Table des matières

Partie 4: La Preuve Moléculaire 3
Prédiction 4.1: Redondance fonctionnelle des protéines 3
Discussion: 4
Confirmation: 5
Critiques 6
Résumé: 6
Réfutation potentielle: 6
Prédiction 4.2: Redondance du codage de l'ADN 7
Confirmation: 7
Réfutation potentielle: 7
Prédiction 4.3: Preuve moléculaire – Les transposons 7
Confirmation: 8
Réfutation potentielle: 9
Prédiction 4.4: Preuve moléculaire - Pseudogènes redondants 9
Confirmation: 10
Réfutation potentielle: 10
Prédiction 4.4: Preuve moléculaire - Retrovirus endogènes 11
Confirmation: 11
Figure 4.4.1. Insertion de rétrovirus endogène humain K (HERV-K) dans des localisations chromosomiques identiques chez différents primates. 11
Réfutation potentielle: 12
Références 12

Titre original : **29+ Evidences for Macroevolution - Part 4: The Molecular Sequence Evidence**

Auteur : Douglas Theobald, Ph.D.

Copyright © 1999-2012 - Version 2.89

Permission is granted to copy and print these pages in total for non-profit personal, educational, research, or critical purposes.

Page originale disponible à l'adresse : http://www.talkorigins.org/faqs/comdesc/section4.html
Navigation :

- Introduction
  - Preuve scientifique et méthode scientifique
  - Glossaire
- Introduction à la phylogénétique
- Partie 1. Un arbre phylogénétique unique et historique
- Partie 2. Histoire passée
- Partie 3. Opportunisme évolutionniste
- Partie 4. Preuve moléculaire
- Partie 5. Changement
- Remarques finales

Partie 4: La Preuve Moléculaire

Nous avons obtenu des estimations de différenciation génétique entre l'homme et les grands singes ne dépassant pas, par exemple, celles observées entre des espèces de mouches à fruits physiquement indistinctes. " Elizabeth J. Bruce et Francisco J. Ayala *"Les humains et les singes sont génétiquement très semblables"* Nature 276: 264, 16 novembre 1978

La preuve de séquence moléculaire fournit la preuve la plus impressionnante et la plus irréfutable du lien de parenté de toute vie. La nature des séquences moléculaires permet des calculs de probabilité extrêmement impressionnants qui démontrent à quel point les prédictions de descendance commune avec modification correspondent réellement aux observations empiriques. La descendance commune est une déduction qui découle directement de prémisses basées sur des preuves moléculaires observées empiriquement. En outre, la connaissance des mécanismes et des structures moléculaires biologiques, associée à la théorie macro-évolutive, a donné des prévisions biomoléculaires très spécifiques, nouvelles et testables.

Prédiction 4.1: Redondance fonctionnelle des protéines

Le soutien à la descendance commune donné par les études de séquences moléculaires peut être formulé comme un argument déductif. Cet argument est unique dans cette FAQ, car il s'agit du seul cas où nous pouvons directement conclure que la similarité implique une relation. Cette conclusion dépend de la similitude des structures biologiques dans un contexte spécifique: la similitude observée entre des gènes omniprésents de différentes espèces.

La discussion suivante est quelque peu technique, elle est donc d'abord présentée dans ses grandes lignes comme un argument déductif, ce qui facilite le suivi de la logique. Voici la liste des prémisses de l'argumentation suivie de la conclusion et de la discussion ultérieure.

L'essentiel de l'argument:

- (P1) Gènes omniprésents: Tous les organismes vivants possèdent certains gènes car ils remplissent des fonctions vitales fondamentales; ces gènes sont appelés gènes omniprésents.
- (P2) Les gènes omniprésents ne sont pas corrélés avec des phénotypes spécifiques à une espèce: les gènes omniprésents n'ont aucune relation avec les fonctions spécifiques d'espèces différentes. Par exemple, peu importe que vous soyez une bactérie, un humain, une grenouille, une baleine, un colibri, une limace, un champignon ou une anémone de mer - vous avez ces gènes omniprésents qui ont tous la même fonction biologique de base, peu importe ce que vous êtes.
- (P3) Les séquences moléculaires de gènes omniprésents sont fonctionnellement redondantes: toute protéine omniprésente donnée possède un très grand nombre de formes fonctionnellement équivalentes différentes (c'est-à-dire des séquences protéiques pouvant remplir la même fonction biochimique).
- (P4) Les gènes omniprésents spécifiques ne sont pas nécessaires dans une espèce donnée: Évidemment, il n'y a aucune raison a priori pour que chaque organisme ait la même séquence ou même des séquences similaires. Aucune séquence spécifique n'est fonctionnellement nécessaire dans un organisme - tout ce qui est nécessaire est l'une des nombreuses formes fonctionnellement équivalentes d'un gène ou d'une protéine omniprésente donné.
- (P5) L'hérédité corrèle les séquences, même en l'absence de nécessité fonctionnelle: il existe un, et un seul, mécanisme observé qui fait que deux organismes différents possèdent des protéines omniprésentes avec des séquences similaires (en dehors de l'extrême improbabilité du pur hasard, bien sûr) . Ce mécanisme est l'hérédité.
- (C) Ainsi, des gènes omniprésents similaires indiquent une relation généalogique: il s'ensuit que les organismes qui ont des séquences similaires pour les protéines ubiquitaires sont apparentés sur le plan généalogique. En gros, plus les séquences sont similaires, plus la relation généalogique est étroite.

Discussion:

Les séquences d'acides aminés des protéines sont souvent utilisées pour établir les relations phylogénétiques des espèces. Les études de séquence des gènes fonctionnels ont porté sur les gènes de protéines (ou d'ARN) omniprésents (c'est-à-dire que tous les organismes possèdent). Ceci est fait pour s'assurer que les comparaisons sont indépendantes du phénotype général de l'espèce.

Par exemple, supposons que nous comparions la séquence protéique d'un chimpanzé et celle d'un humain. Ces deux animaux ont de nombreux caractères et fonctions anatomiques similaires. Nous pouvons donc nous attendre à ce que leurs protéines soient également similaires, qu'elles soient apparentées ou non à la généalogie. Cependant, nous pouvons comparer les séquences de gènes très basiques utilisés par tous les organismes vivants, tels que le gène cytochrome c, qui n'ont aucune influence sur des caractéristiques spécifiques du chimpanzé ou de l'homme.

Le cytochrome c est une protéine essentielle et omniprésente présente dans tous les organismes, y compris les eucaryotes et les bactéries (Voet et Voet, 1995, p. 24). Les mitochondries des cellules contiennent le cytochrome c, où elles transportent des électrons dans le processus métabolique fondamental de la phosphorylation oxydative. L'oxygène que nous respirons est utilisé pour générer de l'énergie dans ce processus (Voet et Voet 1995, p. 577-582).

En utilisant un gène omniprésent tel que le cytochrome c, il n'y a aucune raison de supposer que deux organismes différents devraient avoir la même séquence protéique, voire des séquences protéiques similaires, à moins que les deux organismes ne soient apparentés sur le plan généalogique. Cela est dû en partie à la redondance fonctionnelle des séquences et des structures protéiques. Ici, la "redondance fonctionnelle" indique que de nombreuses séquences protéiques différentes forment la même structure générale et remplissent le même rôle biologique général. Le cytochrome c est une protéine extrêmement redondante sur le plan fonctionnel, car de nombreuses séquences dissemblables forment toutes des protéines de transport d'électrons du cytochrome c. La redondance fonctionnelle n'a pas besoin d'être exacte en termes de performances;

certaines séquences fonctionnelles du cytochrome c peuvent être légèrement meilleures au transport d'électrons que d'autres.

Des décennies de preuves biochimiques ont montré que de nombreuses mutations d'acides aminés, en particulier des résidus de surface, n'avaient que de faibles effets sur la fonction et la structure des protéines (Branden et Tooze 1999, Ch. 3; Harris et al. 1956; Lesk 2001, Chs. 5 et 6, pages 165-228; Li 1997, page 2; Matthews 1996). Un exemple frappant est celui des cytochromes de type c de diverses bactéries, qui n'ont pratiquement aucune similarité de séquence. Néanmoins, ils se replient tous dans une même structure tridimensionnelle et jouent tous le même rôle biologique (Moore et Pettigrew, 1990, p. 161-223; Ptitsyn, 1998).

Même au sein des espèces, la plupart des mutations d'acides aminés sont fonctionnellement silencieuses. Par exemple, il existe au moins 250 mutations différentes d'acides aminés connues dans l'hémoglobine humaine, portées par plus de 3% de la population mondiale, qui ne présentent aucune manifestation clinique chez les individus hétérozygotes ou homozygotes (Bunn et Forget, 1986; Voet et Voet, 1995, page 235). Le phénomène de redondance fonctionnelle des protéines est très général et est observé dans toutes les protéines et gènes connus.

Avec ceci à l'esprit, considérons à nouveau les séquences moléculaires du cytochrome c. Le cytochrome c est absolument essentiel à la vie - les organismes qui en sont dépourvus ne peuvent pas vivre. Il a été démontré que la protéine du cytochrome c humain marche chez une levure (un organisme unicellulaire) dont le gène du cytochrome c natif est supprimé, alors que le cytochrome c de la levure diffère du cytochrome c humain à plus de 40% (Tanaka et al. 1988a; Tanaka et al. 1988b; Wallace et Tanaka 1994). En fait, les gènes du cytochrome c du thon (poisson), du pigeon (oiseau), du cheval (mammifère), de la mouche *Drosophila* (insecte) et du rat (mammifère) fonctionnent tous chez la levure dépourvue du cytochrome c natif (Clements et al. 1989; Hickey et al 1991; Koshy et al 1992; Scarpulla et Nye 1986). En outre, une analyse génétique approfondie du cytochrome c a montré que la majeure partie de la séquence protéique est inutile pour sa fonction in vivo (Hampsey et al. 1986; Hampsey et al. 1988). Environ un tiers seulement des 100 acides aminés du cytochrome c sont nécessaires pour spécifier sa fonction. La plupart des acides aminés du cytochrome c sont hypervariables (c'est-à-dire qu'ils peuvent être remplacés par un grand nombre d'acides aminés fonctionnellement similaires) (Dickerson et Timkovich, 1975). Fait important, Hubert Yockey a effectué une étude minutieuse dans laquelle il calculait un minimum de 2,3 x 10^{93} séquences de protéines de cytochrome c fonctionnelles possibles, sur la base de ces analyses de mutation génétique (Hampsey et al. 1986; Hampsey et al. 1988; Yockey 1992 , Ch. 6, page 254). Pour mettre en perspective, le nombre 10^{93} est environ un milliard de fois plus grand que le nombre d'atomes dans l'univers visible. Ainsi, le nombre de séquences fonctionnelles du cytochrome c est pratiquement illimité, et il n'existe à priori aucune raison pour que deux espèces différentes possèdent les mêmes séquences de protéines du cytochrome c, même légèrement similaires.

En termes d'analyse statistique scientifique, "l'hypothèse nulle" est que l'identité des acides aminés non essentiels dans les protéines du cytochrome c de l'homme et du chimpanzé doivent être aléatoires les uns par rapport aux autres. Cependant, de la théorie de la descendance commune et de notre arbre phylogénétique standard, nous savons que les humains et les chimpanzés sont assez proches. Nous prédisons donc, malgré les probabilités, que les séquences du cytochrome c chez l'homme et le chimpanzé soient beaucoup plus similaires que, par exemple, le cytochrome c de l'humain et de la levure - simplement en raison de l'héritage.

Confirmation:

Les humains et les chimpanzés ont exactement la même séquence protéique du cytochrome c. "L'hypothèse nulle" donnée ci-dessus est fausse. En l'absence de descendance commune, le risque que cet événement se produise est inférieur à 10^{-93} (1 sur 10^{93}). Ainsi, le degré élevé de similitude de ces protéines corrobore de manière spectaculaire la théorie de la descendance commune. De plus, les protéines du cytochrome c humaines et du chimpanzé diffèrent d'environ 10 acides aminés de celles de tous les autres mammifères. Le risque que cela se produise en l'absence d'un mécanisme héréditaire est inférieur à 10^{-29}. La levure *Candida krusei* est l'un des organismes eucaryotes les moins proches de l'homme. Candida présente une différence de 51 acides aminés par rapport à la séquence humaine. Une estimation prudente de cette probabilité est inférieure à 10^{-25}.

Critiques

Une possible, mais peu probable, objection est que les légères différences de performances fonctionnelles entre les divers cytochromes pourraient être responsables de cette similarité de séquence. Cette objection est peu probable en raison du nombre incroyablement élevé de séquences presque équivalentes qu'il serait impossible de distinguer sur le plan phénotypique pour tout niveau de performance requis. De plus, des séquences presque similaires ne donnent pas nécessairement des niveaux de performance presque similaires.

Néanmoins, supposons, dans l'argumentation, qu'un cytochrome c qui transporte les électrons plus rapidement est nécessaire chez les organismes ayant un métabolisme actif ou présentant des taux élevés de contraction musculaire. Si cela était vrai, nous pourrions nous attendre à observer un modèle de similarité de séquence qui soit en corrélation avec la similarité d'environnement ou avec les exigences physiologiques. Cependant, ceci n'est pas observé. Par exemple, le cytochrome c de la chauve-souris ressemble beaucoup plus au cytochrome c humain qu'au cytochrome c du colibri; Le cytochrome c du marsouin ressemble beaucoup plus au cytochrome c humain qu'au cytochrome c de requin. Comme indiqué précédemment dans la prévision 1.3, l'arbre phylogénétique construit à partir des données du cytochrome c récapitule exactement les relations entre les principaux taxons, telles que déterminées par les données morphologiques totalement indépendantes (McLaughlin et Dayhoff 1973). Ces faits ne font que renforcer l'idée selon laquelle les séquences du cytochrome c sont indépendantes de la fonction phénotypique (autre que l'exigence évidente d'un cytochrome c fonctionnel qui transporte des électrons).

Résumé:

Le point de cette prédiction est légèrement différent de la prévision 1.3, “Convergence des phylogénies indépendantes”. Les preuves présentées ci-dessus démontrent que pour de nombreuses protéines fonctionnelles omniprésentes (telles que le cytochrome c), il existe un nombre énorme de séquences équivalentes qui pourraient former cette protéine dans un organisme donné. Chaque fois que nous trouvons que deux organismes ont des séquences identiques, ou très similaires, pour une protéine omniprésente, nous savons qu'il se passe quelque chose de louche. Pourquoi ces deux organismes auraient-ils des protéines omniprésentes aussi similaires si les chances sont astronomiques? Nous ne connaissons qu'une des raisons pour lesquelles deux organismes auraient deux séquences protéiques similaires en l'absence de nécessité fonctionnelle: l'hérédité. Ainsi, dans de tels cas, nous pouvons déduire avec confiance que les deux organismes sont liés de manière généalogique. En ce sens, la similarité des séquences n'est pas seulement un test de la théorie de la descendance commune; la descendance commune est également une *déduction* du principe de l'hérédité et l'observation de la similarité des séquences. Enfin, la similitude observée pour le cytochrome c ne se limite pas à cette protéine omniprésente; *toutes* les protéines omniprésentes qui ont été comparées entre les chimpanzés et les humains sont très similaires et de nombreuses comparaisons ont été effectuées.

Réfutation potentielle:

Sans supposer la théorie de la descendance commune, le résultat le plus probable est que les séquences de la protéine cytochrome c de tous ces organismes seraient très différentes les unes des autres. Si tel était le cas, une analyse phylogénétique serait impossible, ce qui fournirait des preuves très solides pour une origine des espèces sana rapport avec une généalogie, et peut-être simultanée (Dickerson, 1972; Yockey, 1992; Li, 1997).

En outre, le fondement même de cet argument pourrait être facilement compromis si l'on pouvait démontrer (1) que les protéines du cytochrome c spécifiques à une espèce étaient fonctionnelles exclusivement dans leurs organismes respectifs, ou (2) qu'aucune autre séquence du cytochrome c ne pourrait fonctionner dans un

organisme autre que son propre cytochrome natif, ou (3) qu'un mécanisme observé, au delà de l'hérédité, peut corréler de manière causale la séquence d'une protéine omniprésente avec une morphologie spécifique.

Prédiction 4.2: Redondance du codage de l'ADN

Comme la similarité de séquence protéique, la similarité de séquence ADN de deux gènes omniprésents implique également une ascendance commune. Bien entendu, des comparaisons complètes de séquences d'ADN de protéines conservées telles que le cytochrome c prennent également indirectement en compte les séquences d'acides aminés, car la séquence d'ADN spécifie la séquence de protéines. Cependant, avec les séquences d'ADN, il existe un niveau supplémentaire de redondance. Le code génétique lui-même est redondant sur le plan de l'information; en moyenne, il existe trois codons différents (un codon est un triplet de bases ADN) pouvant spécifier exactement le même acide aminé (Voet et Voet, 1995, p. 966). Ainsi, pour le cytochrome c, il existe environ 3^{104}, soit plus de 10^{146} séquences d'ADN différentes (et donc 10^{146} gènes possibles) pouvant spécifier exactement la même séquence protéique.

Ici, nous pouvons être assez spécifiques dans notre prédiction. Toute différence de séquence entre deux gènes fonctionnels du cytochrome c est nécessairement fonctionnellement neutre ou presque. Le taux de base des mutations chez l'homme (et chez la plupart des autres mammifères) a été mesuré à environ 1-5 x 10^{-8} substitutions de bases par site par génération (Mohrenweiser 1994, p. 128-129), et la génération moyenne des primates est d'environ 20 ans. D'après le registre fossile, nous savons que les humains et les chimpanzés ont divergé d'un ancêtre commun il y a moins de 10 millions d'années (une estimation prudente - probablement il y a moins de 6 millions d'années) (Stewart et Disotell, 1998). Ainsi, si les chimpanzés et les humains sont vraiment liés sur le plan généalogique, nous prédisons que la différence entre leurs séquences respectives d'ADN du gène du cytochrome c devrait être inférieure à 3% - probablement même beaucoup moins, en raison de la fonction essentielle du gène du cytochrome c.

Confirmation:

Comme mentionné ci-dessus, les protéines du cytochrome c chez les chimpanzés et les humains sont exactement identiques. La conclusion est que les deux séquences d'ADN codant pour le cytochrome c chez l'homme et les chimpanzés ne diffèrent que par quatre nucléotides (une différence de 1,2%), bien qu'il existe 10^{49} séquences différentes qui pourraient coder pour cette protéine.

Les effets combinés de la redondance codant l'ADN et de la redondance des séquences protéiques rendent les comparaisons de séquences d'ADN doublement redondantes; Les séquences d'ADN des protéines omniprésentes ne sont absolument pas corrélées avec les différences phénotypiques entre les espèces, mais elles sont fortement corrélées à l'hérédité. C'est pourquoi les phylogénies de séquences d'ADN sont considérées comme si robustes.

Réfutation potentielle:

Le résultat le plus probable est que les séquences d'ADN codant pour ces protéines devraient être radicalement différentes. Ce serait une réfutation retentissante de la macroévolution, et il constituerait une preuve très forte que les chimpanzés et les humains ne sont pas étroitement liés sur le plan généalogique. Bien entendu, les réfutations potentielles pour la prédiction 4.1 s'appliquent également aux séquences d'ADN.

Prédiction 4.3: Preuve moléculaire – Les transposons

À bien des égards, les transposons ressemblent beaucoup aux virus. Cependant, ils ne possèdent pas de gènes pour les protéines d'enveloppe virales, ils ne peuvent pas franchir les frontières cellulaires et ne se répliquent donc que dans le génome de leur hôte. Ils peuvent être considérés comme des parasites intragénomiques. Excepté dans les circonstances les plus rares, le seul mode de transmission d'un métazoaire à un autre consiste directement en duplication d'ADN et par héritage (par exemple, vos transposons sont donnés à vos enfants) (Li 1997, p. 338-345).

Se répliquer pour un transposon signifie se copier et insérer l'ADN copié au hasard quelque part dans le génome de l'hôte. La réplication du transposon (également appelée transposition) a été directement observée dans de nombreux organismes, y compris la levure, le maïs, les wallabies, les humains, les bactéries et les mouches. Ces mécanismes ont récemment été bien compris (Li 1997, p. 335-338; Futuyma 1998, pp. 639-641). On sait que des cas spécifiques de rétrotransposition ont causé la neurofibromatose et l'hémophilie chez l'homme (Kazazian et al. 1988; Wallace et al. 1991) et le cancer, entre autres maladies (Deininger et Batzer 1999).

Cette section sur les transposons et les deux sections suivantes qui traitent des pseudogènes et des rétrovirus endogènes sont toutes liées conceptuellement. Les séquences d’ADN dans les régions intergéniques (régions situées entre les gènes codant les protéines dans les génomes) comprennent de très nombreux transposons (comme LINE et SINE), des rétrovirus endogènes (comme HERV), des pseudogènes et d’autres séquences apparentées comme les microsatellites. De nombreux microsatellites sont étroitement associés et générés par des rétrotransposons tels que LINE et SINE (Arcot et al. 1995; Nadir et al. 1996; Wilder et Hollocher 2001; Yandava et al. 1997). Ces séquences intergéniques sont principalement responsables de modèles très spécifiques observés dans les analyses “d'empreintes génétiques”, telles que celles effectuées dans les tests de paternité ou les tests de fratrie. Comme les empreintes digitales, ces régions intergéniques varient considérablement entre les organismes individuels et les modèles sont en grande partie arbitraires. Par exemple, les éléments Alu, un type de rétrotransposon SINE, sont transposés dans un nouvel emplacement génomique environ toutes les 200 naissances humaines (Deininger et Batzer 1999) et les Alus contribuent à une fraction importante de la diversité génétique humaine (Batzer et Deininger 2002). Dans le cas du transposon humain L1, un des nombreux éléments humain LINE, une nouvelle rétrotransposition concerne environ 1 individu sur 20 (Scaringe et al. 2001; Ostertag et Kazazian 2001). Il s’agit d’une estimation prudente, étant donné que chacun de nous a environ 50 L1 compétents en rétrotransposition (Brouha et al. 2003). Les régions intergéniques du génome, comme tout ADN, sont héréditaires et il existe une très forte corrélation entre les parents. Lorsque deux individus partagent des modèles intergéniques spécifiques bien supérieurs à ceux attendus par hasard, il s'agit d'une preuve très forte d'une ascendance commune. C’est en fait la base scientifique sur laquelle reposent les empreintes génétiques.

Comme expliqué ci-dessus, la découverte du même transposon dans le même emplacement chromosomique dans deux organismes différents constitue une preuve directe forte d'une ascendance commune, puisqu'ils s'insèrent de manière assez aléatoire et qu'ils ne peuvent généralement pas être transmis autrement que par héritage. En outre, une fois qu'un ancêtre commun contenant une certaine transposition a été postulé, tous les descendants de cet ancêtre commun doivent également contenir la même transposition. Une exception possible est si cette transposition a été supprimée en raison d'un rare événement de suppression; Cependant, les délétions ne sont jamais nettes et la plupart du temps, une partie de la séquence du transposon est conservée. En utilisant les mêmes principes que les empreintes génétiques, les biologistes ont utilisé des transposons, des pseudogènes et des rétrovirus endogènes pour démontrer que de nombreuses espèces sont génétiquement apparentées, telles que l'homme et d'autres primates. Quelques exemples parmi d’autres sont donnés ci-dessous.

Confirmation:

Le rétroélément SINE est une classe courante de transposon (Li 1997, p. 349-352). Un important transposon SINE est l’élément Alu de 300 pb. Tous les mammifères contiennent de nombreux éléments Alu, y compris l'homme, où ils constituent 10% du génome humain (soit 60 millions de bases d'ADN répétitif) (Smit 1996; Li 1997, p. 354, 357). Des transpositions humaines très récentes de Alu ont été utilisées pour élucider les migrations humaines historiques et préhistoriques, car certains individus ont de nouvelles insertions Alu qui font

défaut à d'autres (Novick et al. 1993; Novick et al. 1995). En fait, il a été démontré que les transpositions Alu courantes constituaient des marqueurs fiables de descendance commune dans les affaires de paternité et en criminalistique (Novick et al. 1993; Novick et al. 1995; Roy-Engel et al. 2001). Plus important encore, dans le groupe humain α-globine, il y a sept éléments Alu, chacun étant partagé avec les chimpanzés dans les sept mêmes localisations (Sawada et al. 1985).

Plus spécifiquement, trois transpositions spécifiques de SINE différentes ont été trouvées dans les mêmes emplacements chromosomiques de cétacés (baleines), hippopotames et ruminants, qui sont tous étroitement apparentés selon l'arbre phylogénétique standard. Cependant, tous les autres mammifères, y compris les chameaux et les porcs, sont dépourvus de ces trois transpositions spécifiques (Shimamura, 1997).

Vous trouverez plus de détails et d'explications à ce sujet dans la FAQ plagiats d'erreurs et génétique moléculaire d'Edward Max.

Réfutation potentielle:

Voir les deux ci-dessous, car les mêmes principes s'appliquent ici.

Prédiction 4.4: Preuve moléculaire - Pseudogènes redondants

D'autres exemples moléculaires qui fournissent des preuves d'une ascendance commune sont les curieuses séquences d'ADN connues sous le nom de pseudogènes. Les pseudogènes sont très étroitement liés aux gènes fonctionnels codant pour les protéines. La similarité implique à la fois la séquence d'ADN primaire et souvent la localisation chromosomique spécifique des gènes. Les homologues fonctionnels des pseudogènes sont des gènes normaux qui sont transcrits en ARNm, qui est à son tour traduit en protéine fonctionnelle. En revanche, les pseudogènes ont des séquences régulatrices défectueuses qui empêchent la transcription du gène en ARNm ou des codons d'arrêt internes qui empêchent la fabrication de la protéine fonctionnelle. En ce sens, les pseudogènes sont des exemples moléculaires de structures vestigiales.

Cependant, les pseudogènes sont inclus ici dans une prédiction séparée car de nombreux pseudogènes sont inhabituels d'une manière supplémentaire. Les vestiges morphologiques ont perdu leur fonction d'origine et l'organisme portant le vestige a également perdu cette fonction. En revanche, les pseudogènes ont perdu leur fonction d'origine, mais l'organisme lui-même peut conserver cette fonction s'il porte la contrepartie fonctionnelle de ces pseudogènes. Les pseudogènes résiduels au sens morphologique, comme le pseudogène de synthèse de la vitamine C, sont pris en compte dans la prédiction 2.3. Le type restant de pseudogène, dans lequel un organisme porte à la fois un gène fonctionnel et un ou plusieurs pseudogènes correspondants, est appelé ci-après "pseudogène redondant".

La plupart des pseudogènes sont en grande partie non fonctionnels. Plusieurs éléments de preuve étayent cette conclusion. Premièrement, la présence ou l'absence des pseudogènes les plus spécifiques n'a pas d'effet mesurable sur le phénotype de l'organisme. Deuxièmement, il existe de bons arguments mécanistes et génétiques indiquant que les pseudogènes n'ont pratiquement aucune fonction. Les pseudogènes ont des séquences complexes très similaires ou identiques à celles nécessaires au bon fonctionnement d'autres protéines enzymatiques ou structurelles. Ces gènes normaux sont activement transcrits et traduits en protéines, alors que les pseudogènes ne sont ni traduits, ni transcrits, ni les deux. Ainsi, les pseudogènes ne peuvent pas remplir les fonctions des protéines qu'ils codent. Si les pseudogènes ont une fonction, ils doivent exécuter des fonctions relativement simples pour lesquelles la protéine codée par eux n'a pas été conçue.

Troisièmement, si un pseudogène a peu ou pas de fonction, alors la plupart des mutations du pseudogène n'auront que des conséquences fonctionnelles mineures et de nombreuses mutations ne seront pas éliminées par la purification de la sélection. Par conséquent, nous nous attendons à ce que les pseudogènes réellement non fonctionnels accumulent les mutations au taux de base de la mutation. Les pseudogènes à fonctions mineures accumuleront des mutations proches de la vitesse de fond. Comme prévu si les pseudogènes n'ont

pratiquement aucune fonction, la plupart des pseudogènes accumulent les mutations à la vitesse la plus rapide connue pour n'importe quelle région de l'ADN du génome animal. De plus, le taux de mutation déduit pour les pseudogènes à partir d'une analyse phylogénétique correspond très étroitement aux taux mesurés de mutations spontanées. Pour plus d'informations et de références, voir Prédiction 5.8.

Enfin, nous comprenons comment les pseudogènes redondants sont créés et nous avons observé la création de nouveaux pseudogènes redondants en laboratoire et dans la nature. Les pseudogènes redondants sont générés par duplication de gènes et mutations ultérieures. De nombreux processus observés sont connus pour dupliquer des gènes, notamment des événements de transposition, une duplication chromosomique et un crossing-over inégal de chromosomes.

Ces faits corroborent fortement la conclusion selon laquelle la plupart des pseudogènes n'ont peu, voire aucune fonction. Comme les transpositions (voir prévision 4.3), la création de nouveaux pseudogènes redondants par duplication de gènes est un événement rare et aléatoire et, bien entendu, tout ADN dupliqué est hérité. Ainsi, la découverte du même pseudogène dans le même emplacement chromosomique chez deux espèces constitue une preuve solide d'ascendance commune.

Confirmation:

Il existe de très nombreux exemples de pseudogènes redondants partagés entre les primates et les humains. L'un est le gène ψη-globine, un pseudogène de l'hémoglobine. Il n'est partagé que par les primates, à l'emplacement chromosomique exact, avec les mêmes mutations qui détruisent sa fonction en tant que gène codant pour une protéine (Goodman et al. 1989). Un autre exemple est le gène de stéroïde 21-hydroxylase. Les humains possèdent deux copies du gène de stéroïde 21-hydroxylase, une fonctionnelle et un pseudogène non traduit. L'inactivation du gène fonctionnel conduit à une hyperplasie congénitale des surrénales (CAH, une maladie génétique rare et grave), ce qui montre clairement que le pseudogène de la 21-hydroxylase n'a pas la fonction qui lui revient. Les chimpanzés et les humains partagent la même suppression de huit paires de bases dans ce pseudogène qui le rend incapable de sa fonction normale (Kawaguchi et al. 1992).

Réfutation potentielle:

Comme expliqué ci-dessus, les duplications de gènes observées sont des événements rares et aléatoires. Il est donc très peu probable que d'autres mammifères aient ces mêmes pseudogènes redondants aux mêmes emplacements chromosomiques, avec les mêmes mutations qui paralysent leurs fonctions normales. Par exemple, il est essentiellement impossible pour les souris de porter les pseudogènes de 21-hydroxylase, dans le même emplacement génomique, avec la même suppression de huit paires de bases qui détruit sa fonction enzymatique.

De plus, une fois qu'un gène est dupliqué et que des mutations en font un pseudogène redondant, il est hérité par tous les descendants. Ainsi, une fois que certains organismes portant le même pseudogène ont été trouvés, une descendance commune exige que tous les organismes phylogénétiquement intermédiaires portent également ce pseudogène. Par exemple, supposons que nous trouvions que les humains et les cercopithécidés (singes de l'ancien monde) partagent un certain pseudogène redondant. Selon la descendance commune, tous les grands singes (y compris les chimpanzés, les gorilles, les orangs-outans et les siamangs) doivent également nécessairement porter le même pseudogène redondant dans le même emplacement chromosomique. Cette conclusion repose sur le prémisse qu'il n'existe aucun mécanisme permettant de supprimer les pseudogènes des génomes (ou que les mécanismes sont très inefficaces). Cela est apparemment vrai pour les vertébrés, mais on sait que certains organismes dont le temps de génération est court, tels que les bactéries, les protistes et *Drosophila*, possèdent des mécanismes qui éliminent les excès d'ADN.

Notez que cette confirmation et cette réfutation potentielle sont indépendantes de la question de savoir si un pseudogène spécifique a une fonction ou s'il est complètement non fonctionnel, pour les mêmes raisons que celles expliquées dans la prédiction sur les vestiges morphologiques. Comme tout élément génétique ou

structure organismique, l'opportunisme évolutif peut s'appuyer sur un pseudogène et l'utiliser dans une nouvelle et différente fonction.

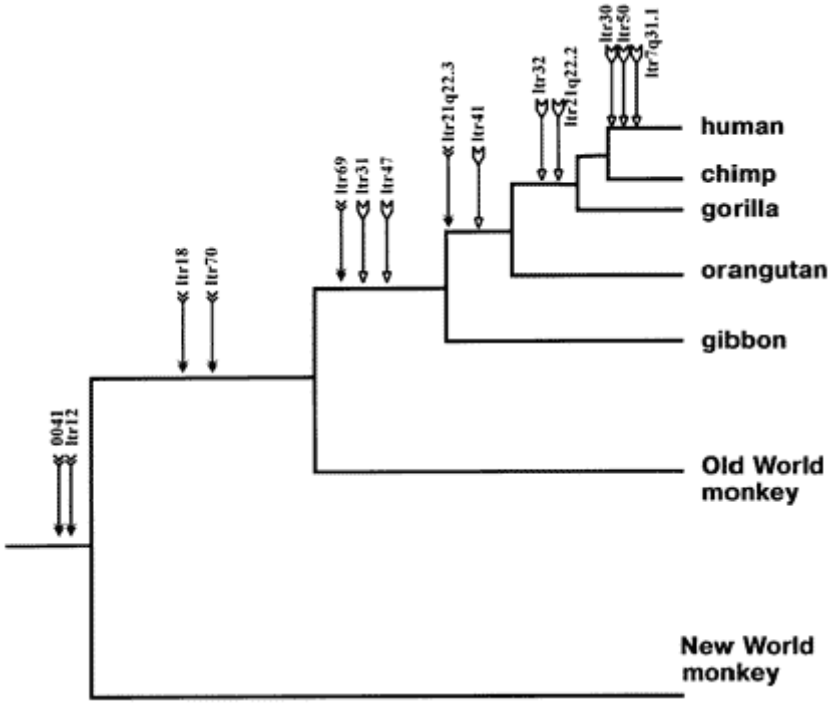
Prédiction 4.4: Preuve moléculaire - Retrovirus endogènes

Les rétrovirus endogènes fournissent un autre exemple de preuve de séquence moléculaire pour une descendance commune universelle. Les rétrovirus endogènes sont les restes moléculaires d'une infection virale parasitaire passée. Parfois, des copies d'un génome de retrovirus sont trouvées dans le génome de son hôte, et ces copies de gènes de rétrovirus sont appelées séquences de rétrovirus endogènes. Les rétrovirus (comme le virus du sida ou HTLV1, qui provoque une forme de leucémie) créent une copie ADN de leur propre génome viral et l'insèrent dans le génome de leur hôte. Si cela se produit dans une cellule de lignée germinale (c'est-à-dire les spermatozoïdes ou les ovocytes), l'ADN rétroviral sera hérité par les descendants de l'hôte. Encore une fois, ce processus est rare et assez aléatoire. Par conséquent, la recherche de rétrogènes dans des positions chromosomiques identiques de deux espèces différentes indique une ascendance commune.

Confirmation:

Chez l'homme, les rétrovirus endogènes occupent environ 1% du génome, soit environ 30 000 rétrovirus différents inclus dans l'ADN génomique de chaque personne (Sverdlov 2000). Il existe au moins sept exemples connus d'insertions rétrogènes communes entre chimpanzés et humains, et ce nombre augmentera certainement à mesure que les génomes de ces organismes sont séquencés (Bonner et al. 1982; Dangel et al. 1995; Svensson et al. 1995; Kjellman et al., 1999; Lebedev et al, 2000; Sverdlov, 2000). La figure 4.4.1 montre un arbre phylogénétique de plusieurs primates, y compris l'homme, issu d'une étude récente ayant identifié de nombreux rétrovirus endogènes partagés dans le génome de ces primates (Lebedev et al., 2000). Les flèches désignent les temps relatifs d'insertion de l'ADN viral dans le génome de l'hôte. Toutes les branches après le point d'insertion (à droite) portent cet ADN rétroviral - ce qui reflète le fait qu'une fois qu'un rétrovirus a été inséré dans l'ADN de la lignée germinale d'un organisme donné, il sera hérité par tous les descendants de cet organisme.

Figure 4.4.1. Insertion de rétrovirus endogène humain K (HERV-K) dans des localisations chromosomiques identiques chez différents primates.



(reproduit de Lebedev et al. 2000, © 2000, avec la permission d'Elsevier Science).

Les félidés (c'est-à-dire les chats) constituent un autre exemple. L'arbre phylogénétique standard montre les petits chats divergeant plus tard que les grands chats. Les petits chats (par exemple, le Chaus, le chat sauvage européen, le chat sauvage africain, le chat à pieds noirs et le chat domestique) partagent une insertion spécifique du gène rétroviral. En revanche, tous les autres carnivores qui ont été testés sont dépourvus de ce rétrogène (Futuyma 1998, pp. 293-294; Todaro et al. 1975).

Réfutation potentielle:

De manière macro-évolutive, cela n'aurait aucun sens si certains autres mammifères (chiens, vaches, ornithorynques, etc.) avaient ces mêmes rétrogènes dans les mêmes emplacements chromosomiques. Par exemple, il serait extrêmement improbable que les chiens portent également les trois insertions de HERV-K uniques à l'homme, comme le montre le coin supérieur droit de la figure 4.4.1, car aucun des autres primates ne possède ces séquences rétrovirales.

Partie 3. Opportunisme évolutionniste
Partie 5. Changement

Références

- Arcot, S.S., Wang, Z., Weber, J.L., Deininger, P.L., and Batzer, M.A. (1995) "Alu repeats: a source for the genesis of primate microsatellites." Genomics. 29: 136-144. PubMed
- Batzer, M. A., and Deininger, P. L. (2002) "Alu repeats and human genomic diversity." Nat Rev Genet. 3: 370. PubMed
- Bonner, T. I., C. O'Connell, et al. (1982) "Cloned endogenous retroviral sequences from human DNA." PNAS 79: 4709. PubMed
- Branden, C. and Tooze, J. (1999) Introduction to Protein Structure. Second Ed., New York, Garland Publishing. Publisher's Site
- Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., Kazazian, H.H. (2003) "Hot

L1s account for the bulk of retrotransposition in the human population." Proc Natl Acad Sci U S A. 100: 5280-5285. PNAS free full text
- Bunn, H. F., and Forget, E. G. (1986) Hemoglobin: Molecular, Genetic, and Clinical Aspects. Saunders.
- Clements, J. M., O'Connell, L. I., Tsunasawa, S., and Sherman, F. (1989) "Expression and activity of a gene encoding rat cytochrome c in the yeast Saccharomyces cerevisiae." Gene 83: 1-14. PubMed
- Dangel, A. W., B. J. Baker, et al. (1995) "Complement component C4 gene intron 9 as a phylogenetic marker for primates: long terminal repeats of the endogenous retrovirus ERV-K(C4) are a molecular clock of evolution." Immunogenetics 42: 41-52. PubMed
- Deininger, P. L., and Batzer, M. A. (1999) "Alu repeats and human disease." Mol Genet Metab. 67: 183-193. PubMed
- Dickerson, R. E. (1972) Scientific American. 226: 58-72.
- Dickerson, R. E., and Timkovich, R. (1975) cytochrome c. The Enzymes. P. D. Boyer. New York, Academic Press. 11: 397-547.
- Futuyma, D. (1998) Evolutionary Biology. Third edition. Sunderland, MA, Sinauer Associates.
- Goodman, M., B. F. Koop, et al. (1989) "Molecular phylogeny of the family of apes and humans." Genome 31(316-335). PubMed
- Hampsey, D. M., Das, G., and Sherman F. (1986) "Amino acid replacements in yeast iso-1-cytochrome c." Journal of Biological Chemistry 261: 3259-71. PubMed
- Hampsey, D. M., Das, G., and Sherman F. (1988) "Yeast iso-1-cytochrome c: genetic analysis of structural requirements." FEBS Letters 231: 275-83. PubMed
- Harris, J. I., F. Sanger, et al. (1956) "Species differences in insulin." Archives of Biochemistry and Biophysics 65: 427-438.
- Hickey, D. R., Jayaraman, K., Goodhue, C. T., Shah,J., Fingar, S. A., Clements, J. M., Hosokawa, Y., Tsunasawa, S., and Sherman, F. (1991) "Synthesis and expression of genes encoding tuna, pigeon, and horse cytochromes c in the yeast Saccharomyces cerevisiae." Gene 105: 73-81. PubMed
- Kawaguchi, H., C. O'hUigin, et al. (1992) "Evolutionary origin of mutations in the primate cytochrome P450c21 gene." American Journal of Human Genetics 50: 766-780. PubMed
- Kazazian, H. H. (1999) "An estimated frequency of endogenous insertional mutations in humans." Nat Genet. 22: 130. PubMed
- Kazazian, H. H., C. Wong, et al. (1988) "Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man." Nature 332: 164. PubMed
- Kjellman, C., H. O. Sjogren, et al. (1999) "HERV-F, a new group of human endogenous retrovirus sequences." Journal of General Virology 80: 2383. http://vir.sgmjournals.org/cgi/content/full/80/9/2383
- Koshy, T. I., Luntz, T. L., Garber, E. A., and Margoliash, E. (1992) "Expression of recombinant cytochromes c from various species in Saccharomyces cerevisiae: post-translational modifications." Protein Expr. Purif. 3: 441-52. PubMed
- Lebedev, Y. B., Belonovitch, O. S., Zybrova, N. V, Khil, P. P., Kurdyukov, S. G., Vinogradova, T. V., Hunsmann, G., and Sverdlov, E. D. (2000) "Differences in HERV-K LTR insertions in orthologous loci of humans and great apes." Gene 247: 265-277. PubMed
- Lesk, Arthur M. (2001) Introduction to Protein Architecture. Oxford, Oxford University Press. Publisher's Site
- Li, W.-H. (1997) Molecular Evolution. Sunderland, MA, Sinauer Associates.
- Matthews, B. W. (1996) "Structural and genetic analysis of the folding and function of T4 lysozyme." FASEB J. 10: 35-41. PubMed
- McLaughlin, P. J., and Dayhoff, M. O. (1973) "Eukaryote evolution: a view based on cytochrome c sequence data." Journal of Molecular Evolution 2(99-116).
- Mohrenweiser, H. (1994) "Impact of the molecular spectrum of mutational lesions on estimates of germinal gene-mutation rates." Mutation Research 304: 119-137. PubMed
- Moore, G. R., and Pettigrew, G. W. (1990) Cytochromes c: Evolutionary, Structural and Physicochemical Aspects. Berlin, Springer-Verlag.
- Nadir, E., Margalit, H., Gallily, T., and Ben-Sasson, S. A. (1996) "Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications." Proc Natl Acad Sci U S A. 93: 6470-6475. PubMed
- Novick, G. E., Gonzalez, T., Garrison, J., Novick, C. C., Batzer, M. A., Deininger, P. L., Herrera, R. J. (1993) "The use of polymorphic Alu insertions in human DNA fingerprinting." EXS. 67: 283-291. PubMed
- Novick, G. E., Novick, C. C., Yunis, J., Yunis, E., Martinez, K., Duncan, G. G., Troup, G. M., Deininger, P. L.,

Stoneking, M., Batzer, M. A., et al. (1995) “Polymorphic human specific Alu insertions as markers for human identification.” Electrophoresis 16: 1596-1601. PubMed
- Ostertag, E. M., and Kazazian, H. H. (2001) “Biology of mammalian L1 retrotransposons.” Annu Rev Genet. 35: 501-538. PubMed
- Ptitsyn, O. B. (1998) “Protein folding and protein evolution: common folding nucleus in different subfamilies of c-type cytochromes.” Journal of Molecular Biology 278: 655. PubMed
- Roy-Engel, A. M., Carroll, M. L., Vogel, E., Garber, R. K., Nguyen, S. V., Salem, A. H., Batzer, M. A., and Deininger, P. L. (2001) “Alu insertion polymorphisms for the study of human genomic diversity.” Genetics 159: 279-290. http://www.genetics.org/cgi/content/full/159/1/279
- Sawada, I., C. Willard, et al. (1985) “Evolution of Alu family repeats since the divergence of human and chimpanzee.” Journal of Molecular Evolution 22(316). PubMed
- Li, X., Scaringe, W. A., Hill, K. A., Roberts, S., Mengos, A., Careri, D., Pinto, M. T., Kasper, C. K., and Sommer, S. S. (2001) “Frequency of recent retrotransposition events in the human factor IX gene.” Hum Mutat. 17: 511-519. PubMed
- Scarpulla, R. C., and Nye, S. H. (1986) “Functional expression of rat cytochrome c in Saccharomyces cerevisiae.” Proc Natl Acad Sci 83: 6352-6. PubMed
- Shimamura, M., et al. (1997) “Molecular evidence from retroposons that whales form a clade within even-toed ungulates.” Nature 388: 666. PubMed
- Smit, A. F. A. (1996) “The origin of interspersed repeats in the human genome.” Current Opinion in Genetics and Development 6: 743-748. PubMed
- Stewart, C. B. and Disotell, T. R. (1998) “Primate evolution - in and out of Africa.” Current Biology 8: R582-588. PubMed
- Svensson, A. C., N. Setterblad, et al. (1995) “Primate DRB genes from the DR3 and DR8 haplotypes contain ERV9 LTR elements at identical positions.” Immunogenetics 41: 74. PubMed
- Sverdlov, E. D. (2000) “Retroviruses and primate evolution.” BioEssays 22: 161-171. PubMed
- Tanaka, Y., Ashikari, T., Shibano, Y., Amachi, T., Yoshizumi, H., and Matsubara, H. (1988a) “Amino acid replacement studies of human cytochrome c by a complementation system using CYC1 deficient yeast.” J Biochem (Tokyo) Sep;104: 477-80. PubMed
- Tanaka, Y., Ashikari, T., Shibano, Y., Amachi, T., Yoshizumi, H., and Matsubara, H. (1988b) “Construction of a human cytochrome c gene and its functional expression in Saccharomyces cerevisiae.” J Biochem (Tokyo) 103: 954-61. PubMed
- Todaro, G.J., Benveniste, R.E., Callahan, R., Lieber, M.M., and Sherr, C.J. (1975) “Endogenous primate and feline type C viruses.” Cold Spring Harb Symp Quant Biol. 39 Pt 2:1159-1168. PubMed
- Voet, D., and Voet, J. (1995) Biochemistry. New York, John Wiley and Sons.
- Wallace, C. J., and Tanaka, Y. (1994) “Improving cytochrome c function by protein engineering?: studies of site-directed mutants of the human protein.” J. Biochem. (Tokyo) 115: 693-700. PubMed
- Wallace, M. R., L. B. Andersen, et al. (1991) “A *de novo* Alu insertion results in neurofibromatosis type 1.” Nature 353: 864-866. PubMed
- Wilder, J., and Hollocher, H. (2001) “Mobile elements and the genesis of microsatellites in dipterans.” Mol Biol Evol. 18: 384-392. PubMed
- Yandava, C. N., Gastier, J. M., Pulido, J. C., Brody, T., Sheffield, V., Murray, J., Buetow, K., and Duyk, G. M. (1997) “Characterization of Alu repeats that are associated with trinucleotide and tetranucleotide repeat microsatellites.” Genome Res. 7: 716-724.
- Yockey, H. P. (1992) Information Theory and Molecular Biology. New York, Cambridge University Press.

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

Permanent link:
https://www.evowiki.fr/comdesc-4?rev=1575654798

Last update: **2019/12/06 18:53**