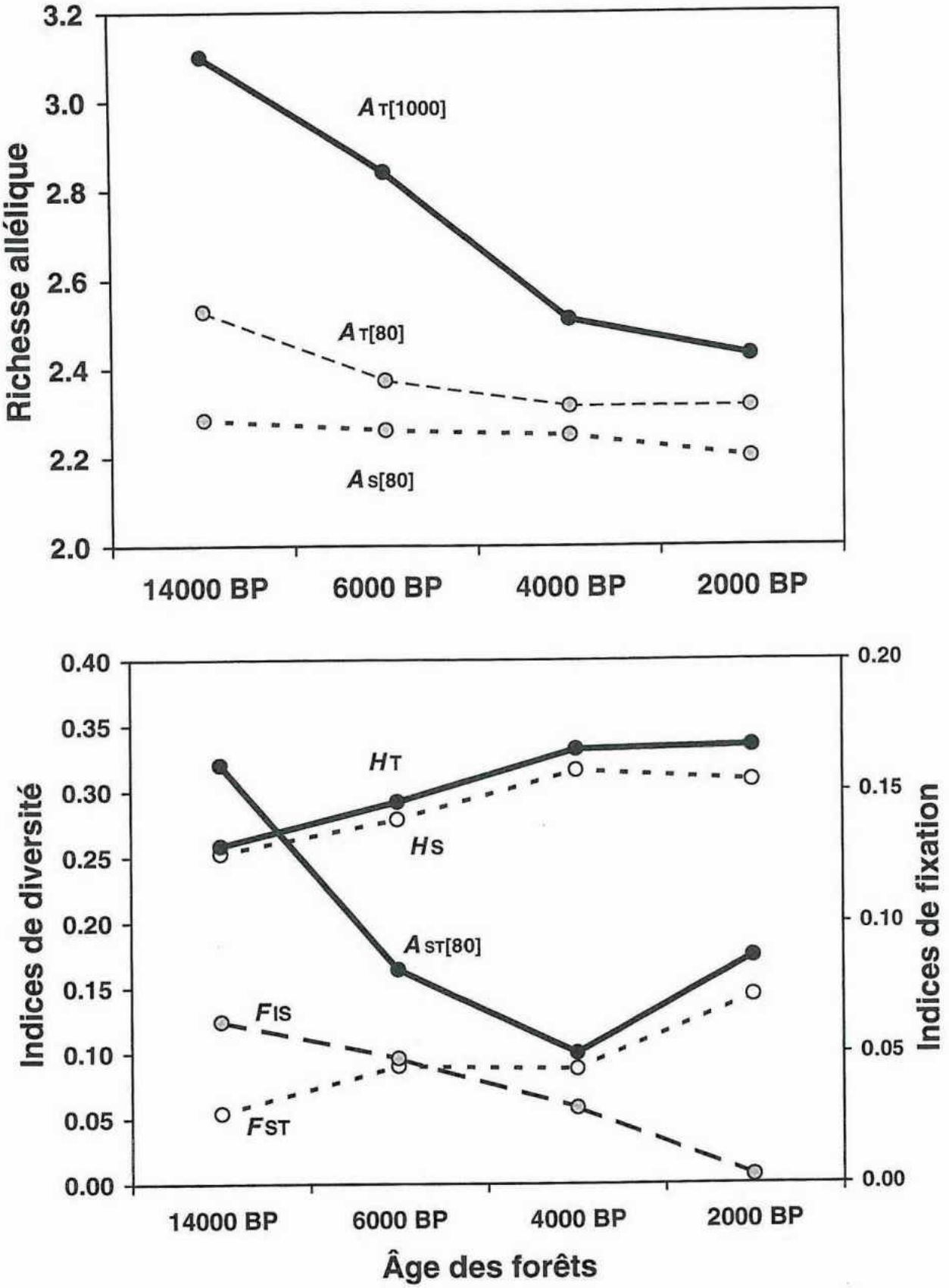


Figure 13. Evolution des principaux indices de diversité et de fixation en fonction de l'âge de la population, chez le hêtre. Noter le chassé-croisé de la richesse allélique et de l'hétérozygotie, la chute du F_{IS} et l'augmentation du F_{ST}.

Croissance des populations après colonisation

L'étude sur la variabilité enzymatique du hêtre, dont il a été question plus haut (Comps et coll. 1999), a précisé encore ces modèles de colonisation et surtout de croissance ultérieure des populations. En effet, les événements de fondation à longue distance ont nécessairement dû entraîner une perte d'allèles lors de la progression de l'espèce. L'application de tests destinés à démontrer l'existence de goulots d'étranglement (chute plus rapide du nombre d'allèles que de l'hétérozygotie dans les populations distantes des zones refuges, Cornuet et Luikart 1996) a montré que des changements démographiques récents avaient effectivement eu lieu pour les populations de hêtre établies plus récemment.

Toutefois, le modèle du goulot d'étranglement ne semble pas parfaitement adapté aux événements en question ; il s'agit plutôt d'événements de fondation successifs dans des populations liées par des flux de gènes. En fait, l'hétérozygotie, chez le hêtre mais aussi chez le chêne sessile, n'a nullement diminué au cours de la migration de ces espèces. Elle a même augmenté ! Ce résultat surprenant est largement discuté dans le papier correspondant (Comps et coll. 1999).

Comment alors caractériser ces événements de fondation ? Les modèles de métapopulations (modèles d'évolution des populations liées par des flux de gènes et où l'extinction et la recolonisation des populations sont prises en compte) peuvent servir de base d'analyse dans un premier temps. Dans ces modèles, la recolonisation peut constituer une forme de flux de gènes (Slatkin 1977), à condition que les colons soient plus nombreux que les migrants échangés entre populations et non apparentés. Dans le cas du hêtre, un rapide calcul indique qu'une trentaine au moins de colons diploïdes (graines) devrait être à l'origine des nouvelles forêts de hêtre pour que l'hétérozygotie commence à augmenter. J'ai calculé, à l'aide de courbes de raréfaction, que dans ces conditions, la richesse allélique continuera de baisser au cours de la colonisation (alors que H augmente), du fait de la dynamique différente des allèles rares ; il serait intéressant de confirmer ce résultat sur le plan analytique ou par simulation. Le nombre élevé de colonisateurs est toutefois en totale contradiction avec les résultats précédents (basés sur l'ADN chloroplastique, et qui avaient conclu que un seul arbre était à l'origine des tâches cytoplasmiques). Les modèles de métapopulation, adaptés à l'étude de populations en expansion (Le Corre et Kremer 1998), sont nécessairement simplifiés et ne tiennent pas compte de la croissance puis de la rencontre (*admixture*) des populations issues de fondateurs différents. Des échanges considérables ont pu avoir lieu à ces stades ultérieurs, par pollen, sans affecter la structuration maternelle. Un résultat classique mais peu connu est que les faibles densités augmentent l'intensité des flux de gènes chez les plantes (ex : Ellstrand et Ellam 1993). Il serait particulièrement intéressant de prendre en compte cette dynamique des flux de gènes, différente à chaque stade de l'évolution des forêts, depuis leur fondation jusqu'à l'équilibre. La figure 13 schématise l'évolution des principaux indices de diversité au cours du temps chez le hêtre, soulignant la complexité de la dynamique et les échelles de temps impliquées (données issues de Comps et coll. 1999). Les modèles développés par Austerlitz et coll. (1997) pourraient être adaptés pour prendre en compte ces changements de flux de gènes.

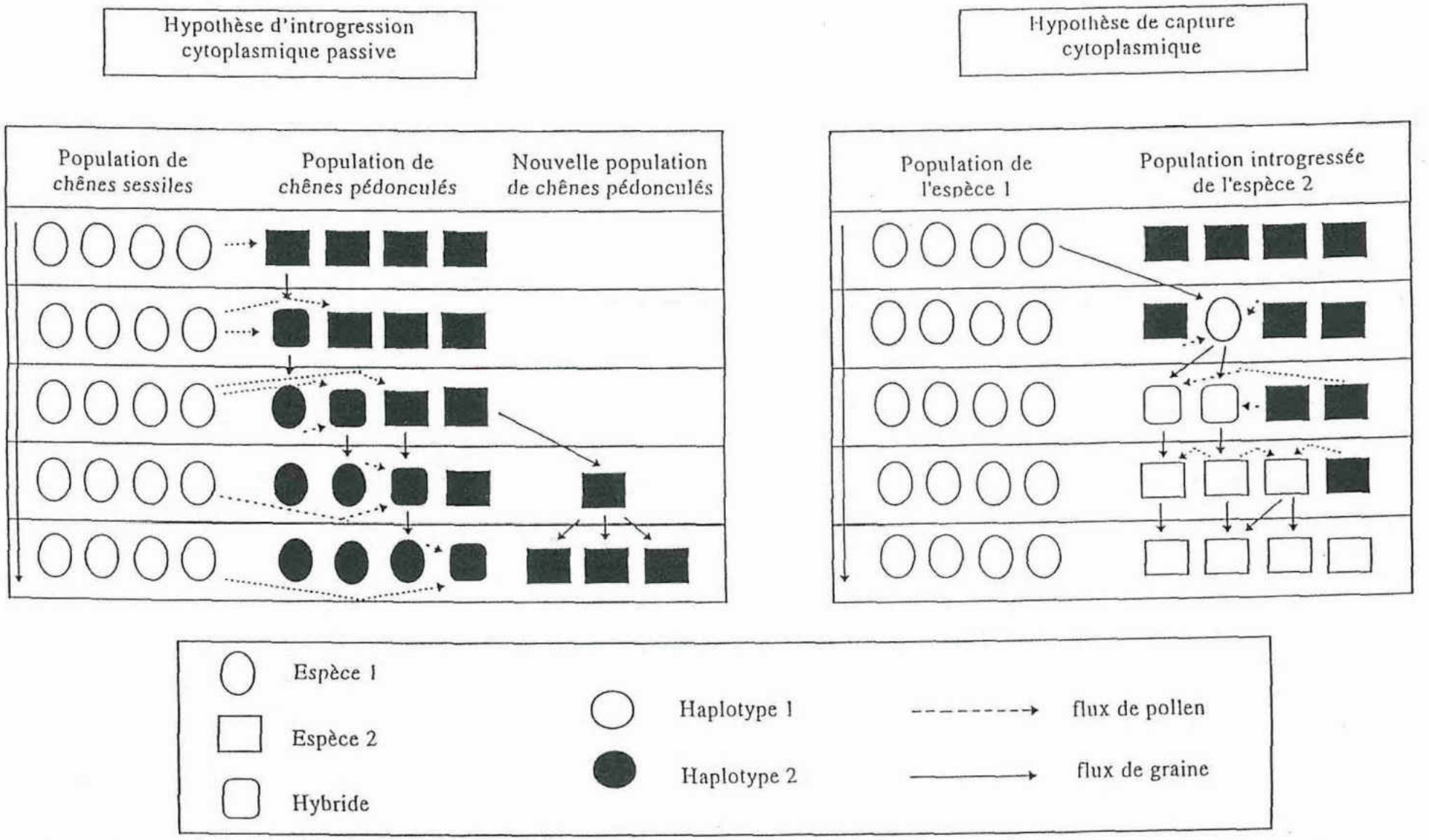


Figure 14 : Les deux scénarios aboutissant à des échanges cytoplasmiques entre espèces de plantes apparentées. L'hypothèse d'introgression cytoplasmique passive que j'ai proposée (à gauche) diffère de la « capture cytoplasmique » décrite par Rieseberg et Soltis (1991) de plusieurs manières. Elle est dynamique : de nouvelles populations se créent, puis des échanges par pollen ont lieu entre ces nouvelles populations et celles d'autres espèces moins pionnières. De plus, les flux interpopulation (intra- et interspécifiques) sont presque uniquement le fait du pollen.

Colonisation et hybridation

Dans le chapitre 2, l'importance des échanges cytoplasmiques entre espèces différentes avait été soulignée (pour les génomes transmis maternellement). Un modèle d'échanges, par capture du génome nucléaire (et non des génomes cytoplasmiques) avait été proposé. Toutefois, ce modèle est plus facile à comprendre dans un cadre dynamique, notamment lors de la colonisation postglaciaire.

Ces échanges posent de plus des problèmes évolutifs qui n'ont pas été réellement résolus à ce jour, et que je voudrais de nouveau souligner ici. Rappelons les faits, dans le cas des chênes : Les fleurs femelles de l'espèce pionnière (dite *r*) sont réceptives au pollen de l'espèce lui succédant généralement dans les forêts plus fermées (espèce *K*) [*r* = chêne pédonculé, *K* = chêne sessile]. Le modèle proposé était donc le suivant (Petit 1992[56] ; Bacilieri et coll. 1996[57] ; Petit et coll. 1997[58]) : le cytoplasme de l'espèce *r*, croisée de manière répétée avec l'espèce *K*, se retrouve finalement dans un environnement nucléaire correspondant à l'espèce *K* (figure 14).

Aucun mouvement physique du cytoplasme n'est nécessaire pour rendre compte de ce passage du cytoplasme d'une espèce à l'autre, en accord avec la faiblesse des flux par graines, et l'importance des flux par pollen. En effet, pour des flux par pollen et par graines aussi inégaux, quelque soit le sens préférentiel de la compatibilité, des hybridations lointaines avec l'espèce en place (par pollen) seront plus efficaces que de nouveaux événements de colonisation par graine.

Toutefois, dans le modèle étudié (chêne sessile/chêne pédonculé), il faut bien constater que ces échanges génétiques accélèrent la succession, et que l'espèce *K* peut 'bénéficier' de l'espèce *r* pour assurer sa colonisation, sans avoir à investir dans de coûteux mécanismes de dispersion des graines à longue distance (pédoncule allongé, glands de grande taille, plus long pouvoir de conservation lié à une germination différée, etc.). Son pollen pourra assurer cette dispersion, falsifiant ainsi la loi énoncée par Harper (1977) : « *[in plants], the haploid phase of environmental exploration cannot act further than has already been colonised by the diploid phase of the species* ». Au contraire, chez les chênes, le pollen peut 'prospecter' les milieux colonisés par d'autres espèces de chênes ; ce mécanisme est peut être à l'origine de leur succès évolutif dans les zones tempérées où la colonisation de nouveaux territoires est essentielle pour se maintenir. Dans ce contexte, il est tentant de voir dans le mécanisme d'hybridation asymétrique une adaptation qui facilite la succession. Notons en effet que la base génétique de cette asymétrie a été prouvée dans des expériences de croisements contrôlés, et pas seulement au travers de la mesure des flux de gènes interspécifiques en forêt. Un calcul à partir de données démographiques et génétiques dans la parcelle mixte de chênes pédonculés et sessiles, à partir des données de Bacilieri et coll. (1996), montrait que les flux interspécifiques étaient aussi efficaces que la démographie (production de glands et survie

[56] Petit R J. 1992. Polymorphisme de l'ADN chloroplastique dans un complexe d'espèces: les chênes blancs européens. Thèse, Paris XI-Orsay, 87 p.

[57] Bacilieri R, A Ducousso, R J Petit et A Kremer. 1996. Mating system and asymmetric hybridization in a mixed stand of European oaks. *Evolution* **50**, 900-908.

[58] Petit R J, E Pineau, B Demesure, R Bacilieri, A Ducousso et A Kremer. 1997. Chloroplast DNA footprints of postglacial recolonization by oaks. *Proc. Natl. Acad. Sci. USA* **94**, 9996-10001.

différentielles) pour modifier la composition spécifique du peuplement étudié. On notera aussi que l'aire de l'espèce *K* est presque entièrement incluse dans celle de l'espèce *r*, et que l'asymétrie de compatibilité a été notée dans au moins trois pays différents.

Pourrait-il donc s'agir d'une adaptation ? On pourrait par exemple penser à un 'effet métapopulation', tel qu'il a été décrit par Olivieri et Gouyon (1997). Dans le cas de l'asymétrie de compatibilité de ces chênes, toutefois, il faudrait étudier si un tel caractère peut effectivement évoluer dans des conditions ou les extinctions et colonisations (voire des expansions et contractions d'aire) sont fréquentes. Pour mesurer le terrain qui resterait à parcourir avant d'aboutir à une démonstration satisfaisante en terme d'adaptation, on peut citer Brandon (1996) : « *an ideally complete adaptation explanation has five components : (1) Evidence that selection has occurred on the relevant traits ; (2) An ecological explanation of that selection ; (3) An account of the underlying genetics of the traits (at minimum, evidence that the relevant traits are heritable) ; (4) Information on population structure (in particular on patterns of gene flow and selective heterogeneity) ; and (5) Phylogenetic information on trait polarity, that is, on which trait values are evolutionarily primitive and which derived.* »

L'étude d'autres espèces de chênes blancs (chêne pubescent, chêne tauzin) a montré que des échanges cytoplasmiques avaient lieu également avec ces espèces (Petit et coll. 1999[59]), mais on ignore si des sens de croisements préférentiels existent.

Un autre exemple d'échanges cytoplasmiques a été mis en évidence entre le chêne vert (*Quercus ilex*) et le chêne liège (*Quercus suber*), au Maroc (Belahbib et coll. 1999[60]) et en Espagne (Jimenez et Petit, données non publiées).

Ces deux chênes appartiennent à des sections différentes du genre *Quercus*. Les résultats indiquent que l'importance des échanges cytoplasmiques est de nouveau très importante. Toutefois, dans cet exemple, il est possible d'identifier l'origine des deux génomes cytoplasmiques, et donc de préciser *a priori* le sens de l'introgression. En effet, deux types chloroplastiques et mitochondriaux, très divergents, ont été détectés ; ces deux types (cpA et cpB, mtA et mtB), sont toujours associés (cpA avec mtA, cpB avec mtB). Chez *Q. ilex*, sauf de très rares exceptions, on ne trouve que les génomes A, avec de nombreuses variantes pour le génome chloroplastique, surtout dans la péninsule ibérique ; chez *Q. suber*, on trouve principalement les génomes B, mais dans de vastes régions où le chêne liège est peu abondant (sud de l'aire marocaine de l'espèce, est de l'Espagne), tous les chênes lièges ont les génomes A. De plus, le génome B est proche de celui du chêne chevelu (*Q. cerris*), qui est apparenté taxonomiquement au chêne liège (mais non au chêne vert). Il est donc probable que les génomes A représentent la lignée de *Q. ilex*, que les génomes B proviennent de la lignée du chêne chevelu, et que les échanges se sont faits dans le sens : pollen de *Q. suber* → fleur femelle de *Q. ilex*.

En fait, les chênes verts fleurissent avant les chênes lièges, empêchant généralement les croisements inter-spécifiques (forts rares, d'après la fréquence des individus morphologiquement intermédiaires).

[59] Dumolin-Lapègue S, A Kremer et R J Petit. Are chloroplast and mitochondrial DNA species-independant in oaks. *Evolution* (sous presse).

[60] Belahbib N, M-H Pemonge, A Ouassou, H Sbay, A Kremer et R J Petit. Chloroplast and mitochondrial DNA exchanges between cork oak (*Q. suber* L.) and holm oak (*Q. ilex* L.) in Morocco. En préparation.

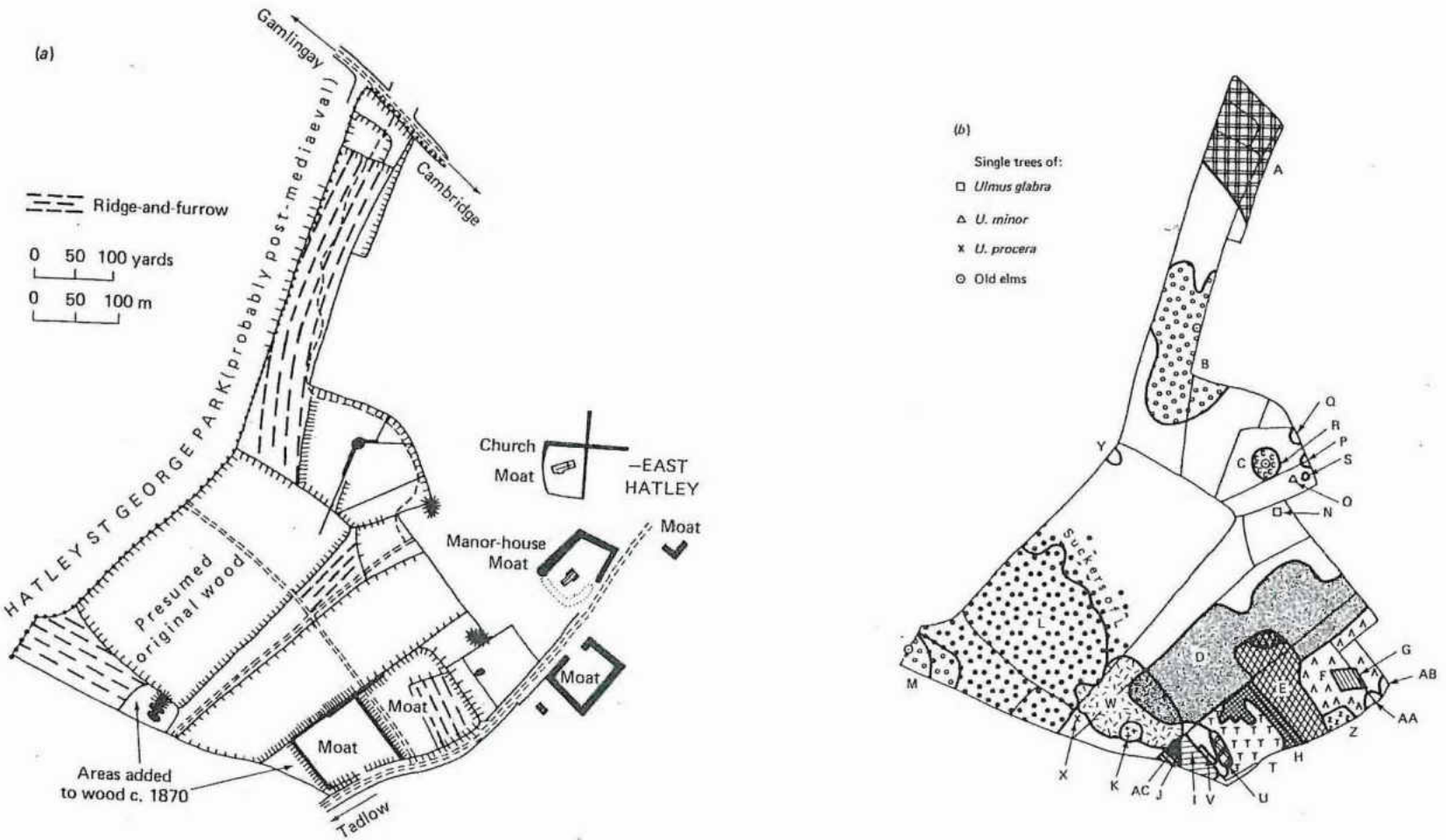


Figure 15 : Histoire écologique d'un bois du Cambridgeshire en Angleterre. Ce petit bois de 16 ha a une histoire complexe. Un noyau de bois ancien est entouré de zones qui ont été cultivées à certaines périodes au moins depuis 1350. A gauche, plan général. A droite, détails de la distribution des ormes (*Ulmus* sp.) dans le bois. Chaque lettre indique une invasion séparée par un clone. Les clones de *Ulmus procera* et *procera*×*minor* sont représentés par des hachurés doubles, *U. glabra* et ses hybrides sont représentés par des hachurés simples. Les clones d'*U. minor* sont figurés par divers pointillés ; les lettres correspondent à des arbres isolés ou de petits clones de ces espèces. D'après Rackham (1975, 1980), in Bridges ans Walters (1984).

Toutefois, les chênes verts tardifs *pourraient se croiser avec des chênes lièges* précoces ; comme ces chênes sont protandres[61], le croisement attendu est bien celui indiqué. Des croisements en retour répétés des hybrides *Q. ilex* (♀) × *Q. suber* (♂) avec du pollen de *Q. suber* auraient ainsi produit des chênes lièges typiques (certaines de ces populations de chêne liège ayant les génomes cytoplasmiques A ont un liège de très grande qualité !). Là aussi, une extension de l'aire du chêne liège par hybridation a donc eu lieu : on notera en effet que l'aire du chêne liège, assez limitée, est largement contenue dans celle du chêne vert, beaucoup plus étendue. D'autre part, le chêne liège est beaucoup moins variable génétiquement que le chêne vert, notamment au niveau cytoplasmique ; il pourrait s'agir d'une espèce apparue récemment ou ayant connu un goulot d'étranglement, et dont les caractéristiques (adaptation contre les incendies par exemple) lui ont permis de se développer, grâce en partie au chêne vert.

CONSEQUENCES DE L'HISTOIRE RECENTE DES FORETS SUR LA STRUCTURATION GENETIQUE DES ESPECES

L'étude de l'histoire récente des forêts est indissolublement liée à l'action de l'homme. La structuration génétique qui s'est mise en place après la dernière glaciation a été fortement perturbée par les déboisements et par les reboisements, qu'ils soient spontanés ou artificiels. Il existe une tradition anglo-saxonne de l'étude de l'histoire récente des forêts et de ses conséquences biologiques (un exemple est présenté figure 15, d'après Rackham, 1975, 1980). Toutefois, peu d'études ont utilisé des marqueurs moléculaires pour reconstituer l'histoire de l'action de l'homme sur les forêts.

L'étude de la variabilité de l'ADN chloroplastique des chênes en France[62] a permis une première analyse de l'action de l'homme sur la structuration génétique de ces espèces. La relation entre les caractéristiques du paysage autour de chacune des 869 forêts étudiées et le niveau de diversité chloroplastique a été analysé. La deuxième étude a consisté à caractériser parcelle par parcelle la variabilité chloroplastique des chênes dans une forêt domaniale pour laquelle on dispose d'un certain nombre d'informations historiques (époque des plantations).

Paysage forestier et structuration de la diversité chloroplastique chez les chênes

La physionomie forestière a été évaluée autour de chaque point d'échantillonnage en France, dans un rayon de 5 km, en estimant à partir des cartes topographiques : (1) la superficie boisée totale, (2) la superficie de la forêt ou du bois échantillonné, et (3) la proportion de carrés d'un km de côté qui étaient boisés. Des indices de fragmentation peuvent ainsi être déduits. Toutes ces caractéristiques reflètent directement l'action de l'homme sur un manteau boisé originellement essentiellement continu. Pour les 262 populations comprenant des chênes pédonculés, une régression multiple pas-à-pas a montré que la combinaison de l'indice de présence et de la superficie forestière totale expliquent significativement une

[61] Protandrie : chez les espèces monoïques (individus présentant à la fois des fleurs mâles et femelles), apparition des fleurs mâles avant les fleurs femelles.

[62] Latouche-Hallé C. 1998. Variabilité de l'ADN chloroplastique des chênes blancs en France. DEA de Biologie, Diversité et Adaptation des Plantes cultivées, Paris XI Orsay.

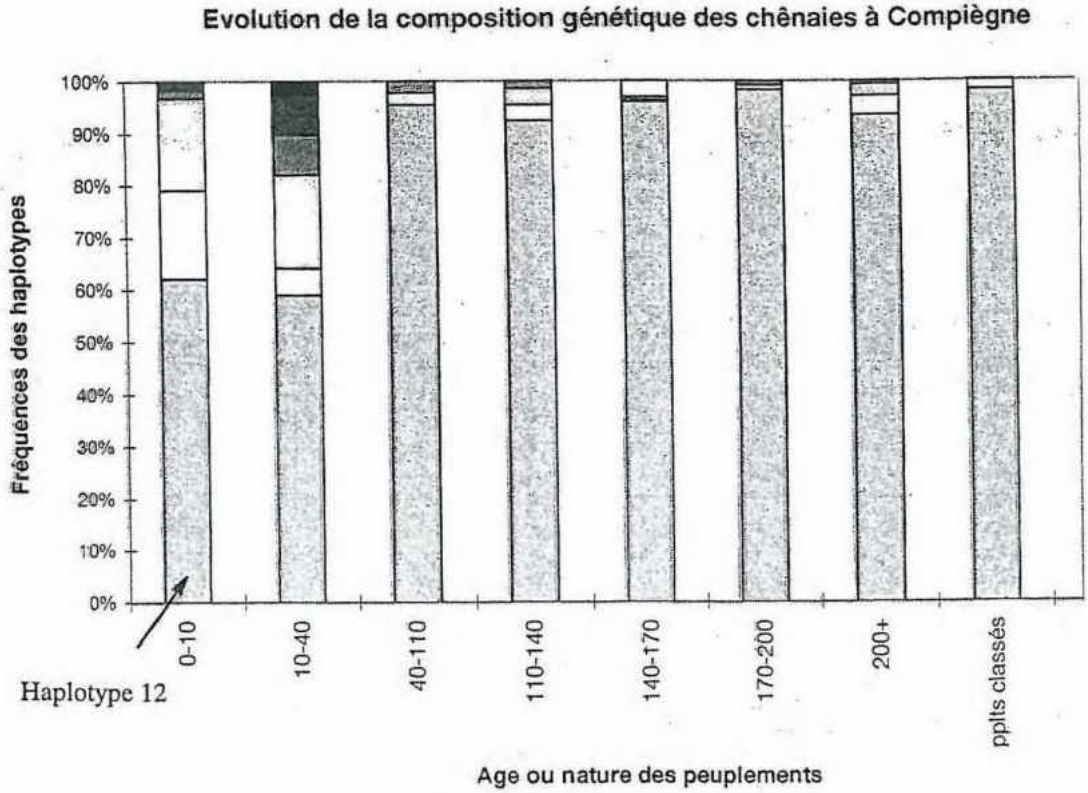


Figure 16 : Proportion de différents haplotypes chloroplastiques trouvés chez les chênes de la forêt domaniale de Compiègne, par classe d'âge. Les peuplements les plus vieux sont les plus homogènes génétiquement ; les haplotypes autres que le n°12 ne sont sans doute pas d'origine locale. Depuis 50 ans, les plantations utilisant des chênes qui ne sont pas originaires de Compiègne a considérablement augmenté. Les plantations, importantes depuis la fin du XVIIIème siècle, étaient vraisemblablement effectuées avec des plants produits dans une pépinière locale, avec des semences ramassées sur place.

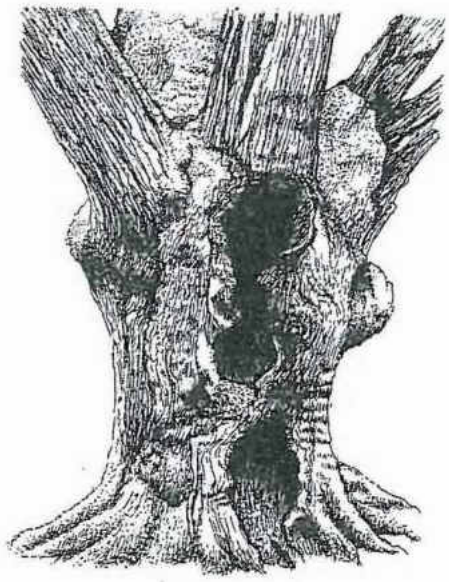


Figure 17 : Le chêne Saint Jean est le chêne (*Q. petraea*) le plus ancien de la forêt de Compiègne (800 ans environ) (illustration tirée de Clémence & Péron 1995). Tout comme les autres chênes remarquables de cette forêt, il est caractérisé par la variant majoritaire n°12, d'origine vraisemblablement locale.

(faible) part de la diversité : les forêts les plus grandes et les plus isolées sont légèrement plus variables que les autres. L'intensité de la sylviculture et d'une manière générale l'influence humaine doivent être particulièrement importants dans de telles forêts isolées. D'autre part, les reboisements en chênes sont surtout le fait des grands massifs domaniaux. Une autre manière d'influer sur la structuration génétique consiste à faciliter les contacts entre espèces différentes de chênes ; en effet, chez le chêne pubescent, la diversité intraspécifique est beaucoup plus grande dans les forêts mixtes (où plusieurs espèces de chênes cohabitent), que dans les chênaies pubescentes pures.

Pour préciser, à une échelle plus fine, l'impact de la sylviculture et notamment des plantations, une étude détaillée d'un massif forestier a ensuite été entreprise.

Variabilité génétique de l'ADN chloroplastique dans la forêt domaniale de Compiègne

Cette forêt de 14 500 hectares, à 80 km au nord-est de Paris, a fait l'objet de nombreuses plantations de chênes ; sous François 1er d'abord (parcelles de chênes sessiles des Beaux-Monts), puis sous Louis XIV, qui avait confié à Pannelier père et fils (choisis grâce à l'influence de Mme de Pompadour) la charge de reboiser plus de 3 000 hectares de 'vides' (landes incultes). En 200 ans (1774-1974), ce seront plus de 8 000 hectares qui seront reboisés artificiellement (Courtois 1975). Les reboisements de chênes par l'Office National des Forêts (il s'agit surtout de chênes sessiles, alors que la forêt est composée principalement de chênes pédonculés) se poursuivent à un rythme important.

Pour cette étude, nous avons procédé à un échantillonnage systématique ; 1642 arbres (dont 8 chênes pubescents, 473 chênes sessiles et 1160 chênes pédonculés) ont été étudiés. Sept haplotypes différents ont été trouvés, dont un (haplotype n°12) représente 87% des arbres (89% des chênes pédonculés, 84% des chênes sessiles et les huit chênes pubescents). Ces chiffres ont été précisés par classe d'âge (figure 16). Les neuf plus vieux chênes de la forêt (dont le chêne St Jean, estimé à 800 ans, voir figure 17) ont tous l'haplotype 12. Sur les 100 chênes provenant de peuplements classés porte-graines, seulement deux n'ont pas le variant 12. Les vieux peuplements sont constitués à 95% par l'haplotype 12. Par contre, dans les classes d'âge les plus récentes (<50 ans), le nombre d'haplotypes différents (voire exotiques, c'est à dire rares ou absents en France) augmente pour atteindre près de 50%. L'absence de structuration spatiale des différents haplotypes au sein du massif de Compiègne conforte l'hypothèse que le variant 12 est le seul variant chloroplastique autochtone (la forêt de Retz, qui commence seulement à trois kilomètres au sud-est de la forêt de Compiègne, est par contre composée presque exclusivement du variant 10, dont la fréquence à Compiègne est de l'ordre de 5%, distribués aléatoirement).

L'étude de plantations récentes à permis de mettre en évidence, dans plusieurs cas (plantations des années 30 à 60 notamment), de véritables *patchworks* d'haplotypes (jusqu'à cinq par plantation), contrastant avec l'uniformité génétique des plus vieux peuplements. Ici, en l'occurrence, l'intervention humaine se traduit par une **augmentation de la diversité** intra-forêt. L'étude détaillée d'une parcelle où des chênes de Slavonie (Croatie) avaient été plantés en mélange avec du matériel local est en cours. Puisque la différenciation entre les deux origines est totale, et qu'il s'agit de la

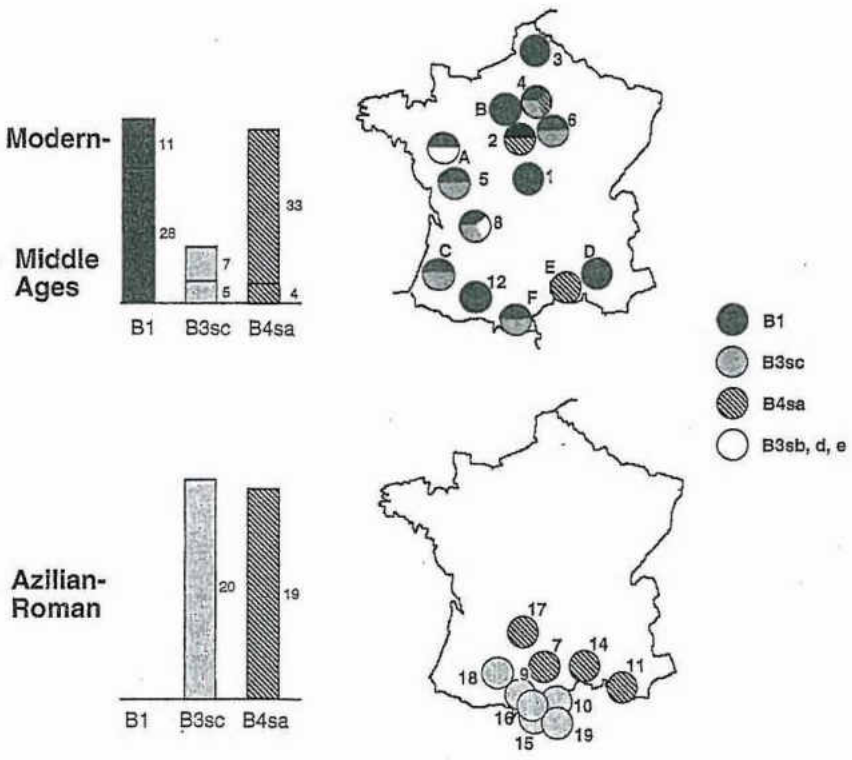


Figure 18. Structuration génétique de l'ADN mitochondrial chez le lapin de garenne : du Moyen-Âge à l'époque moderne (en haut) et de l'époque Azilienne (11800 BP) jusqu'à l'époque Romaine (1600 BP), basée sur l'analyse de restes (os) fossiles. Noter la forte structuration qui existait mais a ensuite disparu. Les histogrammes sur la gauche donnent la composition par type mitochondrial. D'après Hardy et coll. (1995).

première génération (arbres de 70 ans), l'attribution de chaque arbre à l'une des provenances est possible. Ces travaux, dans une forêt réputée pour avoir été particulièrement perturbée par l'homme (importance des plantations), démontrent que c'est seulement récemment que la structure génétique a été modifiée. En fait, l'existence jusqu'au siècle dernier de pépinières forestières au sein même du massif a dû limiter l'introduction de matériel d'origine non locale, si les forestiers s'approvisionnaient sur place en graines, ce qui semble vraisemblable.

LA PALEOGENETIQUE, UN NOUVEL OUTIL POUR ETUDIER L'HISTOIRE DES ARBRES FORESTIERS

Contrairement à l'étude des pollens ou macrorestes fossiles, qui permettent d'accéder à des informations par nature historiques, les interprétations diachroniques basées sur les marqueurs moléculaires sont indirectes car basées sur l'étude des populations d'arbres contemporaines. Elles valorisent le fait que les génomes sont de fantastiques machines à stocker de l'information. Même si on est loin d'avoir extrait toute l'information historique contenue dans le génome des individus vivant actuellement, on peut d'ores et déjà identifier les possibilités qui s'ouvriraient si on pouvait également étudier l'ADN ancien contenu dans les sédiments Quaternaires. Selon Villablanca (1994), « *aDNA [ancient DNA] forges an empirical link between history and two inherently historical fields of research [phylogenetics and population genetics]* ». Quelques exemples, y compris une étude pilote réalisée au laboratoire, permettront d'illustrer les potentialités de ce type d'analyse et les difficultés techniques qui se posent.

Une étude récente de paléogénétique chez le lapin de garenne en France et en Espagne illustre bien l'intérêt de tels travaux (Hardy et coll. 1995, Loreille et coll. 1997). La structuration actuelle de l'espèce en France, étudiée à l'aide d'un fragment d'ADN mitochondrial de 233 pb situé sur le gène cytochrome *b*, est pratiquement nulle. Toutefois, des restes fossiles abondent dans de nombreux sites archéologiques, datant de 1 600 à 10 000 ans, voire plus. L'analyse de la même séquence mitochondriale a permis de montrer que dans le passé les populations de lapins du sud de la France étaient extrêment stables et fortement structurées génétiquement (figure 18). L'époque à partir de laquelle cette structure a commencé à disparaître a pu être datée approximativement (entre l'époque romaine et le Moyen Âge). C'est l'introduction d'un variant auparavant trouvé seulement en Espagne, qui devait correspondre aux premières souches domestiquées, qui est à l'origine de cette perte de structuration.

On peut constater que la plupart des déductions effectuées sur l'histoire de cet animal n'auraient pu être faites sans l'analyse d'ADN ancien. Chez des arbres forestiers ayant été fréquemment plantés, la structuration initiale pourrait ainsi être retrouvée, permettant ainsi l'identification des populations contemporaines d'origine réellement locale –s'il en subsiste–. Des études de moindre ampleur ont été réalisées chez des plantes (ex : Golenberg 1994 (feuilles fossiles extraites de sédiments lacustres), Rollo et coll. 1994 (graines sèches de céréales), Suyama et coll. 1996 (grains de pollen de sapins)). Un point important souligné par Rollo et coll. (1994) est la disponibilité d'amorces PCR adaptées à ce genre d'étude et suffisamment variables.

L'étude des génomes cytoplasmiques, parce qu'ils sont présents en de nombreuses copies, devrait permettre d'obtenir de meilleurs résultats que celle de gènes nucléaires non répétés.

La connaissance acquise sur les variations de séquences de l'ADN chloroplastique du chêne nous a fournit l'occasion d'adapter des amorces afin d'amplifier de l'ADN ancien. Une séquence de 1700 pb a été obtenue pour une vingtaine d'haplotypes chloroplastiques chez les chênes (Dumolin-Lapègue 1998). Des amorces permettant d'amplifier 290 pb ont été mises au point et employées avec succès pour l'amplification d'ADN issu de bois de chêne vieux de 600 ans (pieu d'un barrage)[63]. Nous avons pu montrer que le fragment de bois ancien n'était probablement pas d'origine locale, car les arbres poussant sur le site actuellement avaient tous un autre haplotype.

[63] Dumolin-Lapègue S, L Gielly, P Taberlet et RJ Petit. Amplification of DNA from ancient and modern oak wood. Soumis à *Mol. Ecol.*

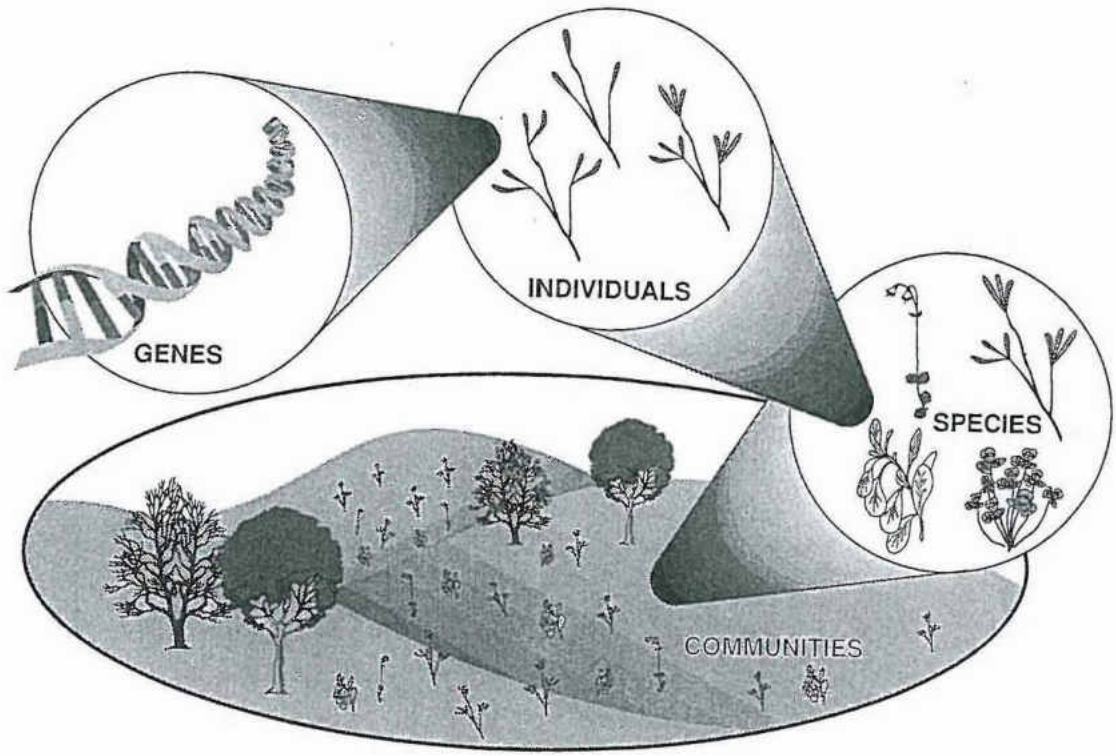


Figure 1 : Les différents niveaux de la diversité, des gènes aux communautés (*in* Frankel et coll. 1995 : *The conservation of plant biodiversity*).

Chapitre 4

Perspectives en matière de gestion et de conservation des ressources génétiques

« La diversité est la base de l'adaptabilité des êtres vivants et peut-être des systèmes écologiques et de la biosphère tout entière, face aux changements qui peuvent affecter leur environnement ». R. Barbault 1992.

« Puisqu'il y a des raisons de penser que la diversité convient à la nature et puisque nous aimons la diversité, [...] on peut donc poser la diversité biologique en norme positive. Elle peut servir de critère pour évaluer l'impact des interventions et des activités humaines sur les systèmes écologiques. Critère objectif, puisqu'il fournit des indications sur le fonctionnement du système considéré... » C. & R. Larrère (1997). Du bon usage de la nature. *Aubier, Paris.*

INTRODUCTION

La biodiversité est un concept récent qui intègre la variété du vivant à tous ses niveaux : les gènes, les populations et les espèces, les écosystèmes et les processus naturels qui assurent la perpétuation de la vie sous toutes ses formes (figure 1). Le niveau qui nous concerne plus particulièrement ici est la variabilité entre populations (niveau intraspécifique), et constitue une partie importante mais généralement négligée dans la constitution des espaces protégés (parcs, réserves naturels). Fréquemment d'ailleurs, les différents niveaux sont étudiés séparément par des spécialistes de disciplines différentes et les résultats restent mal intégrés. Une meilleure synthèse devra émerger à l'avenir, et certaines perspectives que j'évoquerai ici vont dans ce sens.

L'objectif de cette dernière partie est de présenter tout d'abord quelques travaux sur la manière d'identifier les populations pour la conservation, à l'aide de marqueurs moléculaires. J'illustrerai ensuite la discordance entre les deux options qui ont cours actuellement pour conserver la (bio)diversité, l'approche des écologistes, plus globale, basée sur la richesse en espèces ou la diversité des écosystèmes, et celle des généticiens, qui considèrent une seule espèce à la fois. Ce sera l'occasion de présenter un projet européen que j'anime, conçu pour essayer de rapprocher ces deux perspectives. L'échelle dans ce cas est celle du continent, afin d'identifier les grandes priorités en matière de conservation à ce niveau.

L'approfondissement de notre connaissance sur les mécanismes de recolonisation postglaciaire et l'effet des translocations de populations pourra être utile à la compréhension des risques liés à un réchauffement rapide de la planète. Pour la gestion plus locale, puisque l'histoire semble être le facteur clé responsable des niveaux de diversité et de l'existence d'une structuration, il deviendra indispensable

d'associer les connaissances disponibles sur l'évolution de l'occupation des sols, la rapidité du renouvellement des forêts et l'identification des fragments boisés les plus anciens, à la description désormais plus classique de la diversité génétique des espèces forestières.

DIFFERENCIATION, FLUX DE GENES ET CONSERVATION

Différenciation génétique et flux de gènes

Dans le chapitre 3, j'avais illustré les conséquences de plantations de chênes en forêt de Compiègne : la diversité intrapopulation était considérablement augmentée, du fait de ces flux de gènes artificiels. L'idée qu'il faut faciliter les flux de gènes pour maintenir la diversité est solidement ancrée dans l'esprit de nombre de généticiens et d'écologistes et dans celui des gestionnaires. A première vue, l'action de l'homme semble en effet entraîner une rupture des échanges entre populations et donc une perte de diversité génétique : la fragmentation des milieux naturels limite les possibilités d'échanges, la réduction de leur taille engendre une baisse des effectifs des populations qu'ils hébergent, donc une dérive accrue.

Toutefois, pour beaucoup d'espèces, une diminution de taille des populations et de densité, voire un isolement accru vont en fait augmenter les flux de gènes et la fréquence des perturbations. Tout ceci sera exacerbé pour les espèces, très nombreuses dans les pays tempérés, qui sont l'objet de plantations, et d'introductions volontaires ou involontaires par l'homme. Ellstrand et Ellam (1993) rappellent que, chez les espèces cultivées, des expérimentations ont démontré une augmentation des flux de gènes dans les petites populations : « *In conclusion, small populations are expected to receive gene flow at a higher rate than large populations and are more likely to receive gene flow from large populations than from other small ones, even if the latter are in closer proximity* ». Donc, les petites populations risquent non pas de disparaître ni de perdre de la diversité, mais au contraire de se maintenir, voire de se développer, mais en perdant leur 'originalité' génétique, en se banalisant ; ce seront en effet les plus grandes populations, même lointaines, qui vont contribuer désormais à les alimenter en allèles.

La taille efficace d'une population subdivisée, dans le modèle en île à l'équilibre, est augmentée d'un facteur $1/(1-F_{ST})$ relativement à une population de même taille non subdivisée (Barton et Whitlock 1997). Selon ces auteurs, « *This is because in the standard island model, the reproductive success of each population is fixed* ». Ceci vaut pour l'hétérozygotie (qui sert de base au calcul du F_{ST}). Les résultats équivalents ne sont pas disponibles pour la richesse allélique ; on ignore donc l'effet de la subdivision sur la conservation du nombre d'allèles. De même, la relation correspondante n'est pas connue pour le paramètre N_{ST}. Cependant, on peut penser que le même type de relation existe. Or, nous avons démontré que les deux mesures de subdivision correspondantes (A_{ST} et N_{ST}) étaient presque toujours plus élevées que le G_{ST} ou F_{ST}. En outre, la différenciation multilocus, qui tient compte des déséquilibres gamétiques (relations entre allèles de locus différents), est également supérieure à la différenciation monolocus moyenne correspondante (Kremer et coll. 1997). Dans les trois cas, une meilleure prise en compte de l'information moléculaire disponible conduit donc à une augmentation de l'estimation du coefficient de subdivision.

Tout ceci conduit à une sérieuse remise en cause de formulations rapides telles que « seulement 2% (ou 5% ou 10%) de **la** diversité réside entre les populations », puisqu'il existe plusieurs concepts et mesures de diversité, qui affectent à chaque fois le coefficient de différentiation.

On peut donc penser que l'augmentation des flux de gènes au delà du niveau naturellement présent dans les populations menace autant sinon plus la diversité globale (mesurée sur l'ensemble de l'espèce) qu'une diminution des flux de gènes (en dessous de ce niveau de référence). Il me semble en tout cas que les deux faces du problème (défaut et excès de flux de gènes) devraient être considérées systématiquement dans les politiques de conservation. De même que les écologistes ont à l'esprit des plantes rudérales, qui contribuent à banaliser un écosystème, voire à le menacer si elles deviennent trop envahissantes (même si dans un premier temps leur arrivée augmente la richesse spécifique), les gestionnaires des ressources génétiques devraient s'inquiéter de l'arrivée massive d'allèles déjà largement représentés ailleurs et de la disparition du caractère spécifiquement local du pool génétique initial. Outre le maintien du niveau de diversité, Ellstrand et Elam (1993) signalent d'autres problèmes potentiels liés à une augmentation du flux de gènes : l'''*outbreeding depression*' (quand les croisements avec des populations divergentes produisent des hybrides moins bien adaptés localement), et la diminution de la valeur sélective locale de la population. Ils concluent que « *In most cases, recognition and consideration of gene flow as a potential hazard by plant conservation decision-makers will prevent future problems such as costly, unnecessary, and potentially problematic transplantation projects* ». Leur analyse portait sur les espèces rares ; pour les arbres forestiers, la motivation derrière ces projets de transplantation est évidemment tout autre.

Même si les conséquences de ces flux de gènes étaient faibles pour l'adaptation des arbres forestiers (du fait de la sélection naturelle qui peut éliminer les plantules moins bien adaptées, vu les fortes densités initiales), la perte de diversité pourrait devenir une réalité, notamment si des plantations ont lieu dans les zones refuges déjà naturellement assez morcelées et fragmentées. Ces plantations, à terme, peuvent gommer la structuration génétique préexistante (c'est le cas du pin maritime au Portugal par exemple, ou des chênes dans le nord de l'Allemagne, etc.).

Or cette structuration (outre les reconstitutions historiques qu'elle permet !) peut être valorisée. Par exemple, dans le cas des chênes, la forte structuration en tâches de l'ADN chloroplastique va nous permettre de mettre en place un suivi (traçabilité) du bois issu de ces régions[64], qui va bénéficier aux gestionnaires des forêts (ONF, forêts privées) et à l'industrie de la tonnellerie, car il existe une demande pour des produits d'origine régionale.

En effet, le bois de merrain (servant à la fabrication des barriques destinées à l'élevage du vin) se vend à des prix très différents selon son origine et ses qualités, qui dépendent de facteurs écologiques et sylvicoles et peut être également génétiques. La recherche de produits plus différenciés (par espèce de chêne, type de sylviculture ou origine géographique) permettra de disposer d'une gamme plus large pour l'élevage du vin. Certains produits pourront alors être associés à des crûs particuliers de vins, le tout correspondant à une tendance vers une

[64] Thèse en cours de M-F Deguilloux (bourse Cifre) et recrutement de M-H Pemonge, dans le cadre d'un projet en partenariat avec le Centre Technique du Bois et de l'Ameublement.

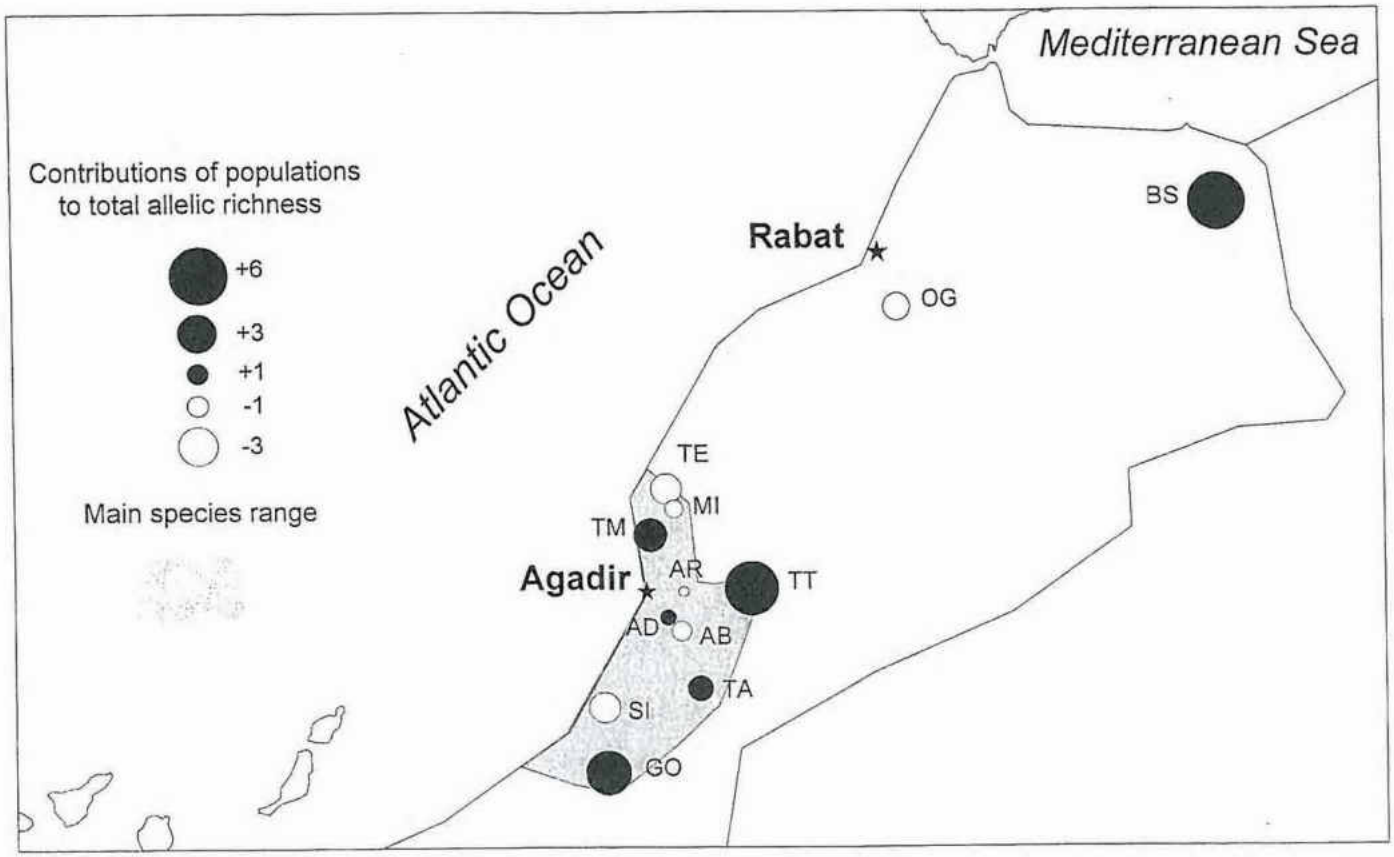


Figure 2 : Contribution de chacune des 12 populations d'arganier étudiées à la richesse allélique totale (exprimé en pourcentage du total). Les contributions positives sont exprimées en noir, les contributions négatives en blanc. L'isolat du nord-est (BS), d'origine vraisemblablement anthropique, contribue le plus à la richesse allélique.

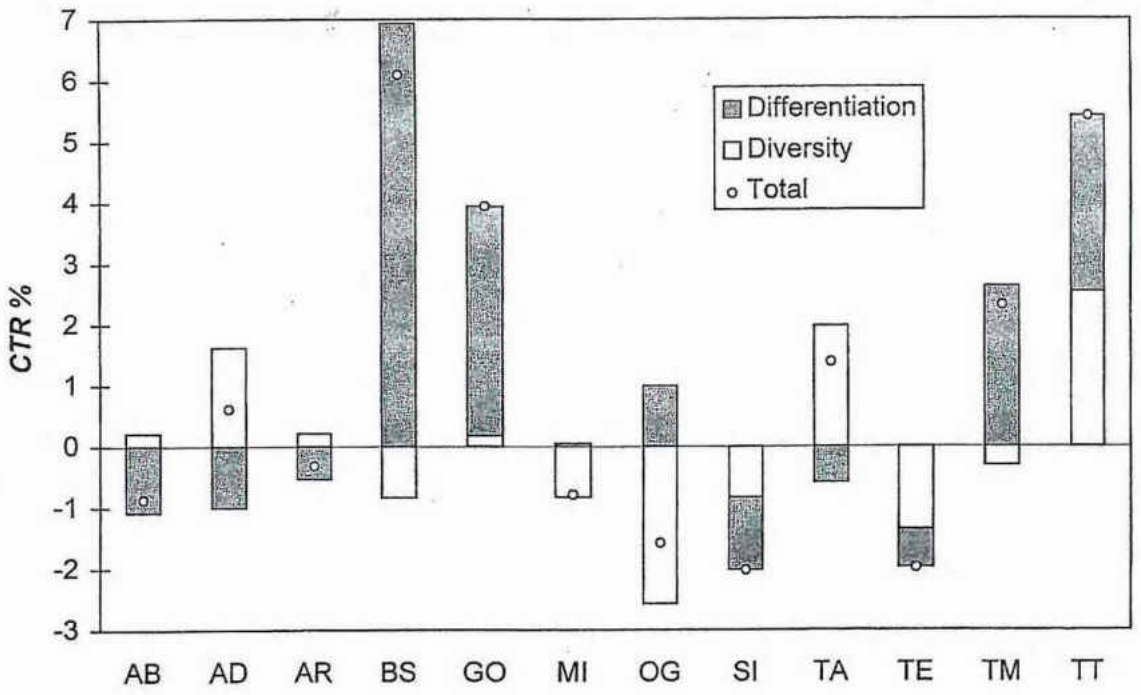


Figure 3 : Décomposition de la contribution à la richesse allélique de chaque population en deux composantes : contribution liée à la diversité de la population et contribution liée à sa divergence. L'isolat du nord-est est moins variable mais plus divergent que les autres populations, et c'est ce dernier terme qui prédomine.

demande pour des produits authentiques, 'du terroir'. La plus-value qui en résultera disparaîtrait si des plantations systématiques avec des plants de chênes d'autres provenances modifiaient la structure actuelle.

Identification des populations pour la conservation

Les paramètres utilisés en génétique des populations n'ont pas été conçus *a priori* pour permettre d'optimiser la conservation des ressources génétiques. Toutefois, dans la mesure ou l'étude de 'diversité' n'est pas purement descriptive mais a pour objectif d'identifier les principaux facteurs de l'évolution, elle devrait permettre d'améliorer les politiques de conservation des ressources. J'ai proposé une manière de synthétiser les résultats des analyses moléculaires adapté aux décisions en matière de conservation, et qui peut également être utile pour interpréter les processus en cause (Petit et coll. 1998[65]). La méthode permet d'estimer la contribution d'une population à la diversité totale (figure 2). N'importe quel paramètre de diversité jugé approprié peut être utilisé. De plus, cette contribution peut être décomposée en deux parties : part liée à la diversité de la population (qui peut être supérieure ou inférieure à la moyenne, donc positive ou négative), et part liée à sa divergence des autres populations (figure 3). Il semble en particulier intéressant d'utiliser dans ce contexte les mesures de richesse allélique (à l'aide de la technique de raréfaction, exposée au chapitre 1).

Par exemple, chez l'arganier *Argania spinosa* (Sapotacée), espèce endémique du Maroc, en forte régression, nous avons montré que les populations du nord de l'aire, très excentrées, considérées jusque là comme des peuplements 'reliques', avaient plus vraisemblablement une origine récente (translocation par l'homme, *cf.* El Mousadik et Petit 1996[66]).

En ce qui concerne les isozymes, la richesse allélique des deux populations avait chuté ; par contre elles avaient fortement divergé, parfois pour des allèles rares ou absents ailleurs ; en conséquence, l'une d'elles, de plus grande taille, contribuait plus que n'importe quelle autre population d'arganier à la diversité totale de l'espèce, démontrant l'effet positif de l'isolement, au moins pour ce critère. Ces tentatives pourraient être complétées par la mise au point d'algorithmes destinés à identifier des sous-ensembles de populations candidates à un statut de conservation particulier, comme cela se fait pour l'identification des *core collections* chez les plantes cultivées (Bataillon et coll. 1996).

GENETIQUE MULTI-ESPECES, ZONES REFUGES ET PRIORITES EUROPEENNES

Les généticiens forestiers mettent généralement en avant la conservation des ressources génétiques. Par contre, les écologistes sont souvent persuadés que la conservation des écosystèmes est prioritaire (Frankel et coll. 1995). Comment réconcilier ces deux positions ? Il est clair que conserver les ressources génétiques d'une espèce en faisant abstraction de son écosystème (conservation *ex situ*) est un pis-aller. Par contre, une approche plus ciblée, pour des espèces présentant un intérêt particulier pour

[65] Petit R J, A El Mousadik et O Pons. 1998. *Identifying populations for conservation on the basis of genetic markers. Cons. Biol.* **12**, 844-855.

[66] El Mousadik A et R J Petit. 1996. Chloroplast DNA phylogeography of the argan tree of Morocco. *Mol. Ecol.* **5** : 547-555.

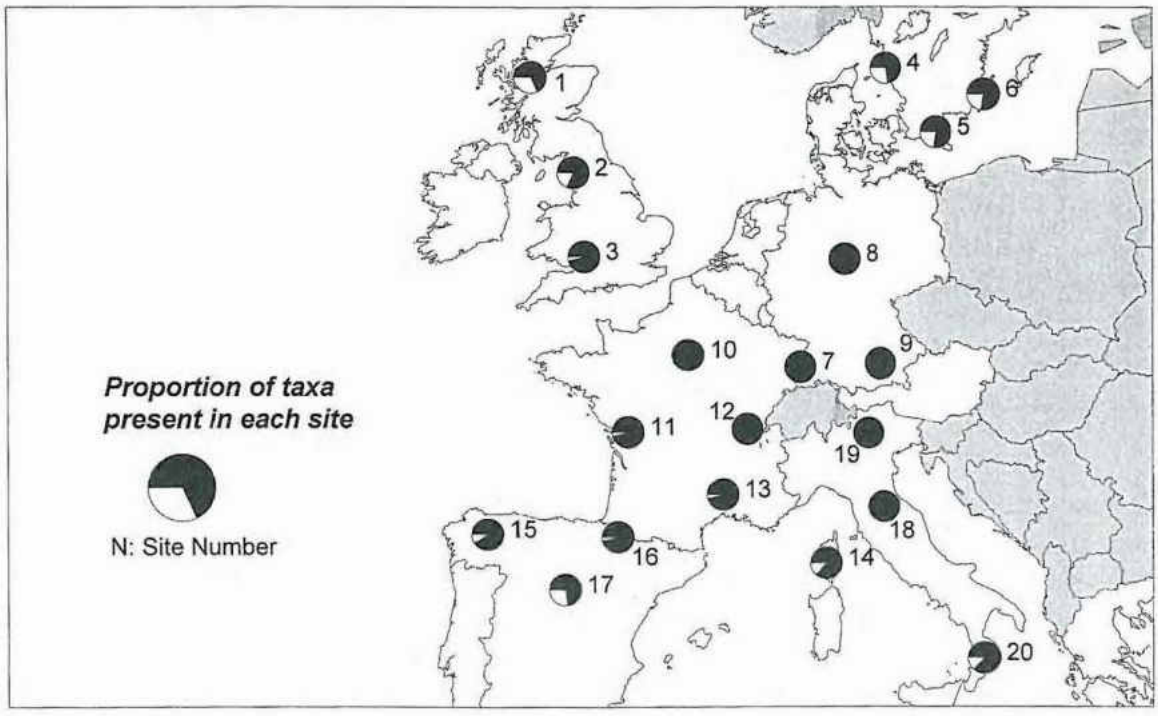


Figure 4 : Distribution géographique des 20 forêts étudiées dans le projet CYTOFOR, et proportion des 22 espèces suivantes présentes en chaque site.

Espèces	**Hybridation possible avec :**	**Famille**	**Reproduction**	**Dispersion des graines**		**Pollinisation**	
				Vent	**Anim**	**Vent**	**Insect**
Acer campestre	*A. monspessulanum*	Aceraceae	Polygame	+			+
Acer pseudoplatanus		Aceraceae	Polygame	+			+
Alnus glutinosa	*A. incana & cordata*	Betulaceae	Monoïque	+		+	
Betula pendula	*B. pubescens*	Betulaceae	Monoïque	+		+	
Calluna vulgaris		Ericaceae	Hermaphrodite	?			+
Carpinus betulus		Corylaceae	Monoïque	+		+	
Corylus avellana		Corylaceae	Monoïque		+	+	
Crataegus monogyna	*C. laevigata*	Rosaceae	Hermaphrodite		+		+
Cytisus scoparius		Leguminosae	Hermaphrodite		+		+
Fagus sylvatica		Fagaceae	Monoïque		+	+	
Fraxinus excelsior	*F. angustifolia*	Oleaceae	Polygame	+		+	
Hedera helix		Araliaceae	Hermaphrodite		+		+
Ilex aquifolium		Aquifoliaceae	Polygame		+		+
Populus tremula	*P. canescens P. alba*	Salicaceae	Dioïque	+		+	
Prunus avium	*cultivars of cherry*	Rosaceae	Hermaphrodite		+		+
Prunus spinosa		Rosaceae	Hermaphrodite		+		+
Quercus robur	*Q. petraea, Q. pubescens*	Fagaceae	Monoïque		+	+	
Rubus fruticosus		Rosaceae	Hermaphrodite		+		+
Salix caprea	*S. aurita, cinerea...*	Salicaceae	Dioïque	+			+
Sorbus torminalis	*S. aria*	Rosaceae	Hermaphrodite		+		+
Tilia cordata	*T. platyphyllos*	Tiliaceae	Hermaphrodite	+			+
Ulmus minor	*U. glabra*	Ulmaceae	Hermaphrodite	+			+

Table 1. Caractéristiques biologiques des 22 espèces étudiées dans le cadre du projet.

l'homme, pourra être plus efficace. Dans beaucoup de pays, en effet, les parcs nationaux sont situés dans des zones extrêmes, difficilement exploitables par l'homme et pas forcément représentatives de la diversité génétique des populations d'arbres de plaine. Une manière de rapprocher les conceptions des généticiens et des écologistes est de travailler avec les outils des généticiens (marqueurs moléculaires dans le cas présent) sur un grand nombre d'espèces, afin de voir si des tendances générales se dégagent, à l'échelle géographique considérée (le continent). Les zones de fort endémisme ou particulièrement riches en espèces sont-elles aussi d'un intérêt particulier pour la conservation des ressources génétiques ? Peut-on prédire, pour des espèces dont on ne connaît que les attributs biologiques (mode de dispersion des graines, du pollen...), le niveau et la distribution de la diversité génétique, et donc avoir un raisonnement génétique applicable en routine dans les opérations de conservation ?

Pour tenter de répondre à ces interrogations, j'ai donc proposé, avec des collègues de huit autres pays, de travailler en parallèle sur un grand nombre d'arbres et d'arbustes, dans le cadre d'un projet soutenu par l'Union Européenne[67].

Nous avons sélectionné 22 espèces (plus de nombreuses espèces apparentées), qui vont être échantillonnées dans les même sites (20 forêts dans un premier temps, réparties en Europe). Leur diversité sera étudiée avec les mêmes outils, pour faciliter les comparaisons. Les marqueurs retenus dans cette étude pilote (j'ai qualifié ce genre de travaux de *génétique des populations multi-espèces*) sont les polymorphismes de l'ADN chloroplastique et mitochondrial. Nous souhaitons en effet examiner si d'autres espèces que les chênes se prêtent à des études de traçabilité géographique (vu notamment les perspectives en matière d'écocertification[68]). La carte des forêts étudiées est donnée en figure 4, ainsi que la liste des espèces concernées (table 1).

MIGRATIONS ET CHANGEMENTS CLIMATIQUES

En savons nous assez pour prédire ce qui se passerait concernant la distribution des arbres forestiers en cas de changement climatique ? Pitelka et coll. (1997) posent cette question et précisent que les prédictions simplistes qui ne tiennent pas compte des capacités réelles de migration des espèces dans un paysage rendu souvent inhospitalier par l'homme risquent d'aboutir à des conclusions erronées. L'étude détaillée des mécanismes de migration des plantes est nécessaire pour prédire les conséquences de ces changements.

Les résultats présentés dans le chapitre 3 montrent que l'étude rétrospective de l'effet du réchauffement à l'Holocène peut fournir des informations de base pour prédire les changement futurs. Il faut aboutir à des scénarios de migration de plus en plus réalistes, qui rendent compte non seulement de la progression physique des espèces mais aussi simultanément des modifications de leur structure génétique, la description de l'un enrichissant la compréhension de l'autre. Les études futures ne devront pas seulement porter sur les populations les plus nordiques, en phase d'expansion. Elles

[67] Petit R J (coordinateur). CYTOFOR : *Measuring molecular differentiation of European deciduous forests for conservation and management.* Projet FAIR soutenu par l'Union européenne 1998-2001.

[68] Genre de label vert pour les produits forestiers, garantissant que les forêts dont ils sont issus sont gérées durablement.

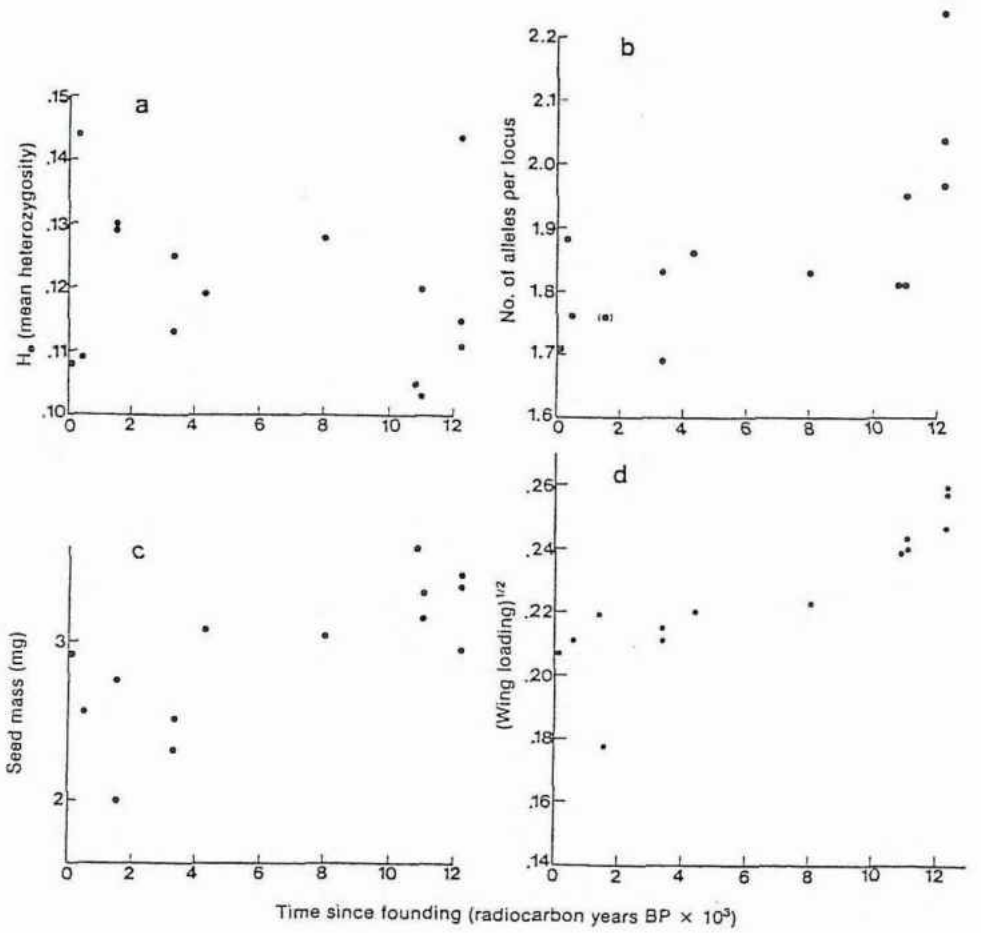


Figure 5 : Relation entre le temps qui s'est écoulé depuis la fondation d'une population et divers caractéristiques de la population : hétérozygotie, nombre d'allèles par locus, poids et portance des graines, chez *Pinus contorta*. D'après Cwynar & McDonald (1987).

devront s'intéresser aussi aux populations restées dans les refuges et qui ont constitué jusqu'ici le 'camp de base' permanent des espèces forestières au cours des oscillations climatiques du Quaternaire. Ces zones refuges apparaissent à la fois plus riches et plus menacées. Les populations qui s'y trouvent pourront-elles subsister plus en altitude, si le climat se réchauffe ?

Un sujet encore sous-exploré est la relation entre migration et caractères adaptatifs, particulièrement importante dans un contexte de conservation. Il serait notamment utile d'évaluer les conséquences de migrations à venir sur l'adaptation des espèces. Cwynar et MacDonald (1987) ont mis en évidence une sélection associée à la dispersabilité des graines (les graines ailées du pin qu'ils étudiaient avaient une portance plus importante dans les populations les plus récentes ; *cf.* figure 5). Est-ce un cas rare et isolé, ou d'autres exemples existent-ils, y compris pour d'autres caractères adaptatifs, plus importants pour le forestier ?

Même quand on dispose de scénarios détaillés de la recolonisation postglaciaire, avec idéalement la date d'apparition de l'espèce dans chaque forêt, des corrélations éventuelles entre ancienneté de la population et caractères adaptatifs resteront difficiles à interpréter. Par exemple, des tests de provenance ont démontré que les populations de hêtre du sud-est de l'Europe (là où se trouvent les forêts de hêtre les plus anciennes) étaient plus précoces (von Wuehlisch et coll. 1995).

Des résultats similaires ont été trouvés pour le chêne sessile (Kremer et coll. 1998[69]).

Cette structuration pour la phénologie est-elle une conséquence de la colonisation, ou l'effet de gradients climatiques ? D'une manière générale, de nombreux caractères adaptatifs sont contrôlés par le climat et varient donc latitudinalement et/ou longitudinalement (exemple : figure 6). Comme les zones refuges sont situées dans les régions méridionales, les effets de la sélection et de l'histoire ne peuvent être facilement distingués. Toutefois, nos études montrent que la colonisation postglaciaire n'a pas consisté en des mouvements uniformes vers le nord. La possibilité de distinguer les deux effets (histoire et sélection) est donc réelle. Une analyse des associations entre marqueurs moléculaires (ADN chloroplastique, considéré comme un marqueur de l'histoire) et caractères adaptatifs pourrait donc permettre d'évaluer l'importance relative des deux facteurs et l'effet résiduel de l'histoire sur la structuration de caractères adaptatifs (Kremer et coll. 1998).

CONSERVATION ET GESTION DES FORETS ANCIENNES

Dans nos régions tempérées, les forêts semblent constituer un élément immuable du paysage, et être issues telles quelles du fond des âges. En étudiant de près leur histoire, notamment grâce aux cartes topographiques ou aux plans de propriétés plus anciens, force est de constater que le taux de renouvellement des forêts est en fait très important. Vers 1800, seulement environ 10% du territoire métropolitain était couvert de forêts, contre près de 27% actuellement. Il

[69] Kremer A, R J Petit, A Ducousso, V Le Corre (1998) General trends of variation of genetic diversity in *Quercus petraea* (Matt.) Liebl. *In*: Diversity and adaptation in oak species, KC Steiner, ed. Proceedings of the second meeting of Working Party 2.08.05, Genetics of *Quercus*, of the International Union of Forest Research Organizations. October 12-17 1997. University Park (State College), Pennsylvania, USA

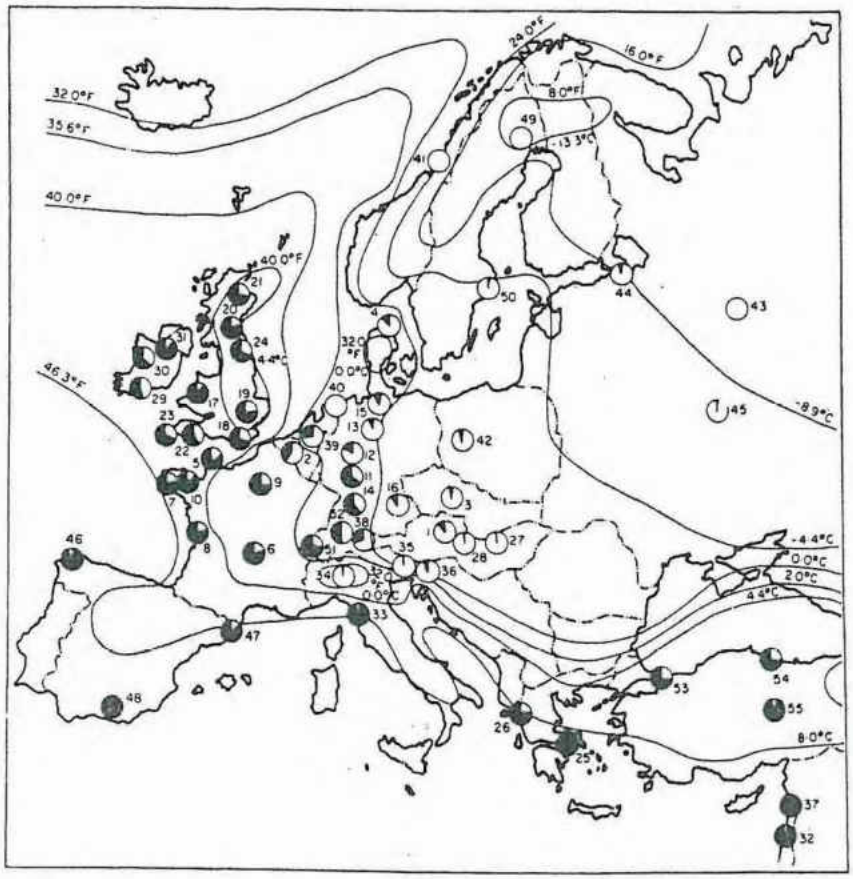


Figure 6 : Exemple de caractère dont la distribution géographique semble déterminé par la sélection (et non l'histoire) : fréquence des variants cyanogènes chez *Trifolium repens* en Europe. D'après Daday, 1954, in Briggs & Walters (1984). Les isothermes de janvier sont également indiqués.

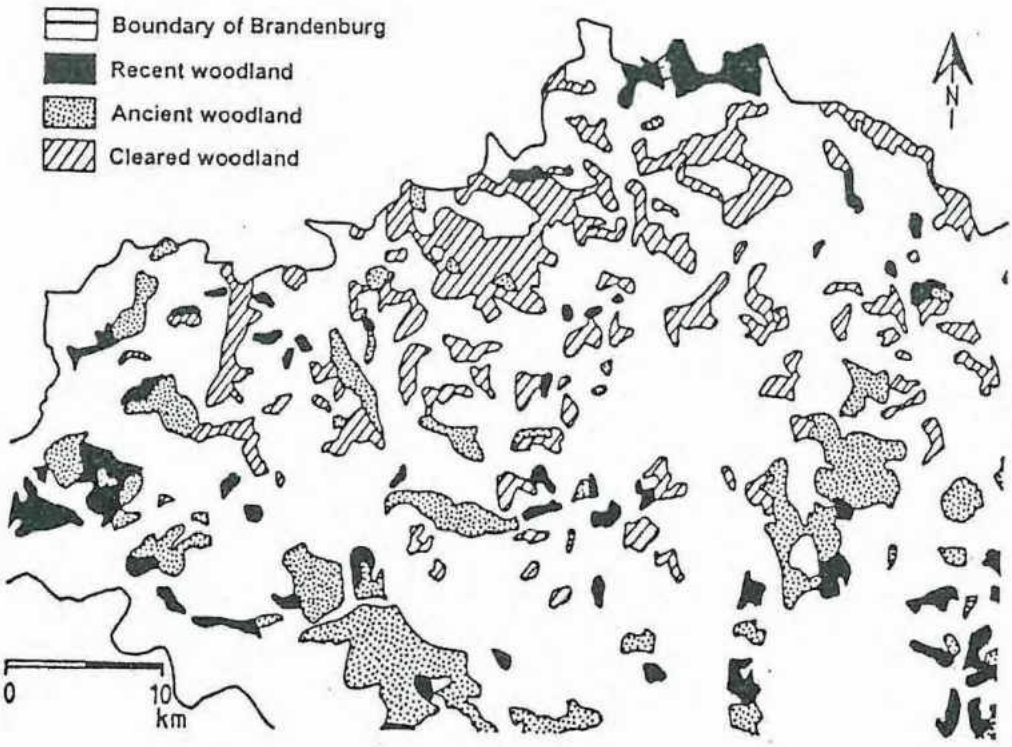


Figure 7 : Distribution des zones de forêts anciennes, de reboisement et de déboisement dans le nord-ouest de la région de Prignitz (Brandenburg), entre 1780 et 1937 (Wulf 1998).

ne faudrait absolument pas en conclure que un bon tiers des forêts actuelles sont 'anciennes'. La distribution des zones agricoles et forestières a évolué au cours du temps et a été influencée par de nombreux facteurs, naturels (qualité des sols, reliefs) ou sociaux (croissance démographique, industrialisation, etc.). Grâce à un intérêt assez récent pour ces études historiques, des statistiques permettant d'estimer les taux de renouvellement des forêts au cours des derniers siècles commencent à être disponibles. En outre, des bases de données reliées à des systèmes d'information géographiques permettant d'identifier les fragments de forêts anciennes et de visualiser les changements de distribution des zones boisées au cours du temps et à différentes échelles spatiales, sont en cours de constitution (ex : Koerner et coll. 1997, Wulf 1998) (figure 7).

Ces travaux ont été initiés notamment par des géobotanistes et des écologistes s'intéressant à la composition floristique de ces forêts. De l'ancienneté des forêts dépend la présence de certaines espèces de plantes (espèces dites indicatrices de forêts anciennes, Peterken 1974, Wulf 1997). Il s'agit essentiellement d'espèces herbacées, par exemple en Europe du nord : *Carex sylvatica*, *Galium odoratum*, *Paris quadrifolia*, *Melica uniflora*, etc., bien que certains arbres soient parfois cités (*Prunus avium*). Le nombre d'espèces indicatrices citées dans chaque région où une telle étude a été menée est beaucoup plus élevé, mais le degré d'association avec les forêts anciennes varie souvent d'un pays à l'autre. En fait, pour les espèces les plus rares, il est impossible de tester l'association avec l'âge de la forêt. Même si la taille des forêts, la nature des sols et les facteurs climatiques contribuent à expliquer la distribution des espèces, l'âge des forêts reste l'un des facteurs les plus importants à prendre en compte (Wulf 1997). La principale explication pour l'absence des espèces caractéristiques des forêts anciennes dans les forêts les plus récentes semble être leur faible capacité de migration à longue distance : en effet des bois récents situés à proximité immédiate de forêts anciennes sont seulement très progressivement envahis par ces espèces. En particulier, les espèces myrmécochores (disséminées par les fourmis) et autochores (absence de mécanisme particulier de dispersion) sont surreprésentées parmi ces espèces indicatrices ; toutefois ces catégories basées sur la morphologie des graines sont peu précises, et des stratégies mixtes peuvent exister. Ces conclusions mériteraient donc d'être validées par une autre approche.

Les études sur la distribution des plantes forestières sont importantes pour la conservation de la biodiversité. En effet, la liste des espèces associées aux forêts anciennes inclut de nombreuses espèces rares et menacées. Puisqu'un facteur essentiel pour interpréter ces résultats est – une fois de plus– l'histoire, et que l'étude de la colonisation de ces espèces semble être l'aspect clé de la question, il apparaît donc qu'à cette échelle aussi l'utilisation des marqueurs moléculaires pourrait se révéler particulièrement intéressante. Non pas pour identifier les zones de plus forte diversité régionale ou locale, résultats d'un intérêt très limité, mais pour vérifier les hypothèses des écologistes (faibles aptitudes à la dispersion des graines, donc forte différenciation pour des marqueurs transmis maternellement chez ces espèces), reconstruire les voies de migration par graines entre les fragments plus anciens et les bois ou forêts plus récentes, et étudier les conséquences génétiques de ces colonisations. Toutes ces informations, combinées à l'étude directe de la dispersion des graines, pourront être très utiles pour gérer ces forêts. La possibilité de

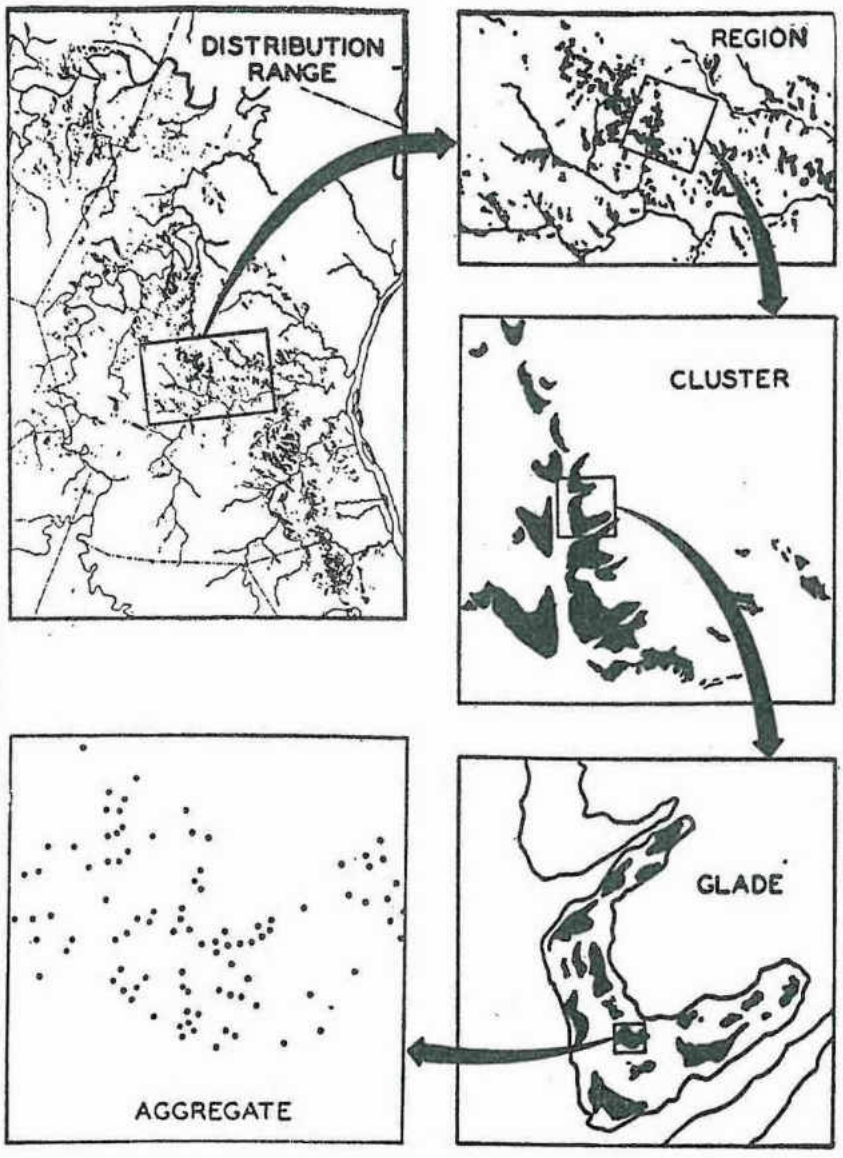
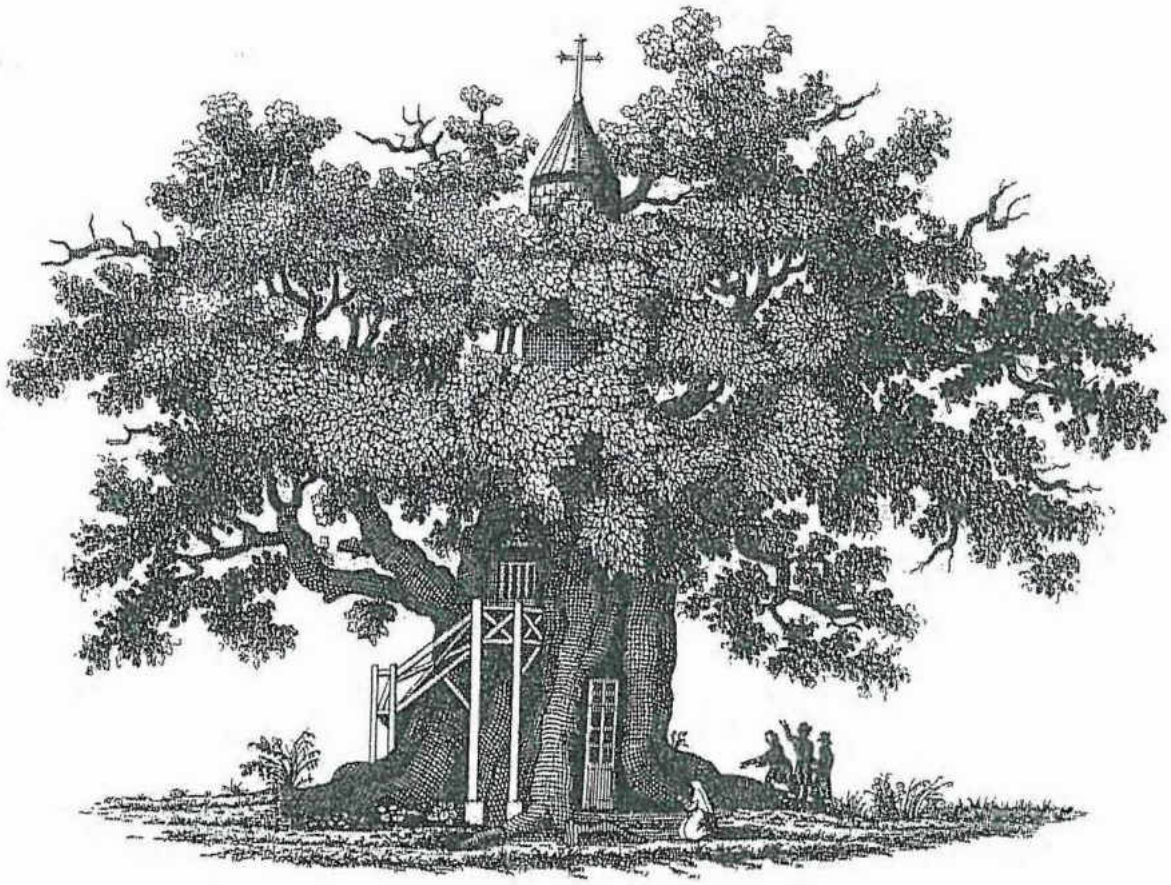


Figure 8 : Diagramme illustrant l'organisation de l'aire de distribution de *Clematis fremontii* var. *riehlii* à différentes échelles hiérarchiques. D'après Erickson (1945) in Stebbins (1950).

modéliser les processus étudiés en se basant sur des données réelles (distribution spatio-temporelle des forêts à l'échelle locale ou régionale, prise en compte de la qualité de l'environnement) devrait permettre des progrès significatifs dans cette discipline récente qu'est l'écologie du paysage (Wiens 1997). Des cartes de distribution plus précises des espèces aux différentes échelles spatiales seront également nécessaires pour ces études (un exemple déjà ancien de ce qui est souhaitable est illustré figure 8), or pour la plupart des espèces végétales (et même pour les arbres) les aires de distribution françaises et régionales sont extrêment approximatives, même si des efforts sont en cours.



Le chêne-chapelle d'Allouville-Bellefosse, en Seine-Maritime, d'après une gravure de 1822. C'est le plus gros chêne pédonculé de France : 15 m de circonférence à la base, 10 m à 1 m du sol. Il a été « christianisé » en 1696 (Lieutaghi 1998).

Conclusion

Dans l'introduction, je constatais que, pour rendre compte de la structuration génétique des populations animales et végétales, les facteurs historiques étaient de plus en plus souvent évoqués. Ceci est particulièrement sensible chez les arbres forestiers, organismes particulièrement longévifs. Cette tendance est récente : bien que des explications historiques aient toujours reçu une certaine attention de la part des généticiens des populations, ils n'occupaient toutefois pas une place centrale. Les modèles à l'équilibre de Wright et Fisher, depuis les années 30, servaient bien plus fréquemment de base aux interprétations.

Les progrès de la paléobotanique et l'émergence de la phylogéographie chez les plantes (grâce à l'étude de l'ADN chloroplastique) a contribué à recentrer les études de variabilité génétique autour de l'histoire. Même pour des marqueurs qui, comme les isozymes, ne permettent pas d'approche phylogénétique, la distribution de la diversité ne peut se comprendre sans référence à l'histoire des populations.

Les arbres forestiers dont l'histoire est particulièrement bien connue grâce à leur abondance et à leur bonne représentation dans les sédiments du Quaternaire, peuvent être considérés comme des organismes modèles pour étudier l'effet de l'histoire sur la structuration génétique. Cela peut paraître paradoxal d'utiliser ces organismes à longue durée de vie, qui apparaissent tardivement dans les successions, comme modèles pour étudier l'effet de la colonisation ! On songerait *a priori* plutôt à des espèces annuelles récemment introduites par l'homme. Toutefois, l'échelle de temps considérée doit être en rapport avec la longueur des générations. Les arbres forestiers permettent en fait les mêmes analyses que des espèces à plus courts cycles, tout en remontant plus loin dans le temps.

Les arbres constituent-ils également de bons modèles pour la conservation des écosystèmes forestiers ? Peut-on déterminer la meilleure manière de conserver la diversité génétique des espèces forestières en interpolant aux autres organismes les résultats obtenus sur les arbres, organismes pour lesquels des études particulièrement détaillées existent ? Peut-on par exemple identifier des zones refuges méritant une attention particulière, aussi bien pour la conservation de la diversité génétique des espèces que pour celle de la biodiversité ? Le projet basé sur l'analyse comparative d'une trentaine d'espèces d'arbres et arbustes que j'anime doit permettre d'évaluer dans quelle mesure l'histoire de chaque espèce d'arbres fut unique, ou si au contraire des tendances générales existent. Toutefois, par la suite, l'étude d'espèces végétales plus menacées aux capacités de dispersion très faibles pourrait compléter utilement notre vision des problèmes se posant au gestionnaire des milieux naturels forestiers.

J'ai tenté de présenter au cours des pages précédentes plusieurs perspectives de recherches qui permettraient d'aller plus loin encore vers ce recentrage autour de l'histoire, qui me semble nécessaire aujourd'hui. En effet, l'homme a engendré



Le chêne-chapelle d'Allouville-Bellefosse, en Seine-Maritime, d'après une gravure récente : seul un fragment de la cime montrée sur la figure précédente subsiste (Clémence & Péron 1995). La conservation de la « biodiversité » doit inclure également des aspects éthiques, esthétiques et culturels…

des déséquilibres (fragmentation de la forêt, plantations, changements du climat) qui semblent avoir un effet majeur sur la structuration génétique des espèces forestières. Il convient, par une étude historique indépendante, de caractériser précisément ces perturbations avant toute analyse génétique. C'est à cette condition qu'on pourra évaluer leur effet sur l'évolution des espèces.

Pour continuer de progresser dans ce sens, de nouveaux outils d'analyse devront voir le jour ou se développer. L'acquisition des données moléculaires devra être plus automatisée que par le passé ; la qualité des données devra encore progresser, pour permettre l'application plus systématique des concepts de la phylogénie ; la paléogénétique doit pouvoir être appliquée avec profit à l'étude de nombreuses questions difficilement accessibles sans elle ; l'analyse des données, qui s'accumuleront à un rythme plus rapide, devra également mieux valoriser l'information recueillie, dépasser le stade de la génétique des paires, raisonner sur des généalogies entières, afin notamment de valoriser les progrès effectués dans le domaine de la théorie de la coalescence.

L'étude des conséquences de l'histoire sur la structuration des caractères adaptatifs est très peu développée. Est-ce parce que ces relations sont faibles, qu'il n'y a jamais « maladaptation », que quelque soit le pool génétique de départ les plantes s'adapteront très vite et de la même façon à leur milieu ? Linhart et Grant (1996) concluent leur revue sur la signification évolutive de la differentiation génétique locale des plantes en proclamant que « *natural selection is the principal force shaping genetic architecture in natural plant populations* ». Dans ce travail, pratiquement aucune référence directe n'est faite à l'histoire des populations. Le développement de travaux permettant d'évaluer cette relation entre histoire et adaptation semble désormais essentiel.

De nouvelles approches de plus en plus multidisciplinaires devront désormais voir le jour. Les collaborations déjà engagées vont dans ce sens, mais devront se poursuivre. Au delà de l'application des outils moléculaires dont il est provisoirement l'un des principaux utilisateurs, le généticien des populations devra apporter, dans ces échanges interdisciplinaires, ses connaissances sur les mécanismes d'évolution des populations. A partir de cette base, il pourra tenter de prédire comment cette évolution pourra être affectée par les interventions humaines. Cela devrait permettre, en concertation avec ses collègues des autres disciplines et les gestionnaires des forêts, de proposer des solutions d'aménagement et de gestion durables.

Références

Allard RW Miller RD & Kahler AL (1978).- The relationship between degree of environmental heterogeneity and genetic polymorphism. *In* : Freysen and Woldendorp (eds.), *Structure and functioning of plant populations*. North Holland Publishing Company, pp 49-73.

Anderson E (1936).- The species problem in *Iris*. *Ann. Missouri Bot. Gard.*, 23 : 457-509.

Archie JW (1985).- Statistical analysis of heterozygosity data: independent sample comparisons. *Evolution*, 39 : 623-637.

Arnold J Asmussen MA & Avise JC (1988).- An epistatic mating system model can produce permanent cytonuclear disequilibria in a hybrid zone. *Proc. Natl. Acad. Sci. USA*, 85 : 1893-1896.

Arnold ML Robinson JJ Buckner CM & Bennet BD (1992).- Pollen dispersal and intraspecific gene flow in Louisiana irises. *Heredity*, 68 : 399-404.

Asmussen MA Arnold J & Avise J (1987).- Definition and properties of disequilibrium statistics for associations between nuclear and cytoplasmic genotypes. *Genetics*, 115 : 755-768.

Atchison BA Whitfeld PR & Bottomley W (1976).- Comparison of chloroplast DNAs by specific fragmentation with EcoR1 endonuclease. *Mol. Gen. Genet.*, 148 : 263-269.

Avise JC Giblin-Davidson C Laerm J Patton JC et coll (1979).- Mitochondrial DNA clones and matriarchal phylogeny within and among populations of the pocket gopher, *Geomys pinetis*. *Proc. Natl. Acad. Sci. USA*, 76 : 6694-6698.

Avise JC (1994).- *Molecular Markers, Natural History and Evolution*. Chapman & Hall, New York.

Avise JC Arnold J Ball RM Bermingham E Lamb T Neigel JE Reeb CA & Saunders NC (1987).- Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Ann. Rev. Ecol. Syst.*, 18 : 489-522.

Banks JA & Birky, Jr CW (1985).- Chloroplast DNA diversity is low in a wild plant, *Lupinus texensis*. *Proc. Natl. Acad. Sci. USA*, 82 : 6950-6954.

Baradat P & Marpeau-Bézard A (1988).- Le pin maritime *Pinus pinaster* Ait.: biologie et génétique des terpènes pour la connaissance et l'amélioration de l'espèce. Thèse, Univ. Bordeaux I.

Barbault R (1992).- *Ecologie des peuplements. Structure, dynamique et évolution*. Masson, Paris, p. 273.

Bartels H (1991).- Genetic control of multiple esterases from needles and macrogametophytes of *Picea abies*. *Planta (Berl.)* 99 : 283-289.

Barton NH & Whitlock MC (1997).- The evolutionof metapopulations. *In* : I. A. Hanski & M. E. Gilpin (eds), *Metapopulation Biology. Ecology, Genetics, and Evolution*. Academic Press, San Diego, California, pp 183-210.

Bataillon TM David JL & Schoen DJ (1996).- Neutral genetic markers and conservation genetics: simulated germplasm collections. *Genetics*, 144 : 409-417.

Baur E (1909) Das Wesen und die Erblichkeitsverhältnisse der "*Varietates albomarginatae* hort." von *Pelargonium zonale*. *Zeit. Ind. Abst. Vereb.* **1**: 330-351.

Bedbrook JR & Bogorad L (1976). Endonuclease recognition sites mapped on *Zea mays* chloroplast DNA. *Proc. Natl. Acad. Sci. USA*, 7 : 4309-4313.

Bennett KD (1985).- The spread of *Fagus grandifolia* across eastern north America during the last 18 000 years. *J. Biogeo.*, 12 : 147-164.

Bennett KD (1988).- Holocene geographic spread and population expansion of *Fagus grandifola* in Ontario, Canada. *J. Ecol.*, 76 : 547-557.

Bennett KD (1997).- *Evolution and ecology. The pace of life*. Cambridge University Press, Cambridge, U.K. 241 p.

Berger A L (1976).- Obliquity and precession for the last 5 000 000 years. *Astronomy and Astrophysics* 51 : 127-135.

Bergmann F (1973).- Genetische Untersuchungen bei *Picea abies* mit Hilfe der Isoenzym-Identifizierung III. Geographische Variation an 2 Esterase- und 2 Leucin-aminopeptidase-Loci in der schwedischen Fichtenpopulation. *Silv. Genet.*, 22 : 63-66.

Berkaloff A, Bourguet J, Favard P, Favard N & Lacroix J-C (1981).- *Biologie et physiologie cellulaires III. Chloroplastes, peroxysomes, division cellulaire*. Hermann, Paris.

Birks HJB (1989).- Holocene isochrone maps and patterns of tree-spreading in the British Isles. *J. Biogeo.*, 16 : 503-540.

Birky, Jr CW (1978).- Transmission genetics of mitochondria and chloroplasts. *Ann. Rev. Genet.*, 12 : 471-512.

Birky, Jr CW (1989).- Organelle evolution. *Genome*, 31 : 1095-1097.

Birky, Jr CW Maruyama T & Fuerst P (1983).- An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics*, 103 : 513-527.

Birky, Jr CW Fuerst P & Maruyama T (1989).- Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics*, 121 : 613-627.

Birky, Jr CW (1988).- Evolution and variation in plant chloroplast and mitochondrial genomes. *In* : Leslie D. Gottlieb and Subodh K. Jain (eds.), *Plant Evolutionary Biology*. Chapman and Hall, London and New York, pp 23-53.

Brandon RN (1996). *Concepts and methods in evolutionary biology*. Cambridge University Press, Cambridge. 221 p.

Brewer G J & Sing C F (1970).- An introduction to isozyme techniques. Academic Press, New York.

Briggs D & Walters S M (1984).- *Plant variation and evolution*. Second edition, Cambridge University Press, Cambridge.

Brown AHD (1978).- Isozymes, plant population genetic structure and genetic conservation. *Theor. Appl. Genet.*, 52 : 145-157.

Brown AHD & Schoen DJ (1994).- Optimal sampling strategies for core collections of plant genetic resources. *In* : V. Loeschcke, J. Tomiuk & S. K. Jain (eds), *Conservation genetics*. Birkäuser Verlag, Basel, Switzerland, pp 357-370.

Brown AHD & Weir BS (1983).- Measuring genetic variability in plant populations. *In* : S.D. Tanksley and T.J. Orton (eds.), *Isozymes in plant genetics and breeding, Part A*. Amsterdam, pp 219-239.

Camefort H (1966).- Observations sur les mitochondries et les plastes d'origine pollinique après leur entrée dans une oosphère chez le Pin noir (*Pinus laricio* Poir. var. *austriaca* = *Pinus nigra* Arn.). *C. R. Acad. Sci., Paris*, 263 : 1371-1374.

Camefort H (1968).- Sur l'organisation du néocytoplasme dans les proembryons tétranucléés du *Larix decidua* Mill. (*Larix europea* D. C.) et l'origine des mitochondries et des plastes de l'embryon chez cette espèce. *C. R. Acad. Sci., Paris*, 266 : 88-91.

Caspari E, Watson G S & Smith W (1966).- The influence of cytoplasmic pollen sterility on gene exchange between populations. *Genetics*, 53 : 741-746.

Carson HL (1990).- Increased genetic variance after a population bottleneck. *Rev. Géo. Pyr. S-O.*, 5 : 228-230.

Chakraborty R Smouse PE & Neel JV (1988).- Population amalgation and genetic variation: observations on artificially agglomerated tribal populations of Central and South America. *Am. J. Hum. Genet.*, 43 : 709-725.

Chakraborty R & Danker-Hopfe H (1991).- Analysis of population structure: a comparative study of different estimators of Wright's fixation indices. *In* : C. R. Rao and R. Chakraborty (eds.), *Handbook of Statistics*, Vol. 8. Elsevier Science Publishers B. V., pp 203-254.

Chakraborty R & Fuerst PA (1979).- Some sampling properties of selectively neutral alleles. *Genet. Res. Camb.*, 34 : 253-267.

Chan P-H & Wildman SG (1972).- Chloroplast DNA codes for the primary structure of the large subunit of Fraction I protein. *Biochim. Biophys. Acta*, 277 : 677-680.

Chesnoy L (1969).- Sur la participation du gamète mâle à la constitution du cytoplasme de l'embryon chez le *Biota orientalis* Endl. *Rev. Cytol. Biol. vég.*, 32 : 273-294.

Chesnoy L (1987).- L'origine des organites du cytoplasme embryonnaire chez les Gymnospermes. *Bull. Soc. Bot. Fr.*, 134 : 51-56.

Chesnoy L & Thomas MJ (1971).- Electron microscopy studies on gametogenesis and fertilization in Gymnosperms. *Phytomorphology*, 21 : 50-63.

Cipriani G Testolin R & Morgante M (1995).- Paternal inheritance of plastids in interspecific hybrids of the genus *Actinidia* revealed by PCR-amplification of chloroplast DNA fragments. *Mol. Gen. Genet.*, 247 : 693-697.

Clark JS (1998).- Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. *Am. Nat.*, 152 : 204-224.

Clark JS Fastie C Hurtt G Jackson ST Johnson C King GA Lewis M Lynch J Pacala S Prentice C Schupp EW Webb III T & Wyckoff P (1998).- Reid's paradox of rapid plant migration. *BioScience*, 48 : 13-24.

Clegg MT (1987).- Preface, plant molecular evolution. *Amer. Natur.* 130 : S1-S5.

Clegg MT Brown AHD & Whitfeld PR (1984).- Chloroplast DNA diversity in wild and culivated barley: implications for genetic conservation. *Genet. Res. Camb.*, 43 : 339-343.

Clémence J-F & Péron F (1995).- *Le chêne*. Actes Sud, Arles.

Cockerham CC & Weir BS (1993).- Estimation of gene flow from *F*-statistics. *Evolution*, 47 : 855-863.

Conkle MT (1971).- Inheritance of alcohol dehydrogenase and leucine amino-peptidase isozymes in knobcone pine. *Forest Sci.* 17 : 190-194.

Cornuet JM & Luikart G (1996).- Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, 144 : 2001-2014.

Correns C (1900).- G. Mendels Regel über das Verhalten der Nachkommenschaft Rassenbastarde. *Ber. Deutsch. Bot. Ges.* 18:158-168.

Correns C (1908).- Die Rolle der männlichen keimzellen bei der Geschlechtsebestimmung der Gynodiöcischen Pflanzen. *Ber. Deutsch. Bot. Ges.* 26, 686-701.

Correns C (1909).- Verebungsversuche mit blass (gelb) grünen und buntblättrigen Sippen bei *Mirabilis*, *Urtica* und *Lunaria*. *Zeit. Ind. Abst. Vererb.* 1, 291-329.

Courtois B (1975).- Plantations en forêt de Compiègne. 1774-1974. Mémoire de 3ème année d'ENITEF.

Crow JF & Mange AP (1965).- Measurement of inbreeding from the frequency of marriages between persons of the same surname. *Eugen. Quart.* 12 : 199-203.

Cuguen J Merzeau D & Thiebaut B (1988).- Genetic structure of the European beech stands (*Fagus sylvatica* L.): F-statistics and importance of mating system characteristics in their evolution. *Heredity*, 60 : 91-100.

Cwynar LC & MacDonald GM (1987).- Geographical variation of lodgpole pine in relation to population history. *Am. Nat.*, 129 : 463-469.

Daday H (1954).- Gene frequencies in wild populations of *Trifolium repens*. 1. Distribution by latitude. *Heredity*, 8 : 61-78.

Davis MB Woods KD Webb SL & Futyma RP (1986).- Dispersal versus climate: Expansion of *Fagus* and *Tsuga* into the Upper Great Lakes region. *Vegetatio*, 67 : 93-103.

DeBenedictis (1973).- On the correlations between certain diversity indices. *Am. Nat.*, 107 : 295-302.

Dobzhansky Th (1955).- A review of some fundamental concepts and problems of population genetics. *Cold Spring Harbor Symp. Quant. Biol.* 20 : 213-224.

Dong J & Wagner DB (1994).- Paternally inherited chloroplast polymorphism in Pinus: Estimation of diversity and population subdivision, and tests of disequilibrium with a maternally inherited mitochondrial polymorphism. *Genetics*, 136 : 1187-1194.

East EM 1910 A Mendelian interpretation of variation that is apparently continuous. *Amer. Nat.* 44 : 65-82.

Ellstrand NC & Elam DR (1993).- Population genetic consequences of small population size: implications for plant conservation. *Ann. Rev. Ecol. Syst.*, 24 : 217-242.

Ennos RA (1994).- Estimating the relative rates of pollen and seed migration among plant populations. *Heredity*, 72 : 250-259.

Ewens WJ (1972).- The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.*, 3 : 87-112.

Excoffier L Smouse PE & Quattro JM (1992).- Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131 : 479-491.

Faith DP (1992).- Conservation evaluation and phylogenetic diversity. *Biol. Cons.*, 61 : 1-10.

Ferris C Oliver RP Davy AJ & Hewitt GM (1993).- Native oak chloroplasts reveal an ancient divide across Europe. *Molecular Ecology*, 2 : 337-344.

Fisher RA (1937).- The wave of advance of advantageous genes. *Ann. Eugen.* 7 : 355-369.

Frankel OH Brown AHD & Burdon JJ (1995).- *The conservation of plant biodiversity*. Cambridge University Press, Cambridge, pp. 299.

Frankel R Scowcroft WR & Whitfeld PR (1979).- Chloroplast DNA variation in isonuclear male-sterile lines of *Nicotiana*. *Mol. Gen. Genet.*, 169 : 129-135.

Gliemeroth AK (1997).- Paläoökologische Aspekte der Einwanderungsgeschichte einiger Baumgattungen während des Holozäns nach Europa. *Angewandte Botanik*, 71 : 54-61.

Golenberg EM (1994).- DNA from plant compression fossils. *In* : B. Hermann, S. Hummel (eds.), *Ancient DNA*. Springer-Verlag, New York, pp 237-256.

Gouyon P-H Henry J-P & Arnould J (1997). *Les avatars du gène*. Belin, Paris. 336 p.

Govindaraju DR Wagner DB Smith GP & Dancik BP (1988).- Chloroplast DNA variation within individual trees of a *Pinus banksiana-Pinus contorta* sympatric region. *Can. J. For. Res.*, 18 : 1347-1350.

Grun P (1976).- *Cytoplasmic genetics and evolution*. Columbia University Press, New York.

Hamrick JL Linhart YB & Mitton JB (1979).- Relationships between life history characteristics and electrophoretically detectable genetic variation in plants. *Ann. Rev. Ecol. Syst.*, 10 : 173-200.

Hamrick JL Godt MJW & Sherman-Broyles SL (1992).- Factors influencing levels of genetic diversity in woody plant species. *New Forests*, 6 : 95-124.

Hardy C Callou C Vigne J-D Casane D Dennebouy N Mounolou J-C & Monnerot M (1995).- Rabbit mitochondrial DNA diversity from prehistoric to modern times. *J. Mol. Evol.*, 40 : 227-237.

Harper JL (1977).- *Population biology of plants*. Academic Press, London/New York/Toronto/Sydney/San Francisco. 892 p.

Heck, JrKL Belle G & Simberloff D (1975).- Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. *Ecology*, 56 : 1459-1461.

Hedgpeth J Goodman HM, Boyer HW (1972).- DNA nucleotide sequence restricted by the R1 endonuclease. *Proc. Natl. Acad. Sci. USA* 69 : 3448-3452.

Hiratsuka J Shimada H Whittier R Ishibashi T Sakamoto M Mori M et coll (1989).- The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes account for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.*, 217 : 185-194.

Holsinger KE (1996).- The scope and the limits of conservation genetics. *Evolution*, 50 : 2558-2561.

Hudson RR (1983).- Properties of a neutral allele model with intragenic recombination. *Theor. Appl. Genet.*, 23 : 183-201.

Hudson R R (1990).- Gene genealogies and the coalescent process. *Oxford Surv. Evol. Biol.* 7 : 1-44.

Huntley B & Birks HJB (1983).- *An atlas of past and present pollen maps for Europe, 0-13000 years ago*. Cambridge Univ. Press, Cambridge, U.K.

Hurlbert SH (1971).- The nonconcept of species diversity: a critique and alternative parameters. *Ecology*, 52 : 577-586.

James JW (1971).- The founder effect and response to artificial selection. *Genet. Res. Camb.*, 16 : 241-250.

Johnson LB & Palmer JD (1989).- Heteroplasmy of chloroplast DNA in *Medicago*. *Plant Mol. Biol.*, 12 : 3-11.

Kingman JFC (1982).- The coalescent. *Stochastic Processes and their Applications*, 13 : 235-248.

Kirk JTO (1971).- Chloroplast structure and biogenesis. *Ann. Rev. Biochem.* 40 : 161-196.

Kirk JTO & Tilney-Bassett RAE (1978).- *The plastids: their chemistry, structure, growth, and inheritance*. 2d ed. Elsevier, Amsterdam.

Koerner W Dupouey JL Dambrine E & Benoît M (1997).- Influence of past land use on the vegetation and soils of present day forest in the Vosges mountains, France. *J. Ecol.*, 85 : 351-358.

Kolodner R & Tewari KK (1972).- Molecular size and conformation of chloroplast deoxyribonucleic acid from pea leaves. *J. biol.Chem.* 10 : 6355-6364.

König A Groppe K & Ziegenhagen B (1998).- First results of chloroplast-DNA investigations in German populations of *Quercus petraea* and *Q. robur*. *In*: Diversity and adaptation in oak species, KC Steiner, ed. Proceedings of the second meeting of Working Party 2.08.05, Genetics of *Quercus*, of the International Union of Forest Research Organizations. October 12-17 1997. University Park (State College), Pennsylvania, USA.

Konnert M & Bergmann F (1995).- The geographical distribution of genetic variation of silver fir (*Abies alba, Pinaceae*) in relation to its migration history. *Pl. Syst. Evol.* 196 : 19-30.

Kremer A (1994).- Diversité génétique et variabilité des caractères phénotypiques chez les arbres forestiers. *Genet. Sel. Evol.*, 26, Suppl : 105s-123s.

Kremer A & Zanetto A (1997).- Geographical structure of gene diversity in *Quercus petraea* (Matt.) Liebl. II: Multilocus patterns of variation. *Heredity*, 78 : 476-489.

Kremer A & Zanetto A (1997).- Geographical structure of gene diversity in *Quercus petraea* (Matt.) Liebl. II: Multilocus patterns of variation. *Heredity*, 78 : 476-489.

Lamattina L Weil J-H & Grienenberger J-M (1989).- RNA editing at a splicing site of NADH dehydrogenase subunit IV gene transcript in wheat mitochondria. *FEBS*, 258 : 79-83.

Landsteiner K (1900).- Zur Kentnnis der antifermentativen lytischen und agglutinierenden Wirkung des Blutserum und der Lymph. *Zentr. Bakteriol. Parasitenk.* 27, 357-362.

Landy A Ruedisueli E, Robinson L, Foeller C & Ross W (1974).- Digestion of deoxyribonucleic acids from bacteriophage T7, λ and ϕ 80 h with site-specific nucleases from *Haemophilus influenzae* Strain Rc and Strain Rd. *Biochemistry* 13 : 2134-2142.

Larrère C & Larrère R (1997).- *Du bon usage de la nature. Pour une philosophie de l'environnement*. Aubier, Paris. 355 p.

Le Corre V & Kremer A (1998).- Cumulative effects of founding events during colonisation on genetic diversity and differentiation in an island and stepping-stone model. *J. Evol. Biol.*, 11 : 495-512.

Legay J-M & Vernay M (1999).- L'instructive histoire des noms de familles. *Pour la science*, 255 : 58-65.

Levings, III CS & Pring DR (1976).- Restriction endonuclease analysis of mitochondrial DNA from normal and Texas cytoplasmic male-sterile maize. *Science* 193 : 158-160.

Lewontin RC (1972).- The apportionment of Human diversity. *Evol. Biol.*, 6 : 381-398.
Lewontin RC (1991).- Twenty-five years ago in GENETICS: Electrophoresis in the development of evolutionary genetics: milestone or millstone?. *Genetics*, 128 : 657-662.
Lieutaghi P (1998).- *La plante compagne. Pratique et imaginaire de la flore sauvage en Europe occidentale*. Actes Sud, Arles. 299 p.
Linhart YB & Grant MC (1996).- Evolutionary significance of local genetic differentiation in plants. *Ann. Rev. Ecol. Syst.* 27 : 237-277.
Liston A (1992).- Variation in the chloroplast genes rpoC1 and rpoC2 of the genus Astragalus (Fabaceae): evidence from restriction site mapping of a PCR-amplified fragment. *Amer. J. Bot.*, 79 : 953-961.
Loreille O Mounolou J-C & Monnerot M (1997).- Histoire des lapins et ADN ancien. *C. R. Soc. Biol.* 191 : 537-544.
Luikart G Sherwin WB Steele BM & Allendorf FW (1998).- Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change. *Mol. Ecol.*, 7 : 963-974.
Lynch M & Crease TJ (1990).- The analysis of population survey data on DNA sequence variation. *Mol. Biol. Evol.*, 7 : 377-394.
MacDonald GM (1993).- Fossil pollen analysis and the reconstruction of plant invasions. *Advances in Ecological Research*, 24 : 67-110.
MacDonald GM & Cwynar LC (1991).- Postglacial population history of *Pinus contorta* ssp. *latifolia* in the western interior of Canada. *J. Ecol.* 79, 417-429.
Marshall DR & Jain SK (1969).- Genetic polymorphism in natural populations of *Avena fatua* and *A. barbata*. *Nature* 221 : 276-278.
Marshall DR & Brown AHD (1975).- Optimum sampling strategies in genetic conservation. *In* : Frankel & Hawkes (eds), *Crop genetic resources for today and tomorrow*, pp 53-80.
McCauley DE (1995).- The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Trends Ecol. Evol.*, 10 : 198-202.
Milligan BG Leebens-Mack J & Strand AE (1994).- Conservation genetics: beyond the maintenance of marker diversity. *Molecular Ecology*, 3 : 423-435.
Mollison D (1977) Spatial contact models for ecological and epidemic spread. *J. Roy. Stat. Soc. Ser. B*, 39 : 283-326.
Moody ME & Mack RN (1988).- Controlling the spread of plant invasions: the importance of nascent foci. *J. Appl. Ecol.* 25 : 1009-1021.
Muller HJ (1950) Our load of mutations. *Amer. J. Hum. Genetics* 2, 111-176.
Nagylaki T (1998).- Fixation indices in subdivided populations. *Genetics*, 148 : 1325-1332.
Neale DB Marshall KA & Sederoff RS (1989).- Chloroplast and mitochondrial DNA are paternally inherited in *Sequoia sempervirens* D.Don Endl. *Proc. Natl. Acad. Sci. USA*, 86 : 9347-9349.
Neale DB Saghai-Maroof MA Allard RW Zhang Q & Jorgensen RA (1986).- Chloroplast DNA diversity in populations of wild and cultivated barley. *Genetics*, 120 : 1105-1110.
Neale DB Marshall KA & Sederoff RR (1988).- Inheritance and evolution of conifer organelle genomes. *In* : J.W. Hanover and D.E. Keathley (eds.), *Genetic manipulation of woody plants*. Plenum Press, New York, pp 251-264.
Neel JV Satoh C Smouse P Asakawa J-I Takahashi N Goriki K Fujita M Kageoka T & Hazama (1988).- Protein variants in Hiroshima and Nagasaki: Tales of two cities. *Am. J. Hum. Genet.*, 43 : 870-893.
Nei M (1973).- Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA*, 70 : 3321-3323.
Nei M (1978).- Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89 : 583-590.
Nei M (1982).- Evolution of human races at the gene level. In: B. Bonne-Tamir, T. Cohen & RM Goodman (eds.). *Human genetics*, part A: *The unfolding genome*. Alan R. Liss, New York, pp 167-181.
Nei M (1986).- Definition and estimation of fixation indices. *Evolution*, 40 : 643-645.
Nei M & Chesser RK (1983).- Estimation of fixation indices and gene diversities. *Ann. Hum. Genet.*, 47 : 253-259.
Nei M & Maruyama T (1975).- Lewontin-Krakauer test for neutral genes. *Genetics*, 80 : 395.
Ohba K Iwakawa M Okada Y & Murai M (1971).- Paternal transmission of a plastid anomaly in some reciprocal crosses of sugi, *Cryptomeria japonica* D. Don. *Silv. Genet.*, 20 : 101-107.
Ohyama K Fukuzawa H Kohchi T Shirai H Sano T Sano S Umesono K (1986).- Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature*, 322 : 572-574.
Olivieri I & Gouyon P-H (1997).- Evolution of migration rate and other traits. The metapopulation effect. *In* : I. A. Hanski & M. E. Gilpin (eds), *Metapopulation Biology. Ecology, Genetics, and Evolution*. Academic Press, San Diego, California, pp 293-323.
Palmer JD (1985).- Comparative organization of chloroplast genomes. *Ann. Rev. Genet.* 19 : 325-354.

Palmer JD (1987).- Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. *Am. Nat.*, 130 : S6-S29.

Palmer JD & Zamir D (1982).- Chloroplast DNA evolution and phylogenetic relationships in *Lycopersicon*. *Proc. Natl. Acad. Sci. USA*, 79 : 5006-5010.

Patil GP & Taillie C (1982).- Diversity as a concept and its measurement. *J. Am. Stat. Assoc.*, 77 : 548-561.

Peet RK (1974).- The measurement of species diversity. *Ann. Rev. Ecol. Syst.*, 5 : 285-307.

Peterken GF (1974).- A method of assessing woodland flora for conservation using indicator species. *Biol.Conserv.* 6 : 239-245.

Picard JF (1998).- La génétique cytoplasmique (et le singulier destin de la génétique en France). *In*: Le petit pois déridé. XXème réunion annuelle du groupe de biologie et génétique des populations. Université de Lille I. 4 p.

Pitelka LF & the plant migration workshop group (1997).- Plant migration and climate change. *American Scientist*, 85 : 464-473.

Potter SS Newbold JE Hutchison III CA & Edgell MH (1975).- Specific cleavage analysis of mammalian mitochondrial DNA. *Proc. Natl. Acad. Sci. USA*, 72 : 4496-4500.

Quetier F & Vedel F (1977).- Heterogeneous populations of mitochondrial DNA molecules in higher plants. Nature 268 : 365-368.

Rackham O (1975).- *Hayley wood. Its history and ecology*. Cambridge: Cambridgeshire and Isle of Ely Naturalists Trust.

Rackham O (1980).- *Ancient woodland: its history, vegetation and uses in England*. Arnold, Londres.

Reid C (1895).- The dispersal of acorns by rooks. *Nature*, 53 : 6.

Reid C (1899).- *The origin of the British flora*. Dulau, London.

Rhoades MM (1933) The cytoplasmic inheritance of male sterility in *Zea mays*. *Journ. Gen.* 27, 71-93.

Rieseberg LH Hanson MA & Philbrick CT (1992).- Androdioecy is derived from dioecy in Datiscacea: evidence from restriction site mapping of PCR-amplified chloroplast DNA fragments. *Syst. Bot.*, 17 : 324-336.

Rieseberg LH & Soltis DE (1991).- Phylogenetic consequences of cytoplasmic gene flow in plants. *Evol. Trend. Plants*, 5 : 65-84.

Rollo F Venanzi FM & Amici A (1994).- DNA and RNA from ancient plant seeds. *In* : B. Hermann & S. Hummel (eds), *Ancient DNA*. Springer-Verlag, New York, pp 218-236.

Rousset F (1997).- Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, 145 : 1219-1228.

Sakai K-I Miyazaki Y & Matsuura T (1971).- Genetic studies in natural populations of forest trees. I. Genetic variability on the enzymatic level in natural forests of *Thujopsis dolabrata*. *Silv. Genet.*, 20 : 168-173.

Sandbrink JM Van Bruggen AC & Van Brederode J (1990).- Patterns of infraspecific chloroplast DNA variation in species of *Silene* sect. *Elisanthe*. *Biochem. Syst. Ecol.*, 18 : 233-238.

Sanders HL (1968).- Marine benthic diversity: a comparative study. *Am. Nat.*, 102 : 243-282.

Sapp J 1987 *Beyond the gene. Cytoplasmic inheritance and the struggle for authority in genetics*. Oxford University Press, Oxford, UK.

Sauer JD (1988).- *Plant migration: the dynamics of geographic patterning in seed plant species*. University of California Press, London, England.

Schnabel A & Asmussen MA (1989).- Definition and properties of disequilibria within nuclear-mitochondrial-chloroplast and other nuclear-dicytoplasmic systems. *Genetics*, 123 : 199-215.

Schoen DJ & Brown AHD (1991).- Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proc. Natl. Acad. Sci. USA*, 88 : 4494-4497.

Schuster L (1950).- Über den Sammeltrieb des Eichelhahers (*Garrulus glandarius*). *Vogelwelt*, 71 : 9-17.

Selander RK & Whittam TS (1983).- Protein polymorphism and the genetic structure of populations. In: M Nei & RK Koehn (eds.), *Evolution of genes and proteins*, Sinauer, Sunderland (MA), USA, pp. 89-114.

Shigesada N Kawasaki K & Takeda Y (1995).- Modeling stratified diffusion in biological invasions. *Am. Nat.*, 146 : 229-251.

Shinozaki K Ohme M, Tanaka M et coll. (1986).- The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *The EMBO Journal* 5 : 2043-2049.

Simberloff DS (1972).- Properties of the rarefaction diversity measurement. *Am. Nat.*, 106 : 414-418.

Simpson EH (1949).- Measurement of diversity. *Nature* 163 : 688.

Skellam JG (1951).- Random dispersal in theoretical populations. *Biometrika*, 38 : 196-218.

Slatkin M (1977).- Gene flow and genetic drift in a species subject to frequent local extinctions. *Theor. Popul. Biol.*, 12 : 253-262.

Soltis DE Soltis PS & Milligan BG (1992).- Intraspecific chloroplast DNA variation: Systematic and phylogenetic implications. *In* : P. S. Soltis, D. E. Soltis and J. J. Doyle (eds.), *Molecular Systematics of Plants*. Chapman and Hall, New York, London, pp 117-150.

Soltis DE Mayer MS Soltis PS & Edgerton M (1991).- Chloroplast-DNA variation in *Tellima grandiflora* (Saxifragaceae). *Amer. J. Bot.*, 78 : 1379-1390.

Southern EM (1975).- Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.*, 98 : 503-517.

Stebbins G L, Jr. (1950).- *Variation and evolution in plants*. Columbia University Press, New York.

Sugiura M (1992).- The chloroplast genome. *Plant Mol. Biol.* 19 : 149-168.

Suyama Y Kawamuro K Kinoshita I Yoshimura K Tsumura Y & Takahara H (1996).- DNA sequence from a fossil pollen of *Abies* spp. from Pleistocene peat. *Genes Genet. Syst.*, 71 : 145-149.

Szmidt AE Aldén T & Hällgren J-E (1987).- Paternal inheritance of chloroplast DNA in *Larix*. *Plant Mol. Biol.*, 9 : 59-64.

Taberlet P Fumagalli L Wust-Saucy A-G & J-F Cosson (1998).- Comparative phylogeography and postglacial colonization routes in Europe. *Mol. Ecol.* 7 : 453-464.

Taberlet P Gielly L & Bouvet J (1991).- Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.*, 17 : 1105-1109.

Tajima F (1983).- Evolutionary relationship of DNA sequences in finite populations. *Genetics*, 105 : 437-460.

Takahata N & Palumbi SR (1985).- Extranuclear differentiation and gene flow in the finite island model. *Genetics*, 109 : 441-457.

Thompson R Hughes SG & Broda P (1974).- Plasmid identification using specific endonuclease. *Mol. Gen. Genet.* 133 : 144-149.

Thorpe RS (1984).- Primary and secondary transition zones in speciation and population differentiation: a phylogenetic analysis of range expansion. *Evolution*, 38 : 233-243.

Timothy DH Levings III CS Pring DR Conde MF & Kermicle JL (1979).- Organelle DNA variation and systematic relationships in the genus *Zea*: Teosinte. *Proc. Natl. Acad. Sci. USA*, 76 : 4220-4224.

Vedel F Quétier F & Bayen M (1976).- Specific cleavage of chloroplast DNA from higher plants by *Eco*RI restriction nuclease. *Nature*, 263 : 440-442.

Vedel F Quétier F Dosba F & Doussinault G (1978).- Study of wheat phylogeny by EcoRI analysis of chloroplastic and mitochondrial DNAs. *Plant Science Letters*, 13 : 97-102.

Vendramin GG Lelli L Rossi P & Morgante M (1996).- A set of primers for the amplification of 20 chloroplast microsatellites in *Pinaceae*. *Mol. Ecol.*, 5 : 595-598.

de Vienne D Leonardi A & Damerval C (1988).- Genetic aspects of variation of protein amounts in maize and pea. *Electrophoresis*, 9 : 742-750.

Villablanca FX (1994).- Spatial and temporal aspects of populations revealed by mitochondrial DNA. *In* : B. Herrmann & S. Hummel (eds), *Ancient DNA*. Springer-Verlag, New York, pp 31-58.

Wagner DB Furnier GR Saghai-Maroof MA Williams SM Dancik BP et coll (1987).- Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. *Proc. Natl. Acad. Sci. USA*, 84 : 2097-2100.

Wagner DB Sun Z-X Govindaraju DG & Dancik BP (1991).- Spatial patterns of chloroplast DNA and cone morphology variation within populations of a *Pinus banksiana-Pinus contorta* sympatric region. *Am. Nat.*, 138 : 156-170.

Wakasugi T Tsudzuki J Ito S Nakashima K Tsudzuki T & Sugiura M (1994).- Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proc. Natl. Acad. Sci. USA* 91: 9794-9798.

Wakeley J (1998).- Segregating sites in Wright's island model. *Theor. Popul. Biol.*, 53 : 166-174.

Ward RD Skibinski DOF & Woodwark M (1992).- Protein heterozygosity, protein structure, and taxonomic differentiation. *Evol. Biol.*, 26 : 73-159.

Watson G S & Caspari E (1960).- The behavior of cytoplasmic pollen sterility in populations. *Evolution*, 15 : 56-63.

Watterson GA (1984).- Allele frequencies after a bottleneck. *Theor. Popul. Biol.*, 26 : 387-407.

Webb SL (1987).- Beech range extension and vegetation history: pollen stratigraphy of two Wisconsin lakes. *Ecology*, 68 : 1993-2005.

Weir BS & Cockerham CC (1984).- Estimating *F*-statistics for the analysis of population structure. *Evolution*, 38 : 1358-1370.

Weitzman ML (1992).- On diversity. *Quart. J. Econom.*, 107 : 363-405.

Whittemore AT & Schaal BA (1991).- Interspecific gene flow in oaks. *Proc. Natl. Acad. Sci. USA*, 88 : 2540-2544.

Wiens JA (1997).- Metapopulation dynamics and landscape ecology. *In* : I. A. Hanski & M. E. Gilpin (eds), *Metapopulation Biology. Ecology, Genetics, and Evolution*. Academic Press, San Diego, California, pp 43-62.

Wildman SG (1971).- *In*: NK Boardman, AW Linnane & RW Smillie (eds.), *Autonomy and biogenesis of mitochondria and chloroplasts*, North Holland, Amsterdam, p. 402.

Wolfe KH Li W-H & Sharp PM (1987).- Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. USA*, 84 : 9054-9058.

Woods KD & Davis MB (1989).- Paleoecology of range limits: beech in the upper peninsula of Michigan. *Ecology*, 70 : 681-696.

Wright S (1931).- Evolution in Mendelian populations. *Genetics*, 16 : 97-159.

Wu J Krutovskii KV & Strauss SH (1998).- Abundant mitochondrial genome diversity, population differentiation and convergent evolution in pines. *Genetics*, 150 : 1605-1614.

Wulf M (1997).- Plant species as indicators of ancient woodland in northwestern Germany. *J. Veget. Sc.*, 8 : 635-642.

Wulf M (1998).- Distribution of ancient woodland, afforestation and clearances in relation to Quaternary deposits and soil types in northwestern Brandenburg (Germany). *In* : K.J. Kirby and C. Watkins (eds.), *The ecological history of European forests*. Cab international, UK, pp 301-310.

Zurawski G Clegg MT & Brown AHD (1984).- The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. *Genetics* 106 : 735-749.

Liste des publications

I-PUBLICATIONS SCIENTIFIQUES

I.1 - Articles (ou communications) primaires

I.1.1. Dans périodique à comité de lecture.

(1) **Bacilieri R, Ducousso A, Petit RJ, Kremer A** (1996) Mating system and asymmetric hybridization in a mixed stand of European oaks. *Evolution*, **50**, 900-908.

(2) **Bahrman N, Petit RJ** (1995) Genetic polymorphism in maritime pine (*Pinus pinaster* Ait.) assessed by two-dimensional gel electrophoresis of needle, bud, and pollen proteins. *Journal of Molecular Evolution*, **41**, 231-237.

(3) **Chat J, Chalak L, Petit RJ** Strict paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in intraspecific crosses of kiwifruit. *Theoretical and Applied Genetics*, en révision.

(4) **Demesure B, Comps B, Petit RJ** (1996) Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution*, **50**, 2515-2520.

(5) **Demesure B, Sodzi N, Petit RJ** (1995) A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology*, **4**, 129-131.

(6) **Dumolin S, Demesure B, Petit RJ** (1995) Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theoretical and Applied Genetics*, **91**, 1253-1256.

(7) **Dumolin-Lapègue S, Demesure B, Le Corre V, Fineschi S, Petit RJ** (1997) Phylogeographic structure of white oaks throughout the European continent. *Genetics*, **146**, 1475-1487.

(8) **Dumolin-Lapègue S, Bodénès C, Petit RJ** (1997) Detection of rare polymorphisms in mitochondrial DNA of oaks with PCR-RFLP combined to SSCP analysis. *Forest Genetics*, **3**, 227-230.

(9) **Dumolin-Lapègue S, Pemonge M-H, Petit RJ** (1997) An enlarged set of consensus primers for the study of organelle DNA in plants. *Molecular Ecology*, **6**, 393-397.

(10) **Dumolin-Lapègue S, Pemonge M-H, Petit R J** (1998) Association between chloroplast and mitochondrial lineages in oaks. *Molecular Biology and Evolution*, **15**, 1321-1331.

(11) **Dumolin-Lapègue S, Kremer A, Petit R J** Are chloroplast and mitochondrial DNA variation species-independent in oaks? *Evolution*, sous presse.

(12) **El Mousadik A, Petit RJ** (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics*, **92**, 832-839.

(13) **El Mousadik A, Petit RJ** (1996) Chloroplast DNA phylogeography of the argan tree of Morocco. *Molecular Ecology*, **5**, 547-555.

(14) **Le Corre V, Machon N, Petit RJ, Kremer A** (1997) Colonisation with long-distance seed dispersal and genetic structure of maternally inherited genes in forest trees: a simulation study. *Genetical Research, Cambridge*, **69**, 117-125.

(15) **Petit RJ, Bahrman N, Baradat PH** (1995) Comparison of genetic differentiation in maritime pine (*Pinus pinaster* Ait.) estimated using isozyme, total protein and terpenic loci. *Heredity*, **75**, 382-389.

(16) **Petit RJ, El Mousadik A, Pons O** (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12**, 844-855.

(17) **Petit R J, Kremer A, Wagner D B** (1993) Geographic structure of chloroplast DNA polymorphisms in European oaks. *Theoretical and Applied Genetics*, **87**, 122-128.

(18) **Petit R J, Kremer A, Wagner D B** (1993) Finite island model for organelle and nuclear genes in plants. *Heredity*, **71**, 630-641.

(19) **Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducousso A, Kremer A** (1997) Chloroplast DNA footprints of postglacial recolonisation by oaks. *Proceedings of the National Academy of Sciences, USA*, **94**, 9996-10001.

(20) **Petit RJ, Pons O** (1998) Bootstrap variance of diversity and differentiation estimates in a subdivided population. *Heredity*, **80**, 56-61.

(21) **Petit R J, Wagner D B, Kremer A** (1993) Ribosomal RNA and chloroplast DNA polymorphisms in a mixed stand of *Quercus robur* and *Q. petraea*. *Annales des Sciences Forestières*, **50**, Suppl. 1, s41-s47.

(22) **Pons O, Petit RJ** (1995) Estimation, variance and optimal sampling of gene diversity. I. Haploid locus. *Theoretical and Applied Genetics*, **90**, 462-470.

(23) **Pons O, Petit RJ** (1996) Measuring and testing genetic differentiation with ordered *versus* unordered alleles. *Genetics*, **144**, 1237-1245.

(24) **Vendramin GG, Degen B, Petit RJ, Anzidei M, Madaghiele A, Ziegenhagen B** (1999) High level of variation at *Abies alba* chloroplast microsatellite loci in Europe. *Molecular Ecology*, sous presse.

Articles soumis ou en préparation:

(25) **Burban C, Petit RJ, Carcreff E, Jactel H.** Rangewide variation of the maritime pine bast scale *Matsucoccus feytaudi* Duc. (Homoptera: Matsucoccidae) in relation to the genetic structure of its host. *Mol. Ecol.* Soumis à *Mol. Ecol.*

(26) **Comps B, Gömöry D, Letouzey J, Thiébaut B, Petit RJ**. Imprint of postglacial colonization on the distribution of heterozygosity and allelic richness in the European beech. Soumis à *Genetics*.

(27) **Demesure B, Le Guerroué B, Lucchi G, Prat D, Petit RJ**. Genetic variability of a scattered temperate forest tree: *Sorbus torminalis* L. (Crantz). Soumis à *Ann. For. Sci.*

(28) **Dumolin-Lapègue S, Gielly L, Taberlet P, Petit RJ**. A mplification of DNA from ancient and modern oak wood. Soumis à *Mol. Ecol.*

(29) **Muloko-Ntoutoume N, Petit RJ, White L, Abernathy K.** Chloroplast DNA variation of a rainforest tree (*Aucoumea klaineana*, Burseraceae) in Gabon. Soumis à *Mol. Ecol.*

(30) **Petit RJ, Fineschi S, Vendramin GG**. Genetic structure of chloroplast and mitochondrial DNA polymorphisms in plants. En préparation.

(31) **Belahbib N, Pemonge M-H, Ouassou A, Sbay H, Kremer A, Petit RJ**. Chloroplast and mitochondrial DNA exchanges between cork oak (*Q. suber* L.) and holm oak (*Q. ilex* L.) in Morocco. En préparation.

I.1.3. Rapports diplômants

(32) **Petit RJ** (1988) Ressources génétiques du pin maritime : apport des marqueurs enzymatiques. DEA, Paris XI-Orsay, 106 p.
(33) **Petit RJ** (1992) Polymorphisme de l'ADN chloroplastique dans un complexe d'espèces: les chênes blancs européens. Thèse, Paris XI-Orsay, 87 p.

I.1.4. Communications courtes dans congrès, symposiums

(34) **Burban C, Petit RJ, Jactel H** Rangewide variation of mitochondrial DNA among populations of the maritime pine scale *Matsucoccus feytaudi* Duc. (Homoptera: Margarodidae). **Poster**, 20e Int. congr. of Entomology, Florence, Italie, 25-31 Août 1996
(35) **Dumolin-Lapègue S, Demesure B, Petit RJ** Comparaison de deux phylogénies maternelles (ADNcp et ADNmt) chez le chêne. **Communication orale**, 17ème Colloque de Biologie et Génétique des Populations, Lyon, 29-31 Août 1995.
(36) **Dumolin-Lapègue S, Demesure B, Petit RJ** Structuration spatiale de la variabilité chloroplastique chez les chênes blancs en Europe. **Poster**, 18ème Colloque de Biologie et Génétique des Populations, Grenoble, 27-30 Août 1996.
(37) **Petit RJ, Pons O** Analyse statistique de la différenciation génétique. **Communication orale**, 17ème Colloque de Biologie et Génétique des Populations, Lyon, 29-31 Août 1995.
(38) **Petit RJ** (1995) Méthodes rapides de caractérisation de la diversité cytoplasmique des plantes. 2de rencontre du groupe biologie moléculaire des arbres. 12-13 décembre
(39) **Petit RJ** (1995) Problèmes liés à la mesure de la diversité et de la différenciation avec des marqueurs hypervariables. Réunion Microsatellites et génétique des populations, 18-19 décembre 1995, Montpellier.

I.2. Synthèses scientifiques

I.2.1. Dans périodique à comité de lecture.

(40) **Bahrman N, Plomion C, Petit RJ, Kremer** (1997) Contribution of two-dimensional electrophoresis of proteins to maritime pine genetics. *Annales des Sciences Forestières*, **54**, 225-236.
(41) **Kremer A, Petit R J** (1993) Gene diversity in natural populations of oak species. *Annales des Sciences Forestières*, **50**, Suppl. 1, s186-s202.

I.2.2. Dans périodique sans comité de lecture.

(42) **El Alami S, Sbay H, Petit RJ, Ouassou A** (1998) Contribution à l'étude de la variabilité enzymatique de six provenances marocaines de pin maritime (*Pinus pinaster* Ait.). *Pp* 56-62 *in: Actes du séminaire sur l'Amélioration, la Conservation et l'Utilisation des*

Ressources Génétiques Forestières Marocaines, M Arbez, O M'Hirit, H Sbay, SL El Alami (eds).

(43) El Mousadik A, Fakir S, Petit RJ (1998) Apport des marqueurs moléculaires pour l'étude de la diversité génétique chez l'arganier *Argania spinosa* L. *Pp* 33-49 *in: Actes du séminaire sur l'Amélioration, la Conservation et l'Utilisation des Ressources Génétiques Forestières Marocaines*, M Arbez, O M'Hirit, H Sbay, SL El Alami (eds).

(44) Kremer A, Petit RJ, Ducousso A, Le Corre V (1998) General trends of variation of genetic diversity in *Quercus petraea* (Matt.) Liebl. In: Diversity and adaptation in oak species, KC Steiner, ed. Proceedings of the second meeting of Working Party 2.08.05, Genetics of *Quercus*, of the International Union of Forest Research Organizations. October 12-17 1997. University Park (State College), Pennsylvania, USA.

(45) Kremer A, Petit RJ, Ducousso A (1998) Structure of gene diversity, gene flow and gene conservation in *Quercus petraea*. In: First EUFORGEN meeting on social broadleaves, Bordeaux, France, 23-25 October 1997. J Turok, A Kremer, S de Vries, compilers. International Plant Genetic Resources Institute, Rome.

(46) Ouassou A, Petit RJ, Alfioun N, Sbay H, Kremer A (1998) Evaluation du polymorphisme chloroplastique chez le chêne liège. *Pp* 56-62 *in: Actes du séminaire sur l'Amélioration, la Conservation et l'Utilisation des Ressources Génétiques Forestières Marocaines*, M Arbez, O M'Hirit, H Sbay, SL El Alami (eds).

(47) Petit RJ, Demesure B, Pineau E, Kremer A (1995) Genetische Differenzierung in lokalem und kontinentalem Ausmaß bei europäischen Eichenarten: die Bedeutung geschichtlicher Faktoren. *Mitteilungen aus der Forstlichen Versuchsanstalt Rheinland-Pfalz*, **34**, 22-37.

(48) Petit RJ, Demesure B, Pineau E (1996) Chloroplast DNA differentiation at a local and continental scale in European oak species: the importance of historical factors. *In*: *Inter- and intraspecific variation in European oaks: Evolutionary implications and practical consequences*, A Kremer, H-J Muhs, eds. Office for official publications of the European Communities, Luxembourg, 145-163.

(49) Sork VL, Campbell D, Dyer R, Fernandez J, Nason J, Petit RJ, Smouse P, Steinberg E (1998) Proceedings from a workshop on gene flow in fragmented, managed, and continuous populations. Research Paper n° 3. National Center for Ecological Analysis and Synthesis, Santa Barbara, California. http://www.nceas.ucsb.edu/papers/geneflow/.

I.2.3. Chapitre d'ouvrage

(50) Kremer A, Petit R J, Zanetto A, Fougère V, Ducousso A, Wagner D B, Chauvin C (1991) Nuclear and organelle gene diversity in *Quercus robur* and *Q . petraea*. In: Genetic variation in European populations of forest trees. Ed: G. Müller-Starck, M. Ziehe. Sauerländer's Verlag, Frankfurt am Main, 141-166.

(51) Kremer A, Petit R J, Pons O (1998) Measures of polymorphism within and among populations. In: *Molecular tools for screening biodiversity: plants and animals*. Ed: A. Karp, P G Isaac, D S Ingram, Chapman & Hall, Londres, 301-311.

(52) Petit RJ, Demesure B, Pineau E (1996) Chloroplast DNA differentiation at a local and continental scale in European oak species: the importance of historical factors. *In*: *Inter- and intraspecific variation in European oaks: Evolutionary implications and practical*

consequences, A. Kremer, H-J Muhs, eds. Office for official publications of the European Communities, Luxembourg.

(53) **Petit R J, Dumolin-Lapègue S, Demesure B** (1998) cpDNA and plant mtDNA primers. In: *Molecular tools for screening biodiversity: plants and animals*. Ed: A. Karp, P G Isaac, D S Ingram, Chapman & Hall, Londres, 256-261.

(54) **Petit R J, Kremer A, Bacilieri R, Ducousso A, Zanetto A** (1992) Structuration génétique chez les chênes blancs européens. In: *Complexes d'espèces, flux de gènes et ressources génétiques des plantes*. Bureau des Ressources Génétiques Eds, Lavoisier, Cachan, 155-163.

I.2.6. Conférences dans congrès ou symposium

Communications invitées :

(55) **Kremer A, Petit RJ, Ducousso A, Le Corre V** (1997) General trends of variation of genetic diversity in *Quercus petraea* (Matt.) Liebl. Meeting IUFRO '*Diversity and adaptation of oak species*', 12-17 Oct 1997 (orateur en remplacement d'A. Kremer).

(56) **Petit RJ** (1995) Le point sur les études de diversité génétique chez les chênes. Rencontre annuelle sur les chênes de tonnellerie. 11-13 Septembre 1995, INRA Nancy.

(57) **Petit RJ** (1995) Evolutionary patterns in oaks inferred from chloroplast DNA polymorphisms. Séminaire 'Aspetti genetici e botanici del genre *Quercus*', Porano, Italie, 22 Juin 1995.

(58) **Petit RJ** (1996) Phylogeography, recolonization, and migration routes inferred from cytoplasmic markers. Symposium IUFRO'*Diversity and adaptation in forest ecosystems in a changing world*', Vancouver, Canada, 5-9 Août 1996.

(59) **Petit RJ** (1998) Contribution of cytoplasmic markers to studies of gene flow in plants. Workshop on gene flow in fragmented, managed, and continuous populations. National Center for Ecological Analysis and Synthesis, Santa Barbara, California, 5-9 Janvier 98.

(60) **Petit RJ** (1998) Measuring genetic differentiation with molecular markers to identify populations for conservation. Workshop on gene flow in fragmented, managed, and continuous populations. National Center for Ecological Analysis and Synthesis, Santa Barbara, California, 5-9 Janvier 1998.

I.2.7. Autres supports

(61) **Petit RJ** *Encadré sur l'ADNcp dans l'article « génomes végétaux » rédigé par H. Thiellement dans l'encyclopédie Hachette Axis 97.*

(62) **Petit RJ** A review of the comparative dynamics of two interfertile oak species: *Quercus robur* L. and *Quercus petraea* (Matt) Liebl. Manuscript non publié.

II- DOCUMENTS A VOCATION DE TRANSFERT

II.1. - Travaux personnels

II.1.2. Dans périodiques sans comité de lecture.

(63) Ducousso A, Bacilieri R, Demesure B, Petit R J, Zanetto A, Kremer A (1997) Structuration géographique de la diversité génétique chez les chênes à feuilles caduques européens. *Bulletin technique de l'ONF*, **33**, 7-19.

(64) Ducousso A, Bodénès C, Petit R J, Kremer A (1996) Le point sur les chênes européens. *Forêt-entreprise*, **112**, 49-56.

(65) Ducousso A, Petit R J (1994) Le geai des chênes Premier reboiseur européen. *Forêt-entreprise*, **97**, 60-64.

(66) Ducousso A, Petit RJ (1996) La ghiandaia: primo rimboschitore europeo. *Sherwood*, **12**, 6-10. (Traduction italienne de l'article précédent).

(67) Kremer A, Durel C-E, Petit R J, Villar M (1994) Marqueurs moléculaires et génétique des arbres forestiers. *Biofutur*, **2**, 17-23.

II.1.7. Rapports écrits

(68) Burban C, Petit R, Carcreff E, Jactel H (1998) Phylogéographie d'un ravageur specifique, *Matsucoccus feytaudi* (Homoptera: Matsucoccidae) en relation avec la structuration des populations de son hôte, *Pinus pinaster*. Rapport d'AIP, 15 pages.

(69) Ducousso A, Petit RJ, Kremer A (1996) Création d'un réseau de conservation des ressources génétiques des chênes blancs européens: premières propositions pratiques. Commission Nationale de Conservation des Ressources Génétiques Forestières. 15 p.

(70) Petit RJ, Chaussod R (1995) Conservation et Amélioration des Ressources Génétiques Forestières au Maroc: l'Arganier. Rapport de mission F.A.O., 1ère consultation. 31 p.

II.1.8. Créations informatiques

(71) Petit RJ (1995) HAPLODIV. Programme d'analyse statistique de la diversité génétique en TurboPascal et Pascal pour SUN.

(72) Petit RJ (1998) RAREF. Programme d'analyse de la richesse allélique en Turbo Pascal.

II.2 - Travaux encadrés

II.2.2. Mémoires de stages et thèses

(73) **Ben Brahim M** (1992) Etude de la structuration génétique et du régime de reproduction d'un peuplement naturel de sapin pectiné (Abies alba Mill.) DEA de Biologie Forestière, Nancy I. *(encadrement avec Ph. Baradat)*

(74) **Demesure B** (1996) Analyse de la diversité chloroplastique en utilisant des fragments PCR chez des Fagacées: *Fagus sylvaica* L. et *Quercus* ssp. Thèse de Doctorat, Bordeaux I.

(75) **Dumas S** (1994) Polymorphisme de l'ADN chloroplastique chez l'angélique (*Dicorynia guianensis*) en Guyane française. DEA de Biologie Forestière, Nancy I. *(avec A. Kremer)*

(76) **Dumolin-Lapègue S** (1998) Variabilité chloroplastique et mitochondriale chez les chênes blancs européens à deux échelles spatiales. Thèse de Doctorat, Paris XI Orsay.

(77) **El Alami SD** (1995) Initiation à l'électrophorèse d'enzymes sur gel d'amidon. CNRF, Rabat. Mémoire de 3ème année, école d'ingénieurs forestiers de Salé.

(78) **El Mousadik A** (1997) Organisation de la diversité génétique de l'arganier *Argania spinosa* L. Apport des marqueurs nucléaires et cytoplasmiques. Thèse de Doctorat d'Etat ès Sciences en Biologie, Option Ressources génétiques, Université Ibnou Zohr, Agadir, Maroc.

(79) **Jung-Müller B** (1993) Contribution à la mesure des tailles de populations chez les chênes: compatibilité pollinique et structuration spatiale en peuplement naturel. DEA Ressources Génétiques, Amélioration des Plantes et Création variétale, INA PG. *(avec A. Ducousso)*

(80) **Latouche-Hallé C** (1998) Variabilité de l'ADN chloroplastique des chênes blancs en France. DEA de Biologie, Diversité et Adaptation des Plantes cultivées, Paris XI Orsay.

(81) **Marque G** (1996) Variabilité intra- et interpopulations de l'ADN chloroplastique chez l'angélique de Guyane (*Dicorynia guianensis*). DEA de Biologie forestière, Nancy I.