

Zeitschrift: Saussurea : journal de la Société botanique de Genève
Herausgeber: Société botanique de Genève
Band: 32 (2002)

Artikel: Des familles aux populations : l'enseignement des gènes
Autor: Naciri Graven, Yamama / Manen, Jean-François
DOI: <https://doi.org/10.5169/seals-1098857>

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

Download PDF: 17.04.2026

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Des familles aux populations : l'enseignement des gènes

Contexte général : l'horloge moléculaire.

Le matériel génétique, responsable du fonctionnement des organismes vivants, est transmis de génération en génération à la faveur de la reproduction. Ce matériel, dont la matrice est l'ADN¹, est constitué de messages qu'on appelle gènes. Leur transmission étant fidèle dans la majorité des cas, les gènes de la descendance sont le reflet de ceux des deux parents et contiennent toujours l'empreinte des ancêtres, mais les mécanismes de la sexualité font que les gènes d'un individu sont réassortis - on dit "recombinés" - entre les lignées de ses quatre grands-parents. Cependant, durant la formation des gamètes qui s'unissent au cours de la fécondation, des erreurs de réplication du matériel génétique se produisent naturellement et peuvent être transmises à la descendance. Le sort de ces erreurs (on parle de mutations) se classe en 3 catégories: une majorité n'a pas d'avenir, car les descendants qui les reçoivent ne sont pas viables. Une partie est transmise, car elle n'a pas ou peu d'influence sur la survie des descendants. Enfin, une infime partie de ces erreurs apporte, par hasard, un avantage aux descendants, qui ont de plus grandes chances de les transmettre à leur tour, car les individus qui les portent en tirent un avantage.

Ces trois catégories de mutations sont définies par le rôle que joue la portion d'ADN muté. On parlera de **mutation silencieuse** lorsque celle-ci ne change pas le contenu du message, ou de **mutations neutres** lorsque la mutation apparaît soit dans une séquence d'ADN ne comportant pas de messages, soit dans une séquence comprenant un message dont la modification n'a pas d'influence sur la vie de l'individu et sa capacité à se reproduire. Ces mutations, neutres vis-à-vis de la sélection naturelle, sont transmises à la descendance selon les lois de Mendel ou par reproduction asexuée. D'autres parties de la séquence de l'ADN correspondent à des messages vitaux, par exemple les gènes codant pour la photosynthèse chez les plantes. Des mutations à ces endroits détériorent le message et, dans l'exemple choisi, résultent en un fonctionnement défectueux de la photosynthèse. On parle alors de **mutations défavorables**, voire **létales** si elles sont mortelles. Elles ont pour conséquence l'affaiblissement de la capacité reproductive des individus qui les portent ou leur disparition pure et simple. La dernière catégorie de mutations est beaucoup plus rare, mais très importante pour l'évolution des organismes. Lorsqu'une mutation crée un nouveau message ou change un message existant, en produisant une amélioration de certaines fonctions dans des conditions environnementales normales ou nouvelles, cette mutation se transmettra et donnera un avantage sélectif² aux descendants qui la portent. On parle alors de **mutation favorable**.

Comme les mutations sont le résultat d'erreurs qui se font naturellement au hasard des réplifications, il en résulte que plus des individus sont proches parents (et donc séparés par moins d'évènements de réplication), plus le nombre de

Yamama Naciri Graven
& Jean-François Manen

¹ ADN: L'ADN constitue le matériau de base des chromosomes. Il contient l'information héréditaire transmise de génération en génération. Il s'agit d'une longue molécule construite à partir de 4 motifs chimiques, les bases (Adénine (A), Cytosine (C), Guanine (G) et Thymine (T)), qui correspondent à autant de lettres d'un alphabet qui serait réduit à 4 signes. La taille d'une molécule d'ADN dépend du nombre total de chacun de ces motifs. Elle varie de quelques millions de motifs (bactéries) à quelques milliards (plantes et animaux) selon les organismes. La transmission du matériel génétique nécessite une réplication de l'ADN en une copie conforme qui est transmise à la descendance. C'est au cours de cette réplication que des erreurs (mutations) surviennent naturellement qui créent autant d'allèles ou d'haplotypes différents.

² **Avantage sélectif**: Avantage conféré à un individu par une mutation, qui lui permet de se reproduire de façon plus efficace que la moyenne de la population, et de transmettre ainsi un nombre plus important de copies du gène muté que ses congénères.

mutations qui les distinguent est petit. Le temps s'exprime par le nombre de générations qui séparent deux individus de leur ancêtre commun. Deux sœurs ont ainsi une génération d'écart avec leurs parents qui représentent leurs ancêtres communs. Deux cousines ont deux générations d'écart avec leurs grand-parents, qui représentent leurs ancêtres communs. La théorie de l'horloge moléculaire est basée sur ce principe. En effet, plus deux individus sont proches (plus leur dernier ancêtre commun est récent), moins leurs ADN sont différents (peu de réplifications, donc moins de chances que les mutations aient modifié leurs ADN respectifs), et plus deux individus sont éloignés (plus leur ancêtre commun est ancien), plus on observera de différences entre leurs ADN (beaucoup de réplifications, donc plus de chances que les mutations aient modifié leurs ADN respectifs). La théorie de l'horloge moléculaire suppose également que les mutations s'accumulent à un rythme régulier dans le temps. L'expérimentation, par séquençage³ de l'ADN, montre que, lorsque certaines précautions sont prises, cette théorie est juste.

Ce qui est vrai pour les gènes des individus d'une même espèce peut être étendu à ceux d'espèces plus ou moins différentes, dont Lamarck (*Discours de l'an VIII*, 1800) et Charles Darwin dans *L'origine des espèces* (1859) ont proposé qu'elles descendaient d'espèces ancestrales communes. Plus deux espèces sont proches (plus leur ancêtre commun est récent), moins on observera de différences (mutations peu nombreuses) entre leurs ADN respectifs. Plus deux espèces sont éloignées (plus leur ancêtre commun est ancien), plus on observera de différences (mutations plus nombreuses) entre leurs ADN. On comprend donc que la comparaison des séquences d'ADN entre individus ou espèces représente un outil remarquable dans la recherche scientifique pour un grand nombre de domaines: systématique, génétique des populations, histoire et géographie des peuplements (phylogéographie), écologie, évolution, recherche judiciaire.

C'est le développement de nouvelles techniques d'étude de l'ADN (PCR⁴ et séquençage automatique) qui a permis l'essor des sciences se basant sur l'analyse de la variabilité de l'ADN, que ce soit au niveau interspécifique ou intraspécifique. Le développement des outils informatiques et statistiques, ainsi que du réseau Internet, a aussi contribué à cet essor, car il permet de gérer et de partager les immenses quantités de données générées par les techniques moléculaires.

Phylogénie moléculaire.

La classification des organismes est basée, depuis plusieurs siècles, sur des données morphologiques très élaborées comme, par exemple, les structures de la fleur, de la graine, du pollen et autres organes pour les plantes. Toutefois, il s'avère que les espèces, familles, ordres... partagent des caractères qui ne reflètent pas toujours une parenté proche, mais souvent une convergence évolutive⁵. En raison de l'incertitude liée à cet état de fait (parenté ou convergence) et des désaccords quant à l'importance à donner à tel ou tel caractère dans la classification, des écoles différentes sont nées, chacune définissant un choix de caractères morphologiques et anatomiques décisifs particuliers.

L'utilisation des données moléculaires permet d'appliquer une méthode plus impartiale de comparaison entre les taxons, la phylogénie⁶ moléculaire, et cela pour différentes raisons: (1) Le nombre de caractères générés et comparés est énorme. Une seule réaction PCR permet d'amplifier une séquence de 1500 à 2000 nucléotides, qui représentent autant de caractères pouvant être comparés entre espèces. (2) Comme les mutations que l'on analyse sont généralement neutres, les caractères étudiés sont indépendants du phénotype⁷ et

³ Séquençage de l'ADN:

Technique moléculaire, maintenant partiellement automatisée, permettant de lire la séquence d'un fragment d'ADN, c'est-à-dire l'ordre des bases ou des nucléotides.

⁴ PCR: Pour Polymerase Chain Reaction. Technique in vitro permettant de reproduire à l'identique et en grande quantité un fragment d'ADN donné.

⁵ Convergence évolutive: C'est l'apparition de caractères similaires (identité par état) chez deux ou plusieurs taxons qui n'ont pas de lien de parenté direct (pas d'identité par descendance).

⁶ Phylogénie: Reconstruction hypothétique de l'évolution des espèces et autres taxons, sur la base de leur degré de parenté, en utilisant des caractères morphologiques ou moléculaires.

⁷ Phénotype: Ensemble des caractères visibles résultant de l'expression de l'ensemble des gènes d'un individu (appelé également génotype) dans un milieu particulier.

représentent donc un jeu de données indépendant des critères morphologiques habituellement examinés. Il est également plus facile d'évaluer l'homologie⁸ entre caractères moléculaires qu'entre caractères morphologiques. (3) Les lois régissant les mutations ainsi que leur taux sont mieux connues, ou du moins plus faciles à modéliser, que celles régissant les changements morphologiques. Des modèles d'évolution peuvent donc être testés. (4) Les différences observées entre deux taxons peuvent facilement être quantifiées. C'est la notion de distance génétique qui permet d'évaluer les liens de parenté.

Depuis la dernière décennie, une part de plus en plus importante des travaux de systématique botanique utilise les techniques de comparaison d'ADN en complément des données morphologiques. Tous les niveaux de la systématique, du genre à l'embranchement, sont étudiés (<http://phylogeny.arizona.edu/tree/phylogeny.html>). A ce stade, il est important de choisir des séquences adéquates pour le genre d'étude projeté. L'idéal est d'utiliser une séquence (gène ou séquence non codante) qui montre un taux optimal de mutation permettant de bien distinguer les taxons. Ainsi, si le but est l'étude d'espèces ou de genres, il faut choisir une séquence caractérisée par un fort taux de mutation, de façon à pouvoir observer suffisamment de variations entre taxons (allèles⁹ ou haplotypes¹⁰ différents) et distinguer des taxons séparés par peu d'événements de réplication, car ayant divergé récemment les uns par rapport aux autres. En revanche lorsque le but du travail est d'étudier les grandes lignes d'un embranchement entier, on utilise des gènes ou des séquences non codantes dont le taux de mutations est faible afin de pouvoir comparer des séquences d'organismes phylogénétiquement très éloignés. On parle alors de séquences conservées. En effet, si le taux de mutation est trop élevé, on risque d'observer une saturation importante de mutations dues à l'accumulation de mutations réverses (par exemple A->G->A) qui effacent l'histoire de la séquence étudiée, ce qui rend leur comparaison difficile et toute déduction sur l'histoire des taxons aléatoire.

A l'heure actuelle, la plupart des études menées sur un groupe d'organismes donné ne se limitent pas à l'étude d'une seule séquence d'ADN, mais utilisent plusieurs séquences appartenant à des génomes¹¹ différents (noyau, chloroplaste, mitochondrie). Des phylogénies sont alors produites en utilisant différentes méthodes comme la parcimonie¹² (fig. 5) ou le maximum de vraisemblance¹³ (Lewis, 2001). A ce stade de l'analyse, il faut souligner que les phylogénies obtenues sont des phylogénies de gènes et non des phylogénies d'espèces ou de genres, etc., selon le cas (Nichols, 2001). Aussi n'est-il pas rare d'obtenir des résultats différents lorsque plusieurs gènes sont analysés séparément. Cela résulte du fait que, bien que faisant partie d'un même génome, chaque gène a sa propre histoire (fig. 1), qui dépend de son taux de mutation, de l'histoire des espèces et des populations qui le portent, de sa neutralité, de son hérité maternelle, paternelle ou biparentale, de la recombinaison, etc... En revanche, lorsque les phylogénies obtenues sur différents compartiments cellulaires (noyau, chloroplaste, mitochondrie) sont compatibles, et que le nombre de gènes étudiés est suffisamment important, on peut raisonnablement en déduire que la phylogénie obtenue est une bonne image de l'évolution et de la diversification des espèces étudiées.

Historiquement, la phylogénie moléculaire s'est d'abord intéressée à la classification des taxons d'ordre supérieurs (embranchements, familles, voir par exemple Angiosperm Phylogeny Group, 1998 ; Spichiger et al., 2000) et des incongruences de classification entre phylogénie moléculaire et taxonomie

⁸ **Homologie** : Parenté entre deux ou plusieurs caractères (séquences d'ADN, fonction, phénotype, etc...).

⁹ **Allèle** : Forme particulière que prend un gène, déterminant l'un des états possibles du caractère codé par ce gène. On parle en général d'allèles pour des gènes nucléaires présents en deux copies dans le génome nucléaire (cas d'organisme diploïdes). Ici le terme est utilisé indifféremment pour des gènes nucléaires ou, en lieu et place du terme haplotype, pour des gènes mitochondriaux ou chloroplastiques qui ne sont présents qu'en une seule copie dans la cellule (génome haploïde).

¹⁰ **Haplotype** : Terme utilisé pour décrire une forme particulière de gène (ou de séquence non codante) portée par un génome haploïde, mitochondrie ou chloroplaste.

¹¹ **Génome** : Les cellules végétales contiennent trois génomes différents. Le génome nucléaire regroupe les chromosomes présents dans le noyau. Pour un organisme diploïde, chaque chromosome est au nombre de 2, l'un ayant été transmis par la mère, l'autre par le père. Les génomes mitochondriaux et chloroplastiques ne sont constitués que d'une molécule d'ADN circulaire contenue respectivement dans la mitochondrie et le chloroplaste. En général, ces génomes sont transmis par la mère, bien qu'il existe de nombreuses exceptions (transmission paternelle ou biparentale).

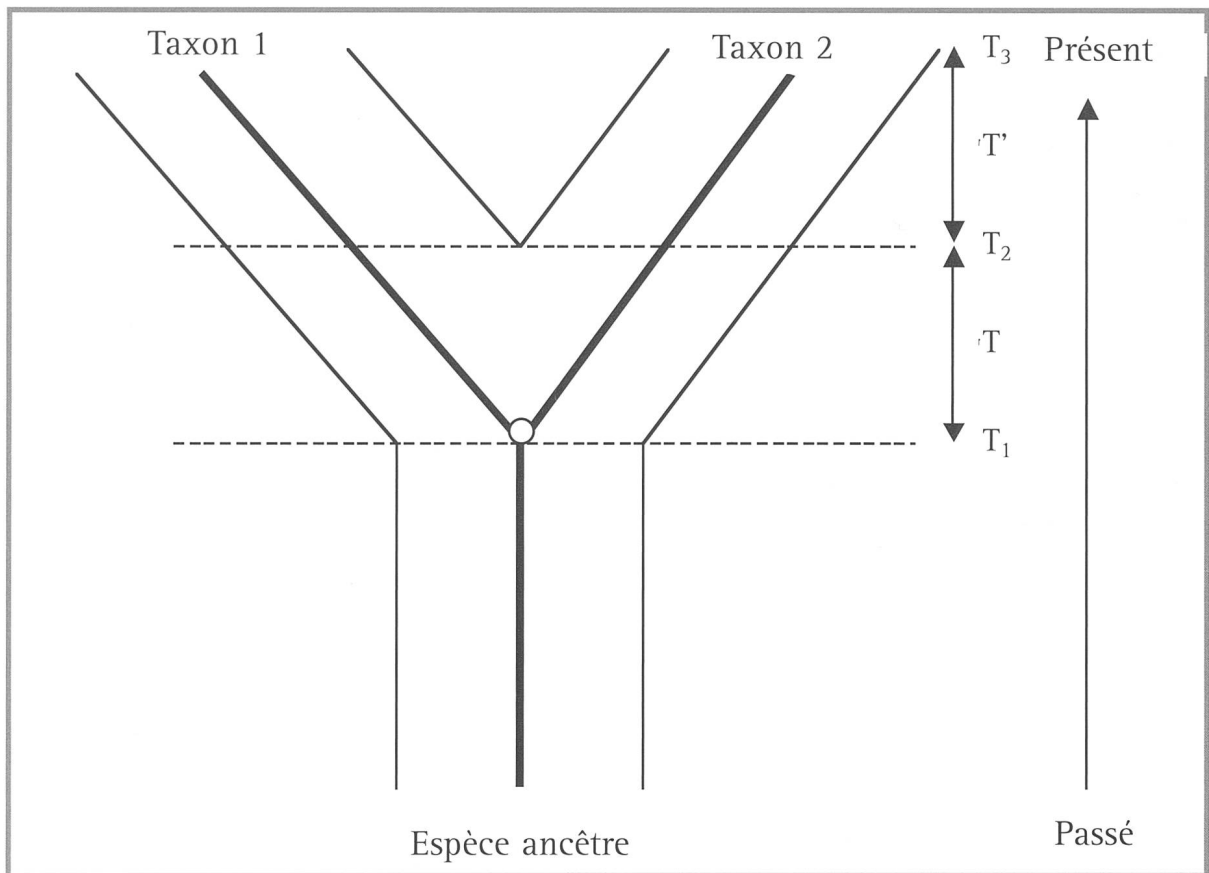


Figure 1. Différence entre histoire des allèles ou des haplotypes et histoire des populations ou des espèces (taxons). T_2 correspond à la date réelle de la fission des taxons (ou de la spéciation) et T_1 à la date à laquelle l'ancêtre commun le plus récent des deux haplotypes existait, date qui est fournie par la phylogénie (voir page 85). Enfin, T_3 correspond au présent. Le cercle situe l'événement de coalescence entre les deux lignées d'haplotypes représentées en trait gras, c'est-à-dire le moment où les deux lignées se fondent en une seule lignée quand on remonte dans le temps. Les traits fins représentent l'évolution des taxons. T correspond à la période de temps écoulée entre la coalescence des lignées et la fission des taxons et T' à la période écoulée entre la fission des taxons et le temps présent. Dans le cas de l'étude de genres ou d'espèces, T est du même ordre de grandeur que T' et ne peut donc être négligé. Par contre, dans le cas de l'étude de taxons de niveaux supérieurs, T est beaucoup plus petit que T' et on peut considérer que T_1 est assimilable à T_2 .

classique, basée sur des caractères morphologiques, ont ainsi été mises à jour (Jeanmonod, 2000). Lors de la comparaison de genres, de familles ou d'embranchements, quelques espèces sont analysées dans chaque taxon, à raison d'un individu par espèce, et ceci pour des gènes à faible taux de mutations (voir plus haut). Le choix d'un faible échantillonnage au sein de l'espèce ou des taxons supérieurs est légitimé par trois postulats : le premier est que la variabilité pour le gène étudié est nulle au sein de l'espèce, ou largement inférieure à ce qui peut s'observer entre espèces différentes (ce qui se conçoit si le gène mute lentement), et le second que les niveaux taxonomiques sont suffisamment distincts pour que tout flux de gènes¹⁴ entre ces derniers (y compris entre espèces) soit absent, et cela depuis longtemps. Ceci a pour conséquence que la mutation, couplée à un long isolement reproductif, aboutit à des groupes d'allèles ou d'haplotypes ne se recouvrant pas entre les taxons étudiés. Le troisième postulat admet la simultanéité des événements de spéciation et de différenciation des gènes. On constate cependant que l'ancêtre commun d'al-

¹² **Parcimonie**: Critère permettant, entre toutes les arbres phylogénétiques possibles, de ne retenir que les arbres les plus courts (c'est-à-dire avec un nombre minimum de mutations). Ce critère, qui est restrictif, répond à une notion d'économie dans l'évolution (fig. 5).

lèles ou d'haplotypes différents, présents dans deux espèces distinctes, est nécessairement antérieur à la spéciation (fig. 1), à moins que des échanges génétiques aient perduré après la spéciation, par exemple par le biais de l'hybridation. Le temps écoulé entre l'existence de cet ancêtre commun et la date réelle de la spéciation est en général considéré comme négligeable au regard du temps écoulé depuis la spéciation jusqu'à nos jours, ce qui se justifie lorsque sont analysés des taxons supérieurs de type ordres ou familles. Ces différents postulats permettent de considérer que les relations entre allèles ou haplotypes échantillonnés chez des individus appartenant à différentes espèces sont hiérarchiques, c'est à dire que chaque haplotype ne dérive que d'un seul autre haplotype (fig. 2a) et que l'approximation entre histoire des gènes et histoire des espèces est acceptable.

Depuis quelques années, il est par contre de plus en plus courant de s'intéresser à la phylogénie des espèces au sein d'un même genre ou au sein de genres voisins. Cet intérêt pour le niveau spécifique nécessite un changement de paradigme que peu d'études prennent en compte actuellement. En effet, plusieurs postulats ayant fait leurs preuves aux niveaux taxonomiques supérieurs ne sont plus valables au niveau spécifique ou intragénérique. Il s'agit entre autres du problème de l'échantillonnage, de la notion de relation hiérarchique entre taxons, de l'approximation entre histoire des gènes et histoire des populations et de l'existence d'allèles ancêtres.

1. *Echantillonnage*: Dans le cas d'analyses intragénérique ou intergénériques, des gènes à taux de mutations plus importants doivent être utilisés de façon à observer suffisamment de mutations permettant de différencier les espèces entre elles. Il n'est alors plus possible de négliger la variabilité intraspécifique, et l'échantillonnage au sein de chaque espèce doit être augmenté de façon à vérifier que toute la variabilité spécifique a bien été prise en compte. Malgré cela, le nombre de mutations différenciant les espèces proches (qui se sont différenciées récemment) est souvent réduit, ce qui ne permet pas de bénéficier d'une puissance statistique suffisante et d'obtenir des topologies¹⁵ robustes (Posada & Crandall, 2001). A un niveau plus général, il faut également s'assurer que toutes les espèces ont été échantillonnées, l'oubli de certaines espèces, en particulier les espèces rares, pouvant biaiser la topologie de l'arbre obtenu (Barraclough & Nee, 2001). De la même façon, le fait que des espèces aient disparu et ne sont donc pas prises en compte dans l'analyse peut biaiser la topologie de l'arbre.

2. *La relation phylogénétique hiérarchique entre taxons*. Elle n'est plus valable dans la mesure où des espèces différentes peuvent partager des allèles ou des haplotypes identiques du fait d'hybridations interspécifiques et de recombinaisons. La notion de relation hiérarchique, qui suppose que les groupes d'allèles ou d'haplotypes portés par chaque espèce ne se recoupent pas, doit alors être remplacée par celle de réseau (fig. 2b). On parle aussi de réticulation.

3. *Assimilation de l'histoire des gènes à celle des populations*. Alors qu'il est acceptable de considérer que le décalage entre histoire des gènes et histoire des espèces est négligeable dans le cas d'analyses d'ordres ou de familles, négliger le temps écoulé entre l'existence de l'ancêtre commun le plus récent des allèles étudiés (inféré par la phylogénie) et la fission des populations peut introduire un biais important dans la datation de l'évènement de spéciation. En assimilant l'histoire des gènes à celle de la spéciation, on a tendance à surestimer la date de la spéciation; ainsi dans le cas d'une paire d'espèces et de deux allèles, la date obtenue par la phylogénie est en moyenne supérieure de 50 % à la date réelle de la spéciation (Edwards & Beerli, 2000).

¹³ **Maximum de vraisemblance**: méthode permettant de construire des arbres phylogénétiques sur la base d'un algorithme prenant en compte un modèle d'évolution du fragment d'ADN étudié. Le modèle d'évolution est basé sur des fréquences et des probabilités, soit: (1) la fréquence de mutation, (2) le modèle de mutation (A a plus de chance de rester inchangé, que de changer en G, que de changer en C... etc), (3) la variance des fréquences de mutation selon les sites. L'arbre ayant le plus de vraisemblance (et qui sera retenu) correspond à celui dont l'ensemble des changements se rapproche le plus du modèle sous-jacent.

¹⁴ **Flux de gènes**: Transmission de gènes d'une population à l'autre par l'intermédiaire de la migration d'individus, de graines ou de pollen.

¹⁵ **Topologie**: Forme caractéristique de tout arbre phylogénétique. Pour un même échantillon de taxons, deux arbres auront des topologies différentes quand les relations hiérarchiques entre certains taxons sont différentes.

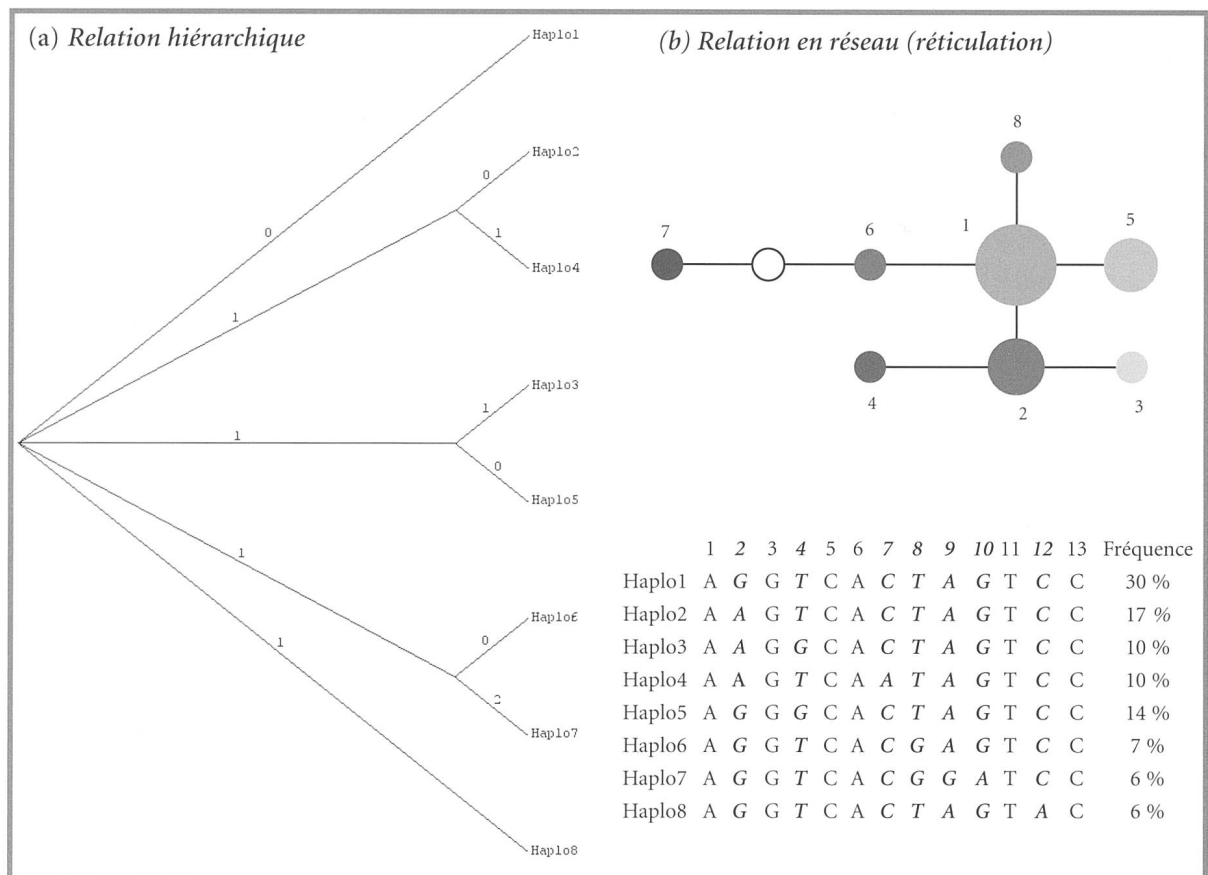


Figure 2. Arbre hiérarchique utilisant le critère de parcimonie (a) et réseau de gènes (b) construits l'un et l'autre sur la base des données ci-dessus. En (b), chaque cercle représente un allèle ou haplotype et la taille des cercles est proportionnelle à la fréquence des allèles ou haplotypes dans la population. Chaque allèle ou haplotype est connecté aux autres par une seule mutation symbolisée par un trait. L'ellipse vide correspond à un allèle ou un haplotype qui n'a pas été échantillonné. En (a), l'haplotype 1 apparaît bien à la racine de l'arbre (la taille des branches est indiquée au-dessus de chaque trait); en revanche, la possibilité que l'haplotype 3 dérive de l'haplotype 2 n'apparaît pas. Or d'après les fréquences, l'haplotype 3 a plus de chance de dériver de l'haplotype 2 que de l'haplotype 5.

4. *Existence d'allèles ancêtres.* Lorsqu'on s'intéresse à la phylogénie de taxons de niveaux supérieurs, il est tout à fait acceptable de considérer que les allèles ou haplotypes ancêtres ont disparu des espèces actuelles du fait de l'accumulation de mutations (c'est même l'une des hypothèses de base de la reconstruction des arbres). Dans le cas d'espèces ayant divergé récemment, il est par contre beaucoup plus hasardeux d'affirmer que les allèles ancêtres ont complètement disparu, ou qu'ils ont tous subi des mutations. De ce fait, l'échantillonnage d'allèles ancestraux n'est pas improbable dans les populations actuelles, ce qui peut introduire de gros biais dans la topologie des reconstructions phylogénétiques (Posada & Crandall, 2001).

Différentes approches permettent de répondre à ces problèmes dans le cadre de la phylogénie moléculaire (Doyle, 1992; Page & Charleston, 1997). Une autre approche, non exclusive, est de prendre en compte l'apport de la génétique des populations et tout particulièrement de la théorie de la coalescence (Kingman, 1982; Hudson, 1990) qui s'applique aux populations, mais également aux espèces proches et à l'estimation de leur date de spéciation. La

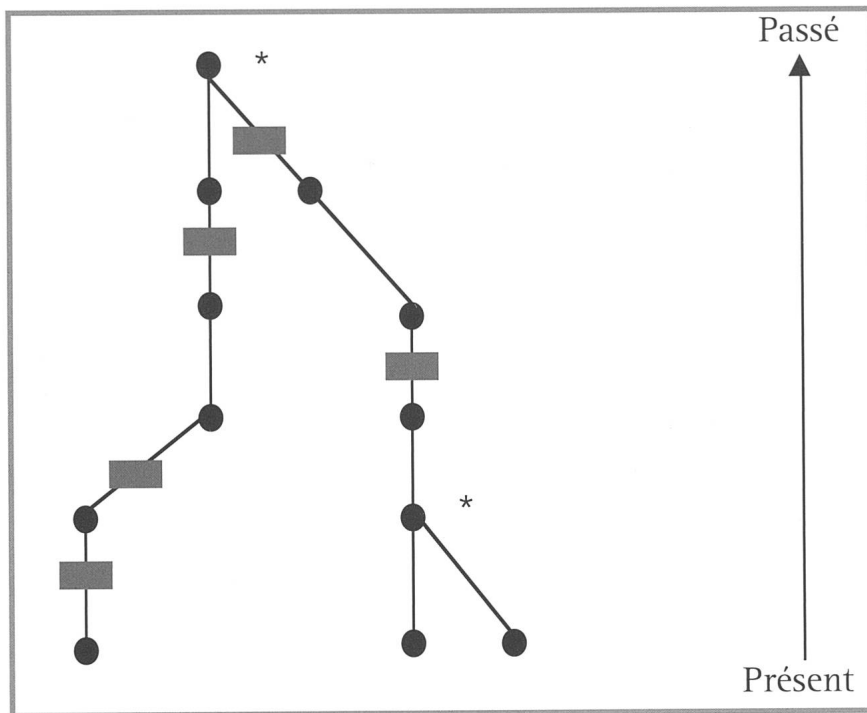


Figure 3. Généalogie d'allèles ou d'haplotypes. Chaque cercle correspond à un allèle ou haplotype et chaque rectangle à une mutation. L'événement de coalescence* correspond au moment où deux lignées d'allèles (ou d'haplotypes) fusionnent dans un même allèle (ou haplotype), c'est-à-dire que l'individu qui le portait en a transmis au moins deux copies. On voit que les événements de coalescence ne sont pas modifiés par la mutation.

théorie de la coalescence permet en effet de visualiser les problèmes posés, au niveau spécifique ou générique, par l'application de la théorie de la phylogénie moléculaire et des postulats qu'elle utilise, et de formaliser les modifications utiles à apporter.

Théorie de la coalescence et génétique des populations

La théorie de la coalescence (Kingman, 1982 ; Hudson, 1990) vise à reconstruire l'histoire généalogique des gènes (allèles ou haplotypes) jusqu'à l'ancêtre commun le plus récent (ACPR ou MRCA en anglais pour Most Recent Common Ancestor). L'approche généalogique en fait une science connexe de la phylogénie moléculaire, mais la prise en compte de l'histoire démographique des populations introduit une différence fondamentale. La théorie de la coalescence est une approche rétrospective. Seul importe l'échantillon des allèles analysés et la modélisation de l'ensemble de la population dont ils sont issus n'est pas nécessaire. En cela, elle se différencie de la génétique des populations, approche essentiellement prospective, qui vise à prédire l'évolution du polymorphisme génétique dans une population sous différentes pressions évolutives et qui fait appel à la théorie mathématique de l'échantillonnage, l'échantillon devant être représentatif de la population.

Le terme de coalescence correspond à l'événement selon lequel deux lignées d'allèles ou d'haplotypes fusionnent dans un même individu ancêtre (ACPR) lorsqu'on remonte dans le temps, cet individu ancestral ayant transmis au

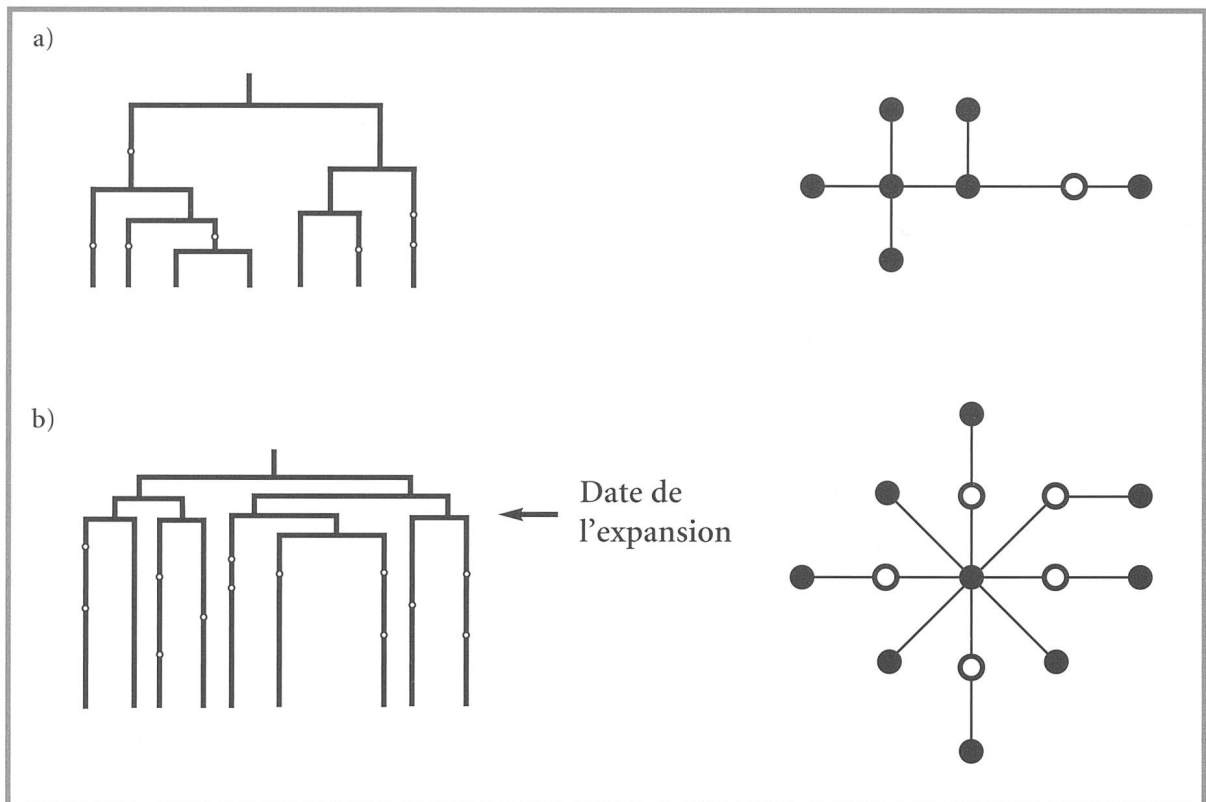


Figure 4. La dynamique de population a un impact très fort sur la structure de la généalogie. (a) : démographie stable : les événements de coalescence sont répartis sur l'ensemble de la généalogie ; (b) : population en expansion (généalogie en peigne) : la grande majorité des événements de coalescence se concentrent au moment de l'expansion démographique. Le type de régime démographique se traduit également par des types de réseaux différents, réseau réticulé pour les démographies stables et réseau en étoile pour les expansions. Les cercles vides représentent les haplotypes absents ou non échantillonnés.

moins 2 copies de son gène à ses descendants. Selon la théorie de la coalescence, la généalogie des allèles et leur coalescence dans le temps est un phénomène purement stochastique qui ne dépend pas du taux de mutation (fig. 3), mais de la taille de la population dont les allèles sont issus. Ainsi, la probabilité de coalescence de deux allèles à la génération précédente est inversement proportionnelle à la taille de la population. De la même façon, et pour une population donnée, la généalogie moyenne d'un gène mitochondrial (haploïde et transmis par la mère uniquement, donc en nombre $N/2$ si $N/2$ est le nombre d'individus de sexe féminin) est attendue 4 fois plus courte que la généalogie d'un gène nucléaire (diploïde et transmis par les deux parents, donc en nombre $2N$, N étant le nombre total d'individus dans la population). Cette différence traduit le fait que la taille efficace¹⁶ (N_e) pour un gène mitochondrial est 4 fois moindre que la taille efficace pour un gène nucléaire, toutes choses égales par ailleurs. Cela a pour conséquence immédiate que, selon le gène étudié et son hérédité, on pourra remonter plus ou moins loin dans le temps. Cela a également comme corollaire que la datation de l'ACPR peut varier fortement selon qu'on s'intéresse à tel ou tel gène, comme cela a été démontré pour l'homme (Excoffier, 2000). Enfin, la forme de la généalogie (et du réseau associé) dépend fortement du régime démographique de la population (fig. 4), ce qui permet en retour de tester différentes hypothèses démographiques.

¹⁶ Taille efficace : dans une population de taille N dans laquelle tous les individus ne participent pas également à la reproduction, taille fictive théorique (N_e) d'une population idéale dans laquelle les deux sexes seraient également représentés, où tous les individus auraient des descendants en nombre équivalent et où la diversité génétique serait équivalente à celle observée dans la population réelle. Le concept de taille efficace permet de comparer l'évolution de la diversité génétique dans des populations différant par leur taille effective, leur mode de reproduction, le nombre de reproducteurs, le sex-ratio, etc...

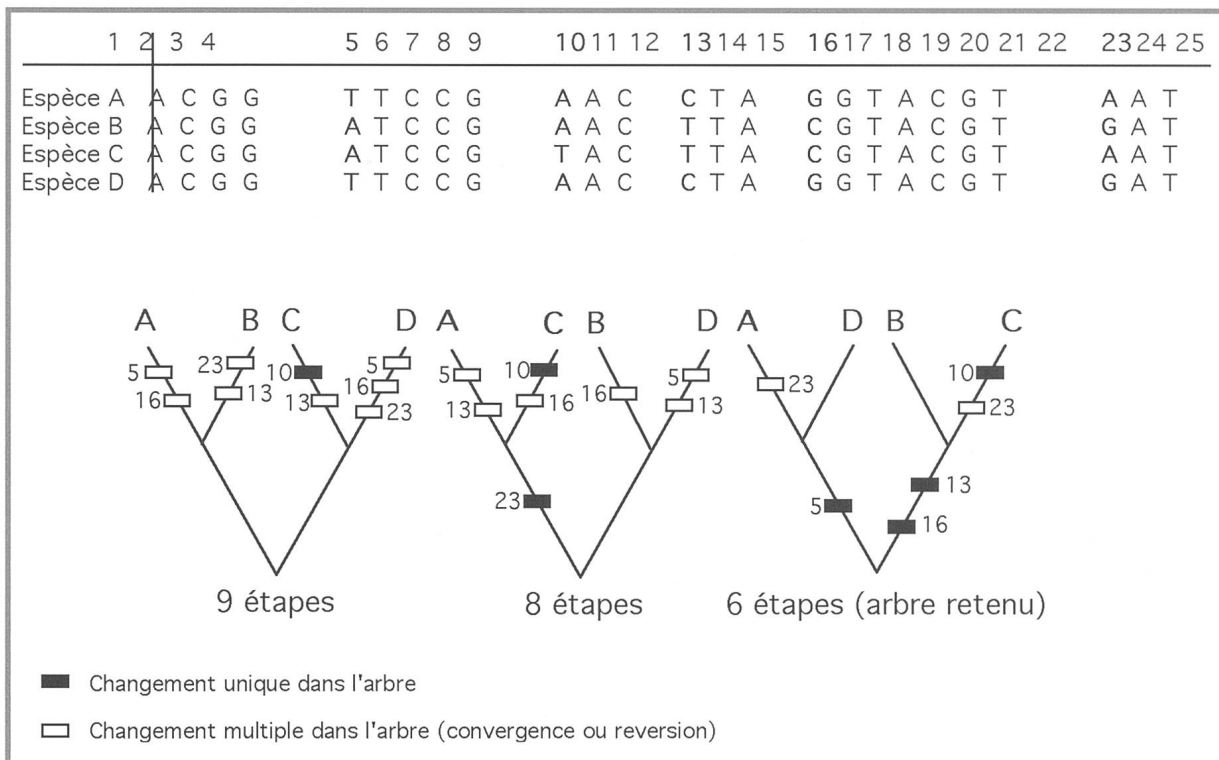


Figure 5. Exemple d'application du critère de parcimonie au choix d'un arbre phylogénétique parmi plusieurs possibles.

L'un des apports de la théorie de la coalescence a été de montrer que la date de l'ACPR est fondamentalement différente de la date de la spéciation comme l'illustre la figure 1. La période qui sépare les deux événements (dans le cas de deux gènes échantillonnés dans les espèces actuelles) est en moyenne de $2N_e$ générations, ce qui est loin d'être négligeable lorsque N_e est grand et que la spéciation est relativement récente (Edwards & Beerli, 2000). L'autre apport a été de proposer de formaliser les relations entre les différents gènes, non sous forme d'arbres hiérarchiques, mais sous forme de réseaux (fig. 2b). Les réseaux permettent en effet de prendre en compte les différents facteurs rendant caducs les arbres hiérarchiques, tels que la réticulation, la persistance d'allèles ou haplotypes ancestraux ou encore le fait que chaque allèle ou haplotype peut donner naissance à plusieurs allèles ou haplotypes dérivés par le biais de mutations différentes. Les résultats théoriques montrent ainsi qu'il y a une relation directe entre l'âge des allèles et leur fréquence dans la population. Plus un allèle aura une fréquence élevée dans la population et plus il sera largement distribué géographiquement (relativement à l'ensemble des populations que compte l'espèce), plus sa probabilité d'être ancien sera grande. De même les allèles anciens auront en moyenne davantage de connexions mutationnelles avec d'autres allèles que les allèles plus récents qui seront en périphérie de réseau (Posada & Crandall, 2001). Ces prédictions, directement tirées de la théorie de la coalescence, peuvent, par exemple, permettre d'enraciner des arbres phylogénétiques intraspécifiques, c'est-à-dire d'identifier les allèles ancêtres à la racine des arbres, ou encore de lever des ambiguïtés de l'arbre (Crandall & Templeton, 1993). Enfin, la théorie de la coalescence fournit plusieurs tests statistiques permettant d'évaluer la neutralité des marqueurs utilisés, sachant que cette neutralité est une des conditions de base de l'application de la théorie, au même titre que pour les reconstructions phylogénétiques.

Jusqu'à présent, la théorie de la coalescence a été surtout appliquée à la génétique humaine, dans le but de retracer l'histoire démographique passée et de trouver l'origine de l'Homme moderne (Sanchez-Mazas, 2001 ; Excoffier & Schneider, 1999). Mais cela n'est bien sûr pas restrictif et de nombreux exemples d'application de cette théorie au règne animal sont disponibles (Avisé, 1994). Les applications au règne végétal sont nettement moins nombreuses, en raison principalement de l'idée qui a longtemps prévalu, selon laquelle le génome chlorophyllien, principalement utilisé chez les plantes, n'est pas variable au sein des espèces, ce qui n'est pas toujours le cas, en fonction des fragments d'ADN utilisés (Taberlet *et al.*, 1991 ; Hamilton, 1999). Les applications sont cependant nombreuses, que ce soit pour délimiter des espèces, inférer des modes de spéciation, ou faire la part entre histoire et structure des populations (phylogéographie).

Références

- ANGIOSPERM PHYLOGENY GROUP (1998). An ordinal classification of the families of flowering plants. *Ann. Missouri Bot. Gard.*, 85, p. 531 - 553.
- AVISÉ, J. C. (1994). *Molecular Markers, Natural History and Evolution*. Chapman & Hall, New York.
- BARRACLOUGH, T. G. & S. NEE (2001). Phylogenetics and speciation. *Trends Ecol. Evol.*, 16 (7), p. 391 - 399.
- CRANDALL, K.A. & A. R. TEMPLETON (1993). Empirical tests of some predictions from coalescent theory with application to intraspecific phylogeny reconstruction. *Genetics*, 134, p. 959 - 969.
- DOYLE, J. J. (1992). Gene trees and species trees: molecular systematics as one-character taxonomy. *Syst. Bot.*, 17 (1), p. 144 - 163.
- EDWARDS, S. V. & P. BEERLI (2000). Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, 54 (6), p. 1839 - 1854.
- EXCOFFIER, L. (2000). Cours de génétique moléculaire des populations. http://anthro.unige.ch/GMDP/Laurent/GMDP_coal.htm
- EXCOFFIER, L. & S. SCHNEIDER (1999). Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. *Proc Natl. Acad. U.S.A.*, 96, p. 10597 - 10602.
- HAMILTON, M. B. (1999). Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molec. Ecol.*, 8, p. 513 - 525.
- HUDSON, R. R. (1990). Gene genealogies and the coalescent process. *Oxford Surv. Evol. Biol.*, 7, p. 1 - 44.
- JEANMONOD, D. (2000). Faut-il brûler votre 'Binz' et votre 'Flora Helvetica' ou que change la systématique moléculaire dans notre flore? *Saussurea*, 31, p. 65 - 72.
- KINGMAN, J. F. C. (1982). The coalescent. *Stochastic Process Appl.*, 13, p. 235 - 248.
- LEWIS, P. O. (2001). Phylogenetic systematics turns over a new leaf. *Trends Ecol. and Evol.*, 16 (1), p. 30 - 37.
- NICHOLS, R. (2001). Gene trees and species trees are not the same. *Trends Ecol. Evol.*, 16 (7), p. 358 - 364.
- PAGE, R. D. M. & M. A. CHARLESTON (1997). From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem. *Molec. Phylogen. Evol.*, 7 (2), p. 231 - 240.

- POSADA, D. & K. A. CRANDALL (2001). Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol. Evol.*, 16 (1), p. 37 - 45.
- SANCHEZ-MAZAS, A. (2001). Les origines de l'homme au cœur de ses gènes. *Pour la Science*, 289, p. 84 - 91.
- SPICHTER, R., V. SAVOLAINEN & M. FIGEAT (2000). *Botanique systématique des plantes à fleurs*. Presses Polytechniques et Universitaires Romandes. Lausanne, Suisse.
- TABERLET, P., L. GIELLY, G. PAUTOU & J. BOUVET (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Pl. Molec. Evol.*, 17, p. 1105 - 1109.

Remerciements

Les auteurs remercient les personnes ayant relu ce manuscrit, dont le Prof. André Langaney et le Dr. Estella Poloni

Texte
Yamama Naciri Graven
& Jean-François Manen

Laboratoire de phylogénie
et de génétique moléculaire

Conservatoire et Jardin
botaniques
1, chemin de l'Impératrice,
CP 60
1292 Chambésy, Genève,
Suisse

Synthèse