Table des matières

Introduction à la Phylogénie 3
Figure 1. L'Arbre de vie phylogénétique qui fait consensus 3
Les arbres phylogénétiques représentent les relations évolutives 4
Figure 2: Les parties d'un arbre phylogénétique. 5
Figure 3: Diverses représentations d'un arbre phylogénétique à 5 taxons. 5
Méthodes de détermination des arbres phylogénétiques: Cladistique et phylogénétique numérique 6
Parcimonie maximale 7
Probabilité maximale 7
Méthodes de distance 8
Support statistique pour les phylogénies 8
L'inférence phylogénétique trouve-t-elle les arbres appropriés? 9
Mises en garde avec l'inférence phylogénétique 10
Jargon Phylogénétique 11
Références 12

Introduction à la Phylogénie

Titre original : 29+ Evidences for Macroevolution - Phylogenetics Primer

Texte de Douglas Theobald, Ph.D.

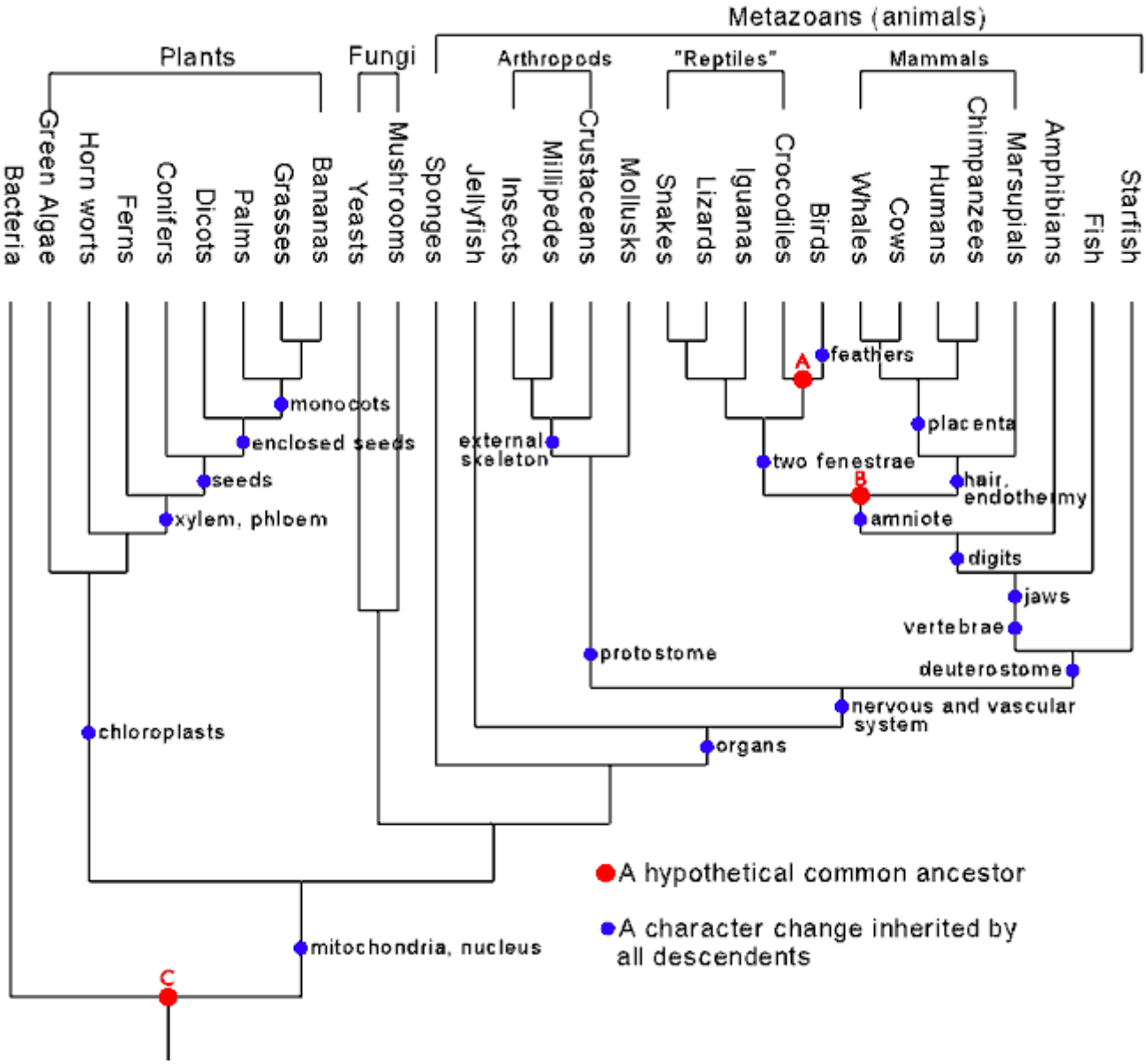
Copyright © 1999-2012

La page originale disponible à l'adresse http://www.talkorigins.org/faqs/comdesc/phylo.html

L'ascendance depuis un ancêtre commun implique un processus de ramification et de divergence, commun à tout processus généalogique. Les généalogies peuvent être illustrées graphiquement par des diagrammes ressemblant à des arbres. C'est pourquoi les biologistes appellent souvent la généalogie des espèces "l'arbre de la vie". Dans la théorie de l'évolution, de tels diagrammes sont appelés arbres phylogénétiques ou phylogénies. L'une des prédictions les plus importantes, les plus puissantes et les plus fondamentales de l'hypothèse de la descendance commune universelle est l'existence d'un arbre phylogénétique universel, historique et unique en son genre pour les espèces qui se reproduisent principalement par des mécanismes génétiques verticaux (un autre type d'héritage, le transfert horizontal de gènes, peut compliquer les phylogénies et même le concept d'espèce, voir Mises en garde ci-dessous). Une compréhension approfondie de la phylogénétique est nécessaire pour comprendre les déductions macro-évolutives. Le modèle qui fait consensus et que les biologistes de l'évolution utilisent pour représenter les branches bien étayées de l'arbre de vie universel, sera qualifié d '"arbre phylogénétique standard". La figure 1 montre un exemple simplifié de certaines des branches les plus connues de l'arbre phylogénétique universel.

Dans la section suivante, vous trouverez un bref aperçu des arbres phylogénétiques et de la façon dont les biologistes les déterminent. Cette vue d'ensemble devient de plus en plus technique à mesure qu'elle avance. Le contenu avant la rubrique *la parcimonie maximale* est essentiel pour comprendre le reste de cette FAQ. La discussion phylogénétique restante est donnée par souci de complétude et pour permettre au lecteur intéressé d'approfondir aussi loin que souhaité.

Figure 1. L'Arbre de vie phylogénétique qui fait consensus

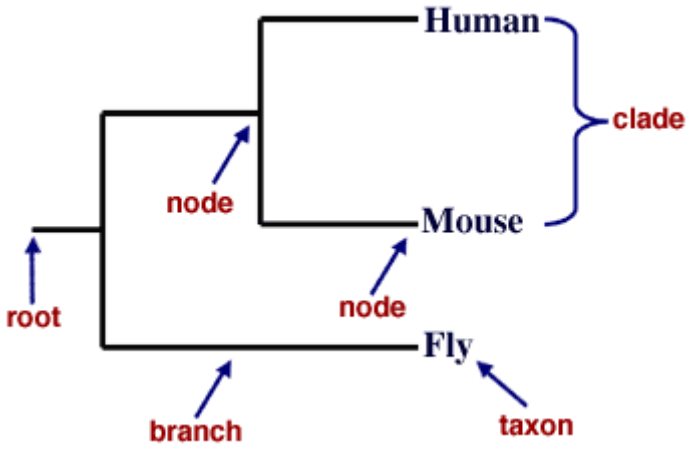


Les arbres phylogénétiques représentent les relations évolutives

La phylogénétique est la discipline scientifique chargée de décrire et de reconstruire les modèles de relations génétiques entre espèces et taxons supérieurs. Les arbres phylogénétiques sont un moyen pratique de représenter visuellement l'histoire évolutive de la vie. Ces diagrammes illustrent les relations inférées entre les organismes et l'ordre des événements de spéciation qui ont conduit des ancêtres communs antérieurs à leurs descendants diversifiés.

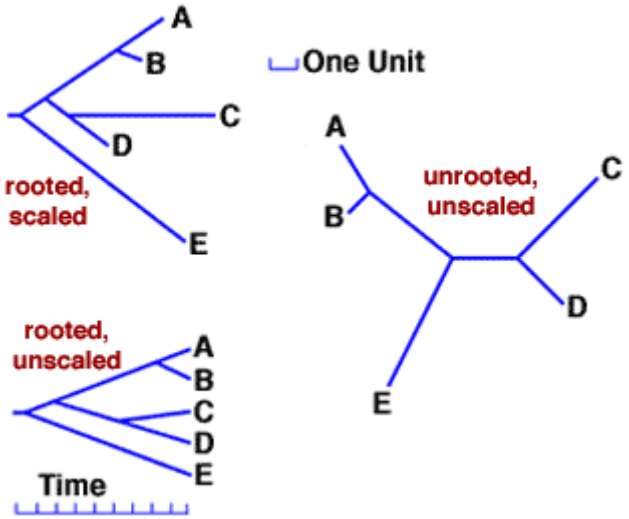
La figure 2 montre un arbre phylogénétique. Les nœuds représentent des unités taxonomiques, telles qu'un organisme, une espèce, une population, un ancêtre commun ou même un genre entier ou un autre groupe taxonomique supérieur. Les branches relient les nœuds de manière unique et représentent les relations génétiques. Le modèle spécifique de branchement détermine la topologie de l'arbre. Les arbres à échelle ont des longueurs de branches proportionnelles à certaines propriétés biologiques importantes, telles que le nombre de changements d'acides aminés entre les nœuds d'une phylogénie protéique (voir la figure 3). Les arbres peuvent également être enracinés ou non. Les arbres enracinés ont un nœud spécial, appelé racine, qui représente un ancêtre commun à tous les taxons affichés dans l'arbre. Les arbres enracinés sont donc directionnels, puisque tous les taxons ont évolué à partir de cette racine. Les arbres non racinés illustrent uniquement les relations, sans référence à des ancêtres communs.

Figure 2: Les parties d'un arbre phylogénétique.



Les taxons dans cet arbre sont “humain”, “souris” et “mouche” (tous leurs génomes ont été complètement séquencés). Plusieurs nœuds sont indiqués, tels que le nœud taxon “mouche” et un nœud interne qui représente l’ancêtre commun des souris et des humains. La racine est indiquée à gauche, représentant l'ancêtre commun des trois taxons énumérés.

Figure 3: Diverses représentations d’un arbre phylogénétique à 5 taxons.



Chacun de ces arbres représente les cinq mêmes taxons modernes: A, B, C, D et E. L'arbre en haut à gauche est enraciné et redimensionné en fonction de la distance évolutive. La racine est à gauche. Les taxons C et E ont tous deux subi des modifications relativement importantes depuis la divergence par rapport à la racine, contrairement aux taxons B et D. L'arbre situé en bas à gauche est enraciné et sans échelle. Ici, les longueurs de branches sont des indicateurs relatifs du temps écoulé depuis la divergence. L'arbre à droite est redimensionné mais non raciné. Dans cet arbre, alors que la racine est inconnue, les relations entre les taxons sont identiques à celles montrées dans les deux autres arbres

Une idée fausse commune est que certaines espèces modernes sont ancestrales à d'autres espèces modernes. Cependant, toutes les espèces modernes se trouvent à l'extrémité des branches de l'arbre, et une espèce moderne est aussi “évoluée” que toute autre. En effet, bien que l'on pense que les mammifères ont évolué à partir de quelque chose qui ressemble aux reptiles modernes, les reptiles modernes sont tout aussi “anciens” sur le plan de l'évolution que les mammifères modernes (Brooks 1991, p.68; Futuyma 1998, p.113).

Méthodes de détermination des arbres phylogénétiques: Cladistique et phylogénétique numérique

Of all clean birds ye shall eat.
But these are they of which ye shall not eat:

The eagle, and the ossifrage, and the ospray,
And the glede, and the kite, and the vulture after his kind,
And every raven after his kind,
And the owl, and the night hawk, and the cuckow, and the hawk after his kind,
The little owl, and the great owl, and the swan,
And the pelican, and the gier eagle, and the cormorant,
And the stork, and the heron after her kind, and the lapwing,
and the bat.

Deuteronomy 14:11-18, KJV

Si les espèces modernes descendent d'êtres ancestraux de cette manière arborescente et ramifiée, il devrait être possible de déduire le véritable arbre historique qui trace leurs chemins de descendance. Les biologistes ont déduit des phylogénie depuis que Darwin a proposé pour la première fois que la vie était unie par la descendance commune, il y a plus de 160 ans. Des méthodologies algorithmiques rigoureuses permettant d'inférer des arbres phylogénétiques sont utilisées depuis 60 ans.

En 1950, le taxonomiste Willi Hennig a proposé une méthode de détermination des arbres phylogénétiques basée sur la morphologie en classifiant les organismes en fonction de leurs caractères dérivés communs, appelés synapomorphies (Hennig, 1966). Cette méthode, maintenant appelée *cladistique*, n'assume pas a priori une parenté généalogique, puisqu'elle permet de classer n'importe quoi en principe, tel que des livres, des voitures ou des chaises qui ne sont évidemment pas reliées biologiquement (Kitching et al. 1998, Ch. 1, page 26;). Cependant, utilisant des arguments évolutifs fermes, Hennig a justifié cette méthode comme étant la technique de classification la plus appropriée pour estimer les relations évolutives générées par descendance linéaire. En fait, la méthode cladistique de Hennig n'est rien de plus qu'une formalisation des méthodes que les biologistes systématiques utilisaient de manière intuitive depuis que Linnaeus a écrit */Systema Naturae*. Les biologistes construisent aujourd'hui leurs arbres phylogénétiques sur la base de la méthode de Hennig. Grâce à la cladistique, ces arbres phylogénétiques sont reproductibles et peuvent être testés de manière indépendante (Brooks, 1991, Ch. 2; Kitching et al., 1998).

Les méthodes cladistiques sont souvent opposées aux méthodes "phénétiques". Les méthodes phénétiques regroupent et classifient les espèces en fonction du nombre de caractères identiques qu'elles partagent, c'est-à-dire en fonction de la similarité globale. De telles méthodes peuvent avoir des problèmes avec des organismes comme les dauphins et le thon, qui présentent de nombreuses similitudes superficielles. Cependant, ces organismes ne sont pas étroitement liés et ne devraient pas être classés ensemble si l'on s'attend à ce que la classification reflète la phylogénie.

En revanche, les phylogénies de type cladistique regroupent les taxons dans des hiérarchies imbriquées, et ils sont déterminés en utilisant uniquement des caractères dérivés d'organismes partagés, et non des caractères primitifs partagés (Brooks 1991, p. 35-36; Kitching et al. 1998, Ch. 1; Maddison et Maddison 1992, page 49). Dans le jargon phylogénétique technique, les caractères primitifs sont appelés plésiomorphies et les caractères dérivés, les apomorphies. Dans la cladistique, les espèces apparentées sont regroupées car elles partagent des caractères dérivés (apomorphies) ayant leur origine dans un ancêtre commun du groupe, mais n'étant pas présentes chez d'autres ancêtres du groupe. Ces fonctionnalités dérivées partagées sont appelées synapomorphies. Primitif et dérivé sont donc des termes relatifs, dépendant du groupe spécifique considéré. Par exemple, les vertèbres sont les caractères primitifs des vertébrés, tandis que les poils sont un caractère dérivé particulier aux vertébrés de mammifère. Cependant, lorsque l'on considère uniquement les mammifères, les poils sont primitifs, tandis qu'un pouce opposable est dérivé.

Dans les analyses phylogénétiques réelles, les caractères dérivés partagés peuvent être en conflit avec d'autres caractères dérivés. Ainsi, des méthodes objectives sont nécessaires pour résoudre ce conflit de caractères (Kitching et al. 1998, Ch. 1; Maddison et Maddison 1992, p. 49). Par exemple, les ailes sont un caractère dérivé d'oiseaux et de chauves-souris. Sur la base de ce seul caractère, la méthode cladistique regrouperait les chauves-souris et les oiseaux, ce que l'auteur du Deutéronome a regroupé dans la citation biblique ci-dessus. Cependant, d'autres caractères dérivés partagés indiquent que les chauves-souris devraient être regroupées avec des mammifères sans ailes et que les oiseaux devraient être regroupés avec des dinosaures sans ailes.

Au cours des 40 dernières années, plusieurs méthodes algorithmiques ont été conçues pour résoudre ce type de conflit de caractères et pour déduire des arbres phylogénétiques corrects (Felsenstein 2004, Ch. 10). Les sections suivantes décrivent certaines des méthodes les plus efficaces. Chaque méthode tente de déduire une phylogénie à partir de données existantes et chacune a ses forces et ses faiblesses. Des années d'essais empiriques et de simulations ont montré qu'en général, ces différents algorithmes, chacun reposant sur des hypothèses sous-jacentes très différentes, convergent vers des arbres très similaires, une fois analysés statistiquement (Li 1997, Chs 5 et 6; Nei et Kumar 2000, Chs 6 , 7 et 8).

Parcimonie maximale

L'une des méthodes les plus anciennes, les plus élémentaires et les plus fréquemment utilisées pour la résolution de caractères est le critère de parcimonie maximale (MP) (Edwards et Cavalli-Sforza, 1963; Kitching et al., 1998). Le critère de parcimonie exige que la meilleure arborescence décrivant les données soit celle qui minimise le nombre de conflits de caractères. Par exemple, considérons un ensemble de données contenant 10 caractères dérivés partagés qui regroupent les chauves-souris avec des singes (plutôt que des oiseaux) et avec un caractère qui les regroupe avec des oiseaux (plutôt que des singes). Selon le critère de parcimonie, l'arbre donnant le premier regroupement devrait être préféré.

Actuellement, la parcimonie est la méthode de choix pour la reconstruction d'arbres morphologiques (Kitching et al. 1998). Il est très rapide sur le plan des calculs et peut résister aux différences de taux d'évolution entre les caractères. Cependant, la parcimonie maximale ne trouve systématiquement la phylogénie correcte que lorsque nous nous attendons à ce que les conflits de caractères soient faibles ou que l'évolution évolue avec parcimonie (Felsenstein 2004, Ch. 9; Kitching et al. 1998, p. 17). Si les taux d'évolution sont lents et les branches courtes, les conflits de caractères seront faibles et la parcimonie fonctionnera bien (Felsenstein 2004, Ch. 9; Felsenstein 1981a; Li 1997, p. 128). Si les conflits de caractères sont modérés ou élevés en réalité, il est alors très peu probable que le véritable arbre présente le moins de conflits de caractères. Lorsque les taux d'évolution sont élevés, ou lorsque certaines branches sont très longues, ou lorsque le nombre d'états de caractère possibles est limité, les conflits de caractères peuvent être courants. Ceci est souvent vrai pour les séquences de nucléotides, qui n'ont que quatre états de caractère possibles (A, C, T ou G). Dans de tels cas, d'autres méthodes phylogénétiques peuvent être plus précises que la parcimonie.

Probabilité maximale

Un autre critère phylogénétique couramment utilisé est le maximum de vraisemblance (Maximum Likelihood - ML), une technique statistique efficace et robuste maintenant utilisée dans tous les domaines scientifiques (Edwards et Cavalli-Sforza 1964; Felsenstein 1981b; Fisher 1912). De nombreux estimateurs statistiques bien connus sont en réalité des estimateurs du maximum de vraisemblance. Par exemple, la moyenne d'un échantillon comme estimation de la signification d'une distribution gaussienne ou l'ajustement par la méthode des moindres carrés d'un ensemble de points sont tous deux des estimations du maximum de vraisemblance. En utilisant ML, on peut déduire les taux d'évolution directement à partir des données et déterminer l'arbre qui décrit le mieux ces données compte tenu de ces taux inférés. En d'autres termes, ML trouve l'arbre et les paramètres évolutifs qui produisent les données observées avec la probabilité la plus élevée. Contrairement à la parcimonie, ML trouve des arbres avec la quantité attendue de conflits de caractères étant donné les taux

d'évolution déduits des données, même si ces taux sont élevés. ML est une méthode de calcul intensive qui peut prendre beaucoup de temps.

Méthodes de distance

En raison de leur vitesse de calcul, les méthodes de matrice de distance sont parmi les plus populaires pour déduire des phylogénies (Nei et Kumar 2000, Ch. 6). Toutes les méthodes de distance transforment les données de caractères en une matrice de paires de distances, une distance pour chaque association possible des taxons étudiés. Les méthodes de matrice de distance ne sont pas cladistiques, car les informations sur les caractères dérivés et primitifs ont été perdues au cours de cette transformation. Les méthodes de distance abordent l'inférence phylogénétique strictement comme un problème statistique et sont utilisées presque exclusivement avec des données moléculaires. Bien qu'elles ne soient pas cladistiques, les méthodes de distance peuvent être considérées comme des approximations des méthodes cladistiques, et il est garanti que plusieurs méthodes convergent mathématiquement vers l'arborescence correcte à mesure que davantage de données sont incluses.

La métrique de distance la plus simple est simplement le nombre de différences de caractères entre deux taxons, tel que le nombre de différences de nucléotides entre deux séquences d'ADN. Il existe de nombreuses autres manières de calculer les distances de séquence moléculaire et la plupart tentent de corriger la possibilité de multiples changements sur un même site au cours de l'évolution. Les méthodes de calcul des distances entre les séquences portent généralement le nom de leurs auteurs, telles que les deux paramètres de Kimura (K2P), Jukes-Cantor (JC), Tamura-Nei (TN), Hasegawa, Kishino et Yano (HKY) et Felsenstein 1984. (F84). Les autres mesures de distance importantes sont General Time Reversible (GTR) et LogDet (Felsenstein 2004, pp. Chs 11 et 13; Nei et Kumar 2000, Chs 2 et 3; Li 1997, Chs 3 et 4).

Une fois qu'une matrice de distance pour les taxons considérés est disponible, il existe plusieurs critères basés sur la distance et d’algorithmes qui peuvent être utilisés pour estimer l'arbre phylogénétique à partir des données (Felsenstein 2004, Ch. 11; Li 1997, Ch. 5). Le critère d'évolution minimum (ME) détermine l'arbre dans lequel la somme de toutes les longueurs de branche est la plus petite. Les critères des moindres carrés pondérés et non pondérés permettent de calculer l'écart entre les distances par paire observées et les distances par paire calculées à partir des longueurs de branche de l'arbre inféré. Les moindres carrés trouvent ensuite l'arbre qui minimise le carré de cet écart. Les méthodes des moindres carrés sont parmi les plus justifiées sur le plan statistique et convergeront vers l'arborescence correcte à mesure que davantage de données seront incluses dans l'analyse (à partir d'une métrique de distance mathématiquement correcte). L'algorithme de jonction entre voisins (NJ) est extrêmement rapide et constitue une approximation des méthodes des moindres carrés et de l'évolution minimale. Si la matrice de distance est une description exacte de l'arbre véritable, alors la jonction entre voisins garanti de reconstruire l'arbre correct. L'algorithme de classification UPGMA (un acronyme déroutant) est également extrêmement rapide, mais il repose sur l'hypothèse improbable selon laquelle les taux d'évolution sont égaux dans toutes les lignées. L'UPGMA est rarement utilisé aujourd'hui, sauf comme outil pédagogique.

Support statistique pour les phylogénies

Une phylogénie est la meilleure approximation de l'arbre correct et historique en utilisant une méthode phylogénétique donnée. Certaines analyses phylogénétiques sont fortement étayées par les données, certaines sont faiblement étayées et différentes parties d’un arbre peuvent avoir plus solides que d’autres. Lorsque l'on compare deux phylogénies déterminées indépendamment, il faut prendre en compte le support statistique attribué à chaque branche des phylogénies. Comme avec toutes les analyses scientifiques, les détails d'un arbre phylogénétique peuvent changer à mesure que de nouvelles informations et données sont incorporées (Maddison et Maddison 1992, p. 112-123; Li 1997, p. 36-146; Felsenstein 1985; Futuyma 1998, p. 99; Hillis et Bull, 1993; Huelsenbeck et al., 2001; Swofford et al, 1996, p. 504-509).

Le *bootstrap* est la méthode statistique la plus répandue pour évaluer la fiabilité des branches d'un arbre

phylogénétique (Felsenstein, 1985). Le bootstrap est une technique statistique permettant d'estimer empiriquement la variabilité d'un paramètre (Efron, 1979; Efron et Gong, 1983). Dans une analyse bootstrap, un jeu de données fictif est créé en échantillonnant de manière aléatoire les données du jeu de données réel jusqu'à ce qu'un nouveau jeu de données de la même taille soit créé. Ce processus est effectué à plusieurs reprises (des centaines ou des milliers de fois) et le paramètre d'intérêt est estimé à partir de chaque jeu de données fictif. La variabilité de ces estimations bootstrap est elle-même une estimation de la variabilité du paramètre d'intérêt.

En phylogénétique, une nouvelle phylogénie est déduite de chaque jeu de données bootstrap (Felsenstein, 1985). Ces phylogénies bootstrap auront probablement des topologies différentes. À partir de ces différents arbres bootstrapés, la variabilité de l'arbre inféré peut être estimée. Une confiance élevée est attribuée aux parties des arbres bootstrap qui sont communes, alors qu'une confiance faible est attribuée aux parties qui varient considérablement. Les arbres construits à partir de données aléatoires ne génèrent pas d'arbres ou de branches de confiance élevée lors du bootstrap. Ainsi, le bootstrap fournit un moyen de vérifier si un arbre phylogénétique est authentique.

L'inférence phylogénétique trouve-t-elle les arbres appropriés?

Afin de déterminer leur validité dans la détermination fiable de phylogénie, des méthodes phylogénétiques ont été testées empiriquement dans les cas où la véritable phylogénie est connue avec certitude, puisque la vraie phylogénie a été directement observée.

- Le bactériophage T7 a été propagé et scindé séquentiellement en présence d'un mutagène, où chaque lignée a été suivie. Sur 135 135 arbres phylogénétiques possibles, le véritable arbre a été correctement déterminé par des méthodes phylogénétiques lors d'une analyse à l'aveugle. Cinq méthodes phylogénétiques différentes ont été utilisées indépendamment, et chacune d'elles a choisi l'arbre approprié (Hillis et al, 1992).
- Dans une autre étude, 24 souches de souris ont été utilisées dans lesquelles les relations généalogiques étaient connues. L'analyse cladistique a reproduit presque parfaitement la phylogénie connue des 24 souches (Atchely et Fitch, 1991).
- Bush et al. a utilisé une analyse phylogénétique pour prédire de manière rétrospective l'arbre évolutif correct du virus humain de la grippe A dans 83% des saisons de propagation allant de 1983 à 1994.
- En 1998, les chercheurs ont utilisé 111 séquences modernes du VIH-1 (virus du sida) lors d'une analyse phylogénétique pour prédire la séquence nucléotidique de l'ancêtre viral dont ils étaient tous les descendants. La séquence ancestrale prédite correspondait étroitement, avec une probabilité statistique élevée, à une séquence VIH ancestrale réelle trouvée dans un échantillon de plasma africain séropositif au VIH-1 collecté et archivé au Congo belge en 1959 (Zhu et al., 1998).
- Au cours de la dernière décennie, les analyses phylogénétiques ont joué un rôle important dans le succès des condamnations dans plusieurs affaires pénales (Albert et al. 1994; Arnold et al. 1995; Birch et al. 2000; Blanchard et al. 1998; Goujon et al. 2000; Holmes et al. 1993; Machuca et al. 2001; Ou, 1992; Veenstra et al. 1995; Vogel, 1997; Yirrell et al. 1997), et les reconstructions phylogénétiques ont maintenant été admises en tant que témoignage juridique d'expert aux États-Unis. (97-KK- 2220 State of Louisiana v. Richard J. Schmidt [PDF]) Aux États-Unis, le critère juridique applicable à la recevabilité du témoignage d'expert est constitué par les directives Daubert (U. S. Supreme Court Case Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 587-89, 113 S. Ct. 2786, 2794, 125 L. Ed. 2d 469, 1993). Les directives de Daubert stipulent qu'un tribunal de première instance devrait prendre en compte cinq facteurs pour déterminer "si le raisonnement ou la méthodologie sous-tendant le témoignage est scientifiquement valable": (1) si la théorie ou la technique en question peut être et a été testée; (2) s'il a été soumis à un examen par des pairs et à une publication; (3) son taux d'erreur connu ou potentiel; (4) l'existence et le maintien de normes contrôlant son fonctionnement; et (5) s'il a été largement accepté par la communauté scientifique concernée (cité presque mot pour mot). L'analyse phylogénétique a officiellement satisfait à ces exigences légales.

Mises en garde avec l’inférence phylogénétique

Comme pour toute méthode scientifique expérimentale, certaines conditions doivent être remplies pour que les résultats soient fiables. Un principe commun à de nombreuses méthodes phylogénétiques moléculaires est que les gènes sont transmis via un héritage vertical, linéaire, c'est-à-dire du parent à la progéniture. Si cette prémisse n’est pas respectée, les arbres géniques ne récapituleront pas une phylogénie d'organisme ou d'espèce. Cette hypothèse est violée dans les cas de transfert horizontal, par ex. dans la transformation d'une bactérie par un plasmide à ADN ou dans l'insertion rétrovirale dans le génome d'un hôte. Au début de l'évolution de la vie, avant l'avènement des organismes multicellulaires, le transfert horizontal était probablement très fréquent (comme c'est le cas aujourd'hui dans l'évolution observée des bactéries et d'autres organismes unicellulaires). Ainsi, on peut se demander si les méthodes phylogénétiques moléculaires sont applicables, même en principe, à la résolution des modèles évolutifs de nombreux microbes, y compris l’évolution précoce près du plus récent ancêtre commun de tous les organismes vivants (Doolittle 1999; Doolittle 2000; Woese 1998).

La liste ci-dessous présente certaines des mises en garde les plus importantes que les scientifiques doivent garder à l'esprit lors de l'interprétation des résultats d'une analyse phylogénétique (Swofford 1996, p. 493-509). En général, la contribution de chacune de ces préoccupations sera “moyennée” en incluant plus de caractères indépendants dans l'analyse phylogénétique, tels que plus de gènes et des séquences plus longues.

- Caractères corrélés: chaque caractère utilisé dans l'analyse de manière optimale doit être génétiquement indépendant. Les caractères fortement corrélés fonctionnellement sont mieux perçus comme un seul caractère. Il existe des tests statistiques permettant de contrôler les corrélations de caractères non reconnus, tels que le bloc bootstrap et le jackknife.
- Véritable convergence structurelle: les structures qui ont subi une évolution convergente peuvent aboutir artificiellement à une topologie arborescente incorrecte. Inclure plus de caractères dans l'analyse aide également à surmonter les effets convergents.
- Inversion de caractères: les caractères qui reviennent à un état ancestral posent un problème similaire à la convergence. Comme l’ADN et l’ARN n’ont que quatre états de caractères différents, ils sont particulièrement sujets aux inversions au cours de l’évolution.
- Caractères perdus: les lignages qui ont perdu des caractères (tels que les baleines et leurs membres postérieurs) peuvent également poser des problèmes cladistiques. Souvent, si une analyse cladistique indique clairement qu'un certain caractère a été perdu au cours de l'évolution, il est préférable de l'omettre lors d'analyses à grande échelle de cette lignée.
- Caractères manquants: les fossiles incomplets sont problématiques, car ils peuvent ne pas comporter de caractères importants. De meilleurs fossiles sont la réponse.
- Nombre incalculable d'arbres phylogénétiques possibles: pour des raisons de calcul, il s'agit de l'un des défis phylogénétiques les plus importants à surmonter. Le but d'une reconstruction phylogénétique est de déterminer le meilleur arbre supporté par les données. Pour une analyse de seulement cinq espèces, il y a 15 arbres possibles. Pour une analyse de 50 espèces, il y a plus de 1074 arbres possibles qui doivent être recherchés, ce qui est impossible en termes de calcul. Ce problème n’est pas aussi grave qu’il semble au premier abord, car la réduction du nombre d’arbres raisonnables peut être triviale dans de nombreux cas (par exemple, en utilisant l’algorithme de séparation et d’évaluation). Plusieurs méthodes ont été développées pour résoudre ce problème avec succès, et au final, les ordinateurs sont devenus plus puissants.
- Hypothèses du maximum de vraisemblance: la méthode du maximum de vraisemblance émet des hypothèses explicites sur le schéma des substitutions de nucléotides basé sur un modèle donné d'évolution des nucléotides. Ces hypothèses reposent sur une base statistique solide; cependant, la validité des modèles doit être prise en compte lors de l'évaluation des résultats.
- Attraction de longues branches: les lignages qui ont divergé il y a relativement longtemps auront tendance à se “regrouper” dans une reconstruction phylogénétique dans certaines conditions. Les raisons

mathématiques sont quelque peu compliquées, mais l'utilisation de gènes (ou de régions de gènes) évoluant plus lentement permet de résoudre le problème.

* Variation du taux entre les lignages: les taux de substitution de nucléotides peuvent différer d'un lignage à l'autre; cela peut contribuer à attirer de longues branches et à créer des topologies d'arbre incorrectes. Cependant, les méthodes du maximum de vraisemblance et des moindres carrés sont particulièrement utiles ici.

- Variation du taux au sein d'un seul gène: les taux de substitution de nucléotides peuvent varier sur la longueur d'un seul gène, ce qui exacerbe également l'attraction de longues branches.
- Les arbres à gènes ne sont pas équivalents aux arbres à espèces: de la simple génétique mendélienne, nous savons que les gènes se séparent individuellement et que les gènes individuels ne suivent pas nécessairement la généalogie de l'organisme (Avise et Wollenberg 1997; Fitch 1970; Hudson 1992; Nichols 2001; Wu 1991).). Un exemple évident est le fait que, même si vous avez les yeux bruns, votre enfant peut posséder les gènes pour les yeux bleus - mais cela ne signifie pas que votre enfant ne fait pas partie de votre descendance, ni que vos enfants aux yeux bruns sont plus proches de vous que vos enfants aux yeux bleus. Inclure plusieurs gènes dans l'analyse est une solution à ce problème. Sur la base de simples calculs génétiques, une analyse de plus de cinq gènes est généralement nécessaire pour reconstruire avec précision la phylogénie d'une espèce (Wu, 1991).

Pour plus d'informations sur la cladistique, vous pouvez consulter l'une des excellentes ressources cladistiques en ligne, telles que SASB Introduction to Phylogenetics, le UC Berkeley's Integrative Biology Phylogenetics Lab, or Diana Lipscomb's stellar Basics of Cladistic Analysis, téléchargeables au format PDF d'Adobe Acrobat. Une bonne et concise description pour les profane sest disponible dans le Journal of Avocational Paleontology. Enfin, vous pouvez lire l'explication de Charles Darwin de "l'Arbre de la vie"dans *L'origine des espèces*, où le concept d'arbre phylogénétique a été introduit.

Jargon Phylogénétique

apomorphie: caractère dérivé d'un groupe d'organismes, non partagé avec les ancêtres du groupe d'organismes. Les apomorphies sont uniques au groupe et définissent donc ce groupe.

bootstrap: procédure statistique technique permettant d'estimer la variabilité d'une mesure. En phylogénétique, le bootstrap implique la production d'un nouvel ensemble de pseudo-données en extrayant de manière aléatoire des points de données de l'ensemble de données d'origine. Une nouvelle phylogénie est déduite pour chaque pseudo-jeu de données. Les séries de données fournissent une estimation des régions de la phylogénie d'origine qui sont bien ou mal supportées.

caractère: caractéristique observable d'un organisme utile pour le distinguer d'un autre. Par exemple, un nucléotide dans une séquence d'ADN, un acide aminé dans une séquence protéique ou des caractères morphologiques tels que poils, plumes ou la présence ou l'absence de certains os.

cladistique: classe de techniques phylogénétiques qui construisent des arbres (cladogrammes) en regroupant des taxons dans des hiérarchies imbriquées en fonction de caractères dérivés partagés (synapomorphies). La cladistique est étroitement associée au critère de parcimonie.

cladogramme: classification hiérarchique des taxons représentés sous forme d'arbre. Les cladogrammes sont formellement indépendants de la théorie de l'évolution, bien qu'en pratique ils soient généralement interprétés comme des phylogénies.

caractère dérivé: Voir apomorphie.

moindres carrés: critère de matrice de distance phylogénétique. Le meilleur arbre est celui avec la plus petite différence au carré entre les distances par paires observées et les distances calculées à partir de l'arbre inféré. Il a une forte justification statistique, car il est basé sur la technique statistique commune linéaire des moindres carrés. Le théorème de Gauss-Markov garantit que les moindres carrés convergent vers la réponse

correcte, lorsque davantage de données sont incluses dans l'analyse si une métrique de distance appropriée est utilisée, c'est-à-dire si les moindres carrés sont cohérents sur le plan statistique. Les versions pondérées corrigent la variabilité aléatoire et les biais dus aux longueurs de branche plus longues.

maximum de vraisemblance: critère cladistique pour déduire des arbres avec des conflits de caractères. Le meilleur modèle d'arbre et d'évolution permet de maximiser la probabilité des données observées. Le maximum de vraisemblance repose sur une base statistique solide. Avec un modèle correct de changement évolutif, sa cohérence statistique est garantie, c'est-à-dire qu'il convergera vers l'arbre correct au fur et à mesure que de nouvelles données seront ajoutées. Le maximum de vraisemblance donne généralement les meilleurs résultats de toutes les méthodes dans les simulations, mais il est très coûteux en calcul. Contrairement à la parcimonie, il repose explicitement sur un modèle évolutif spécifique.

évolution minimale: critère de matrice de distance phylogénétique. Le meilleur arbre est celui dans lequel la somme des longueurs de branche est la plus petite.

neighbor-joining: Un algorithme de matrice de distance pour déduire des arbres. C'est une approximation des méthodes des moindres carrés et de l'évolution minimale.

noeud: Un point dans une phylogénie où les branches se rencontrent ou se terminent. Les nœuds à la pointe ou à la fin d'une branche représentent des taxons. Dans les arbres enracinés, les nœuds internes représentent des ancêtres communs.

parcimonie: critère phylogénétique permettant d'inférer des arbres avec des conflits de caractères. La parcimonie nécessite que le meilleur arbre soit celui avec le moins de conflits de caractères. Il est connu pour produire une phylogénie incorrecte dans certains cas, comme lorsque les taux d'évolution sont élevés ou que certaines branches sont longues.

phénétique: Parfois connues sous le nom de taxonomie numérique, les méthodes phénétiques classifient et groupent les organismes sur la base de leur similarité globale, généralement sans référence explicite à leurs relations phylogénétiques.

phylogénie: diagramme en forme d'arbre représentant les relations généalogiques entre les taxons. Les phylogénies enracinées spécifient des ancêtres communs et ont un axe temporel.

plésiomorphie: caractère primitif d'un groupe d'organismes, partagé avec leurs ancêtres. Comme il est commun à plus que le groupe considéré, une plésiomorphie ne définit pas le groupe.

caractère primitif: Voir plésiomorphie.

racine: Un ancêtre commun de tous les taxons dans une phylogénie. Chronologiquement, la racine est le noeud le plus ancien.

synapomorphie: caractère dérivé partagé par deux groupes d'organismes.

UPGMA: Méthode de classification basée sur une matrice de distance pour la construction d'arbres. Rarement utilisé, il est très rapide mais suppose des taux d'évolution constants dans tout l'arbre (une propriété appelée ultramétricité).

Références

Albert, J., Wahlberg, J., Leitner, T., Escanilla, D. and Uhlen, M. (1994) "Analysis of a rape case by direct sequencing of the human immunodeficiency virus type 1 pol and gag genes." J Virol 68: 5918-24. PubMed]

Arnold, C., Balfe, P. and Clewley, J. P. (1995) "Sequence distances between env genes of HIV-1 from individuals infected from the same source: implications for the investigation of possible transmission events." Virology 211: 198-203. PubMed]

Atchely, W. R., and Fitch, W. M. (1991) “Gene trees and the origins of inbred strains of mice.” Science 254: 554-558. PubMed]

Avise, J. C., and Wollenberg, K. (1997) “Phylogenetics and the origin of species.” PNAS 94: 7748-7755. http://www.pnas.org/cgi/ content/full/94/15/7748

Birch, C. J., McCaw, R. F., Bulach, D. M., Revill, P. A., Carter, J. T., Tomnay, J., Hatch, B., Middleton, T. V., Chibo, D., Catton, M. G., Pankhurst, J. L., Breschkin, A. M., Locarnini, S. A. and Bowden, D. S. (2000) “Molecular analysis of human immunodeficiency virus strains associated with a case of criminal transmission of the virus.” J Infect Dis 182: 941-4. http://jid.oxfordjournals.org/content/182/3/941.long

Blanchard, A., Ferris, S., Chamaret, S., Guetard, D. and Montagnier, L. (1998) “Molecular evidence for nosocomial transmission of human immunodeficiency virus from a surgeon to one of his patients.” J Virol 72: 4537-40. http://jvi.asm.org/cgi/content/full/72/5/4537?view=full&pmid=9557756

Brooks, D. R., and McLennan, D. A. (1991) Phylogeny, ecology, and behavior. Chicago: University of Chicago Press.

Bush, R. M., C. A. Bender, et al. (1999) “Predicting the evolution of human influenza A.” Science 286: 1921-1925. PubMed]

Doolittle, W. F. (1999) “Phylogenetic Classification and the Universal Tree.” Science 284: 2124. PubMed]

Doolittle, W. F. (2000) “The nature of the universal ancestor and the evolution of the proteome.” Current Opinion in Structural Biology 10: 355-358. PubMed]

Edwards, A. W. F. and Cavalli-Sforza, L. L. (1963) “The reconstruction of evolution.” Annals of Human Genetics 27: 105-106.

Efron, B. (1979) “Bootstrap methods: Another look at the jackknife.” Annals of Statistics 7: 1-26.

Efron, B. and Gong, G. (1983) “A leisurely look at the bootstrap, the jackknife, and cross validation.” American Statistician 37: 36-48.

Edwards, A. W. F. and Cavalli-Sforza, L. L. (1964) “Reconstruction of phylogenetic trees.” in Phenetic and Phylogenetic Classification. ed. Heywood, V. H. and McNeill. London: Systematics Assoc. Pub No. 6.

Felsenstein, J. (1981) “A likelihood approach to character weighting and what it tells us about parsimony and compatibility.” Biol J Linn Soc Lond 16: 183-196.

Felsenstein, J. (1981) “Evolutionary trees from DNA sequences: A maximum likelihood approach.” J Mol Evol 17: 368-376. PubMed]

Felsenstein, J. (1985) “Confidence limits on phylogenies: an approach using the bootstrap.” Evolution 39: 783-791.

Felsenstein, J. (2004) Inferring Phylogenies. Sunderland, MA: Sinauer Associates.

Fisher, R. A. (1912) “On an absolute criterion for fitting frequency curves.” Messenger of Mathematics 41: 155-160.

Fitch, W. M. (1970) “Distinguishing homologous from analogous proteins.” Syst. Zool. 28: 132-163.

Futuyma, D. (1998) Evolutionary Biology. Third edition. Sunderland, MA: Sinauer Associates.

Goujon, C. P., Schneider, V. M., Grofti, J., Montigny, J., Jeantils, V., Astagneau, P., Rozenbaum, W., Lot, F., Frocrain-Herchkovitch, C., Delphin, N., Le Gal, F., Nicolas, J. C., Milinkovitch, M. C. and Deny, P. (2000) “Phylogenetic analyses indicate an atypical nurse-to-patient transmission of human immunodeficiency virus type 1.” J Virol 74: 2525-32. http://jvi.asm.org/cgi/content/full/74/6/2525?view=full&pmid=10684266

Hennig, W. (1966) Phylogenetic Systematics. (English Translation). Urbana: University of Illinios Press.

Hillis, D. M., and Bull, J. J. (1993) “An empirical test of bootstrapping as a method for assessing confidence on phylogenetic analysis.” Syst. Biol. 42: 182-192.

Hillis, D. M., J. J. Bull, et al. (1992) “Experimental phylogenetics: Generation of a known phylogeny.” Science 255: 589-592. PubMed]

Holmes, E. C., Zhang, L. Q., Simmonds, P., Rogers, A. S. and Brown, A. J. (1993) “Molecular investigation of human immunodeficiency virus (HIV) infection in a patient of an HIV-infected surgeon.” J Infect Dis 167: 1411-4. PubMed]

Hudson, R. R. (1992) “Gene trees, species trees and the segregation of ancestral alleles.” Genetics 131: 509-513. PubMed]

Huelsenbeck, J. P., Ronquist, F., Nielsen, R., and Bollback, J. P. (2001) “Bayesian inference of phylogeny and its impact on evolutionary biology.” Science 294: 2310-2314. [PubMed]

Kitching, I. J., Forey, P. L., Humphries, C. J., and Williams, D. M. (1998) Cladistics: The Theory and Practice of Parsimony Analysis. Second Edition. The Systematics Association Publication No. 11. Oxford: Oxford University Press.

Li, W.-H. (1997) Molecular Evolution. Sunderland, MA: Sinauer Associates.

Machuca, R., Jorgensen, L. B., Theilade, P. and Nielsen, C. (2001) “Molecular investigation of transmission of human immunodeficiency virus type 1 in a criminal case.” Clin Diagn Lab Immunol 8: 884-90. PubMed]

Maddison, W. P., and Maddison, D. R. (1992) MacClade. Sunderland, MA: Sinauer Associates.

Nei, M. and Kumar, S. (2000) Molecular Evolution and Phylogenetics. New York, NY: Oxford University Press.

Nichols, R. (2001) “Gene trees and species trees are not the same.” Trends Ecol Evol. 16: 358-364. PubMed]

Ou, C. Y., Ciesielski, C. A., Myers, G., Bandea, C. I., Luo, C. C., Korber, B. T., Mullins, J. I., Schochetman, G., Berkelman, R. L., Economou, A. N. and et al. (1992) “Molecular epidemiology of HIV transmission in a dental practice.” Science 256: 1165-71. PubMed]

Swofford, D. L., Olsen, G. J., Waddell, P. J., and Hillis, D. M. (1996) “Phylogenetic inference.” In Molecular Systematics, pp 407-514. Hillis, D. M., Moritiz, C. and Mable, B. K. eds., Sunderland, Massachusetts: Sinauer.

Veenstra, J., Schuurman, R., Cornelissen, M., van't Wout, A. B., Boucher, C. A., Schuitemaker, H., Goudsmit, J. and Coutinho, R. A. (1995) “Transmission of zidovudine-resistant human immunodeficiency virus type 1 variants following deliberate injection of blood from a patient with AIDS: characteristics and natural history of the virus.” Clin Infect Dis 21: 556-60. PubMed]

Vogel, G. (1997) “Phylogenetic analysis: getting its day in court.” Science 275: 1559-60. PubMed]

Woese, C. (1998) “The universal ancestor.” PNAS 95: 6854-6859. http://www.pnas.org/cgi/content/full/95/12/6854

Wu, C. I. (1991) “Inferences of species phylogeny in relation to segregation of ancient polymorphisms.” Genetics 127: 429-435. PubMed]

Yirrell, D. L., Robertson, P., Goldberg, D. J., McMenamin, J., Cameron, S. and Leigh Brown, A. J. (1997) “Molecular investigation into outbreak of HIV in a Scottish prison.” Bmj 314: 1446-50. http://bmj.com/cgi/content/full/314/7092/1446?view=full&pmid=9167560

Zhu, T., B. Korber, et al. (1998) “An African HIV-1 sequence from 1959 and implications for the origin of the epidemic.” Nature 391: 594-597. PubMed]

From:
https://www.evowiki.fr/ - **EvoWiki**

Permanent link:
https://www.evowiki.fr/comdesc-phylo?rev=1575101080

Last update: **2019/11/30 09:04**