UNIVERSITÉ DU QUÉBEC

IMPACT DES CONTRAINTES STRUCTURALES DES POLYPEPTIDES MITOCHONDRIAUX SUR L'ÉVOLUTION DU GÉNOME MITOCHONDRIAL CHEZ LES VERTÉBRÉS

MÉMOIRE PRÉSENTÉ À
L'UNIVERSITÉ DU QUÉBEC À RIMOUSKI
Comme exigence partielle
Du programme de la
Maîtrise en gestion de la faune et ses habitats

Par
ALEXANDRA DUBÉ LOUBERT

Mai 2006

Remerciements

Pierre!
Plus il y a de désordre dans un système (prenons ton bureau comme système de référence), plus il existe de macro-états (piles de livre, articles, idées naissantes ou en maturation) qui correspondent à un seul macro-état (Lavenda, 1997). Ce macro-état de forte entropie dépend du désordre environnant pour nourrir cet ordre unique qu'on appelle idée, découverte. Je voulais par conséquent te remercier d'entretenir ce désordre fécond pour maintenir ce flot d'idées structurées et motivantes. J'espère pouvoir continuer d'être nourrie par ton entropie allumée et inspirée, et peut-être aussi y contribuer encore un peu.

Bernard! France!
Un grand merci d'avoir accepté promptement de faire partie du comité de correction de mon mémoire. Bien que vous ne sachiez pas encore dans qu'elle aventure vous vous êtes embarqués, j'espère qu'elle sera positive et agréable. Merci!

Richard!
Un grand merci pour avoir contribué à ta façon à m'aider à apercevoir dans ce désordre du monde vivant « l'ordre souverain de la nature » (Linné).

Richard! Alain!
Les statistiques auraient prédit cette forte probabilité de mon besoin de consultations répétées en statistique. Merci d'avoir respecté et accepté ce qui était, ma foi, dans le domaine du très probable.

Julien!
Ton amour, ta présence… et ton anglais inconditionnels m'ont porté de par les hauts et les bas de ce projet inspirant. Et s'il me semble aujourd'hui, en fin de parcours, que ce projet fut non seulement inspirant mais inspiré, c'est certainement un peu grâce à toi. Merci!

A STUDENT FOREWORD

QUESTIONNING THE SCIENTIFIC PROCESS

If we recognize that the way we talk about science's objects, far from being solely determined by empirical proof, influences our research enough to determine the type of questions we ask and the type of proof we are susceptible to find (Fox Keller, 1999), we should consider other factors to identify the source of the scientific knowledge we develop and to better understand its implications. If we want a science open to the social world, it is not to depreciate its importance, rather, it is to assert its significance and its reliability. In fact, sciences that confront the diversity of external factors are not only more pertinent, but are also becoming more productive in terms quantity and quality of knowledge (Nowotny, Scott *et al.*, 2001). Therefore, because of preestablished preferences, deeply rooted in our political and social reality, guiding us in our research, our responsibility is to question paradigms that inspire us in our scientific work. (in Dubé-Loubert & Leclerc 2005, to be published soon!)

FROM GENOME STRUCTURE TO PHENOTYPE

In our analysis, we assume that evolution is a genetically-based process where changes in population and species characteristics over time are fundamentally concepts of rate (Arnold, Pfender *et al.*, 2001; Kinnison & Hendry, 2001), where little scale variations would be a reservoir of potential functions (Finta & Zaphiropoulos, 2001). We assume that

correlations between genetic variability and function variation, and between genetic variability and environmental variability, correspond to a causality link between genes and observed functions. However, we should not be to prompt to presume of the causality of genomic changes on the evolution of phenotypes.

It is already difficult to link illnesses to underlying gene mutations (Kunz, 2003), it is consequently much more difficult to give sense to the nature of biological traits on a strictly mechanical view by definable DNA differences. There are relationships, cross factors, synergies, functional levels, each with their own properties and their own environmental sensitivities. Recent developments in genetic and epigenetic (de la Casa-Esperon & Sapienza, 2003; McNairn & Gilbert, 2003; Audit, Vaillant *et al.*, 2004; Chong & Whitelaw, 2004; Hennig, 2004; Rudden, Xiao *et al.*, 2005) allow us to anticipate the amazing complexity of the genome and cellular machinery, with both its determisms and random events. This suggest that different gene types controlled by different processes do not evolve in the same way and thus, do not lead to the same evolutionary modifications. Therefore, this complex system may not solely lie on genetic differential rates of change and variation.

Evolution would depend on the united action of internal and external forces that serve to change individuals particularly during ontogenic development (Balon, 2005, Unpublished manuscript). Species as integrated whole, with *Baupläne*, are so constrained by their phyletic heritage, pathways of development, and general architecture that the

constraints themselves become more interesting and more important in delimiting pathways of change than the selective forces that may mediate change when it occurs (Gould & Lewontin, 1979). Even if morphological and physiological traits present their own characteristics with possibly their own patterns designed by different evolutionary strategies (Hendry & Kinnison, 2001; Kinnison & Hendry, 2001; Blomberg, Garland *et al.*, 2003), using constraints as borders defining evolutionary ways may yield a good conceptual approach to evolutionary questions. We have to remember that species do what they can with what they are, without any need to be optimally designed for a particular environment. Evolution is a florishing scientific field with its controversies and consensuses, with its new theories and new synthesises. Thus, as claimed by Gould & Lewontin (1979), we should support Darwin's own pluralistic approach to identify the agents of evolutionary change.

Résumé

Les banques publiques de gènes permettent aux scientifiques d'exploiter une très grande diversité d'information génétique. Grâce aux savoirs développés sur la biologie des espèces de même que sur la machinerie cellulaire et ses caractéristiques physico-chimiques, ces banques de gènes s'inscrivent dans une démarche holistique – la biologie intégrative – laquelle est nécessaire à une meilleure compréhension des interactions entre les molécules du vivant. Souvent utilisée comme première étape dans le processus scientifique, la bioinformatique permet de mettre en lumière différents patrons comme l'organisation de base de diverses structures cellulaires, et de diriger la recherche vers les processus ayant mis en place ces patrons. Elle est un outil puissant pour questionner les hypothèses sur le déterminisme génétique et les paradigmes en évolution, pour dépasser le réductionnisme génétique et mieux comprendre l'évolution à grande échelle.

La présente étude porte sur l'évolution des treize gènes mitochondriaux codant pour des protéines chez 164 espèces de vertébrés. Les effets des changements physico-chimiques de ces protéines ont permis d'élaborer un modèle évolutif où la tolérance à la variabilité physico-chimique prédit le taux d'évolution des protéines mitochondriales. Le modèle prédit leur capacité à supporter des changements physico-chimiques sans affecter leurs propriétés fonctionnelles, suggérant que plus un gène tolère de grands changements physico-chimiques, plus il est susceptible d'évoluer rapidement. Étant donné, qu'il n'y a pas d'évidence claire que la longueur du gène ou sa position par rapport à l'origine de réplication sont liées au taux d'évolution, la tolérance aux changements physico-chimiques serait le meilleur indice connu pour expliquer les différences de taux d'évolution entre les gènes codées par l'ADNmt. De plus, le modèle rend compte de la singularité des espèces et des gènes, et de leur façon de se détacher de la prédiction générale qui, dans ce contexte, décrit ce qui est considéré comme *normal*. Par exemple, la compilation de l'ensemble des valeurs en dehors de la prédiction du modèle montre que la sélection négative est clairement dominante dans l'ADNmt. De plus, il a permis d'identifier plusieurs patrons évolutifs, notamment l'augmentation de la variabilité physico-chimique des protéines mitochondriales chez les oiseaux, et l'augmentation générale de la sensibilité du complexe NADH déshydrogénase aux variations physico-chimiques. Enfin, l'étude met en lumière d'autres patrons évolutifs émergeant de la structure du génome mitochondrial pouvant constituer des avenues de recherche pour le futur.

Si cette étude peut susciter des questions concernant notre façon de théoriser l'évolution du génome mitochondrial, mais aussi plus largement l'évolution biologique, si elle contribue à rassembler sous une loi des éléments apparemment épars, elle aura accompli ce par quoi un scientifique peut se réaliser; s'étonner, douter, questionner, construire ou reconstruire. Si elle m'a permis de mieux envisager voire comprendre les processus évolutifs responsables des patrons observés, elle m'a surtout permis de découvrir des voies à explorer, de nouveaux facteurs à intégrer à l'étude de l'évolution biologique. Elle est un outil de questionnement, un pavé dans la chaîne ou le réseau des causalités.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

CHAPITRE 1

MITOCHONDRIES ET ÉVOLUTION

1.1 LES MITOCHONDRIES ET L'ADNmt CHEZ LES VERTÉBRÉS

La mitochondrie : de bactéries libres à leur entrée dans les cellules eucaryotes, une relation endosymbiotique s'établit. Ces procaryotes, dits hétérotrophes eubactériens, qui élurent domicile dans des cellules primitives, les archae métanogènes, révolutionnèrent en quelque 2,5 milliards d'années, le métabolisme énergétique en rendant les cellules aptes à consommer l'oxygène (Lang, Paquin *et al.*, 2000; Selosse, 2003; Rand, Haney *et al.*, 2004). Cet organite, commun à tous les organismes eucaryotes, aurait en effet grandement contribué à la complexification des êtres vivants, bref, à leur évolution.

La complexification des organismes, permise notamment par une disponibilité accrue en énergie, s'est illustrée par une diversification sans précédent des tissus, des systèmes de régulation et d'intégration des processus physiologiques, de maintien de l'homéostasie et de thermorégulation. Présentant déjà une longue histoire évolutive leur ayant permis de couvrir l'ensemble des régions biogéographiques, les grands groupes de vertébrés ont

connu ces transformations profondes d'où est né un vaste ensemble de caractères physiologiques clés, comprenant la respiration pulmonaire et branchiale, l'endothermie, etc. Ces caractères reposent sur le métabolisme énergétique, supporté lui-même en grande partie par la respiration mitochondriale (Nicholls & Ferguson, 2001).

À l'intérieur de la mitochondrie, dans la matrice, se trouve un brin d'ADN circulaire, souvenir d'une identité jadis pleinement bactérienne. De même, la matrice abrite le complexe pyruvate déshydrogénase, les enzymes du cycle de l'acide citrique et la plupart des enzymes qui catalysent l'oxydation des acides gras. Vers l'extérieur, la matrice est bordée par la membrane mitochondriale interne. Celle-ci capte les métabolites dans la matrice et leur fait perdre leurs électrons en obligeant ces derniers à passer par les quatre complexes protéiques de la chaîne de transport des électrons (ETS : *electron transport system*) pour fabriquer l'ATP, énergie chimique universelle, à l'aide du complexe protéique ATP synthase, lequel est responsable de la phosphorylation oxydative (OXPHOS) (Nicholls & Ferguson, 2001) (figure 1).

D'une taille variant entre 16 et 18 kb chez les vertébrés, l'ADNmt contient 13 gènes codant des protéines d'ETS et d'OXPHOS (complexes I, III, IV et V), 22 ARN de transfert (contrairement à 30-32 dans le cytosol), 2 ARN ribosomaux et la D-Loop, zone du génome hautement variable contenant l'origine de réplication de l'ADNmt et les principaux promoteurs de la transcription (Scheffler, 1999; Taanman, 1999; Rand, 2001). La réplication, la transcription et la traduction de l'ADN mitochondrial (ADNmt) semblent

exclusivement servir le maintien et l'assemblage d'ETS et de l'ATP synthase. Bien que vitaux, ces mécanismes ne constituent pas les seules fonctions de la mitochondrie (Scheffler, 1999). En effet, en plus d'être au cœur du métabolisme aérobie, la mitochondrie est notamment impliquée dans le cycle de l'urée, dans la biosynthèse des lipides de même que dans l'apoptose (Scheffler, 1999).

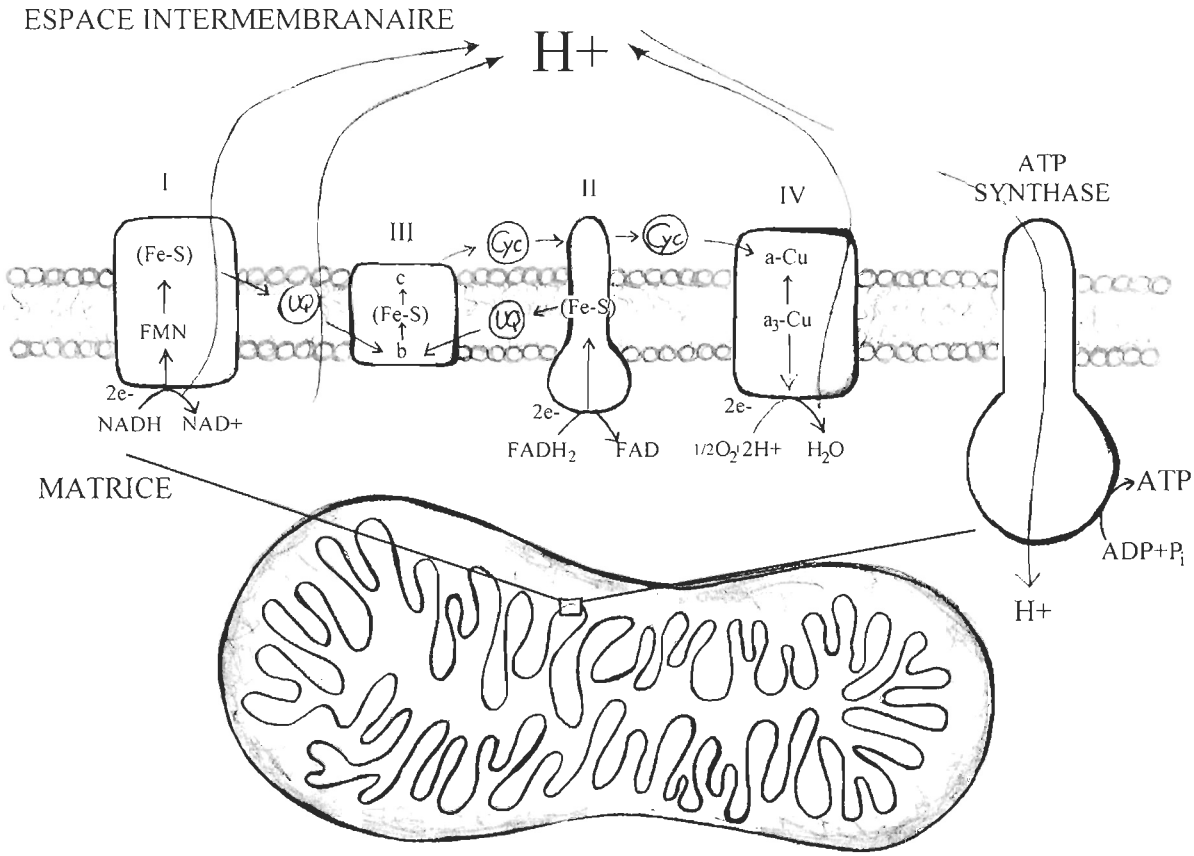


Figure 1. Mitochondrie et structure de la chaîne de transport des électrons avec ses quatre complexes protéiques (I , NADH deshydrogénase ; II, Succinate déshydrogénase; III, Ubiquinone-cytochrome c oxydoréductase; IV, Cytochrome oxydase) et l'ATP synthase.

1.2 L'ÉVOLUTION DE L'ADNmt

La place essentielle qu'occupe la mitochondrie dans les processus vitaux, la transmission uniparentale maternelle, la quasi-absence de recombinaison et la petite taille de son génome ont contribué à placer les études sur l'ADNmt à l'avant-scène des travaux en phylogénie et en génétique des populations. Malgré qu'une grande partie du génome mitochondrial d'origine fut transféré vers le génome de l'hôte eucaryote au cours de l'évolution (Lang, Paquin *et al.*, 2000; Selosse, Albert *et al.*, 2001; Herrmann, 2003), la conservation d'un génome mitochondrial suggère que son rôle central lui aurait valu de fortes contraintes évolutives (Saccone *et al.*, 2000, 2002a, 2002b). De plus, l'importance pour la production d'ATP des 13 peptides codés par l'ADNmt, associés avec les 65 sous-unités nucléaires de l'ETS, laisse suspecter l'action de processus de sélection où l'accumulation aléatoire de mutations devrait ultimement se traduire en des conséquences majeures sur le métabolisme énergétique, en affectant les propriétés fonctionnelles des protéines mitochondriales. L'intégrité fonctionnelle de ces protéines serait aussi dépendante d'une forte co-adaptation des génomes mitochondrial et nucléaire, tous deux responsables de la production des enzymes d'ETS (Blier, Dufresne *et al.*, 2001; Schmidt, Wu *et al.*, 2001; Doiron, Bernatchez *et al.*, 2002; Schmidt, Goodman *et al.*, 2002; Willett & Burton, 2004).

Toutefois, sa petite taille devant entraîner proportionnellement plus d'erreurs de réplication, l'absence d'histones, le cliquet de Muller (*Muller's ratchet*) et l'absence

générale de recombinaison due à la transmission uniparentale maternelle contribueraient au fort taux de mutations observés dans le génome mitochondrial (Rand, 2001; Saccone, Gissi *et al.*, 2002). Cet apparent paradoxe suggère qu'un haut taux de substitutions (i.e. mutations fixées au cours de l'évolution) n'est pas incompatible avec la conservation du génome et des propriétés fonctionnelles de ses gènes. Ainsi, contrairement à l'idée largement partagée qu'un gène évoluant rapidement est moins sujet à être sous forte contrainte évolutive qu'un gène évoluant lentement, les contraintes évolutives pourraient s'exprimer autrement qu'uniquement par des variations du taux de substitutions. Comme il le sera discuté dans l'article présenté au chapitre 2, d'autres facteurs pourraient déterminer ces patrons évolutifs. Enfin, malgré toutes ces restrictions, le maintien d'un génome mitochondrial suggère donc que ces gènes essentiels au métabolisme aérobie et persistants dans le génome mitochondrial sont soit difficilement transférables dans le noyau de par les caractéristiques physico-chimiques des peptides qu'ils codent, soit l'évolution ne leur en a tout simplement pas encore donné l'occasion (voir le cas de l'ATP8, chapitre 3). De même, il se peut que leur présence soit requise pour assurer un contrôle adéquat de leur expression (Hittel & Storey, 2002; Seidel-Rogol & Shadel, 2002; Villena, Carmona *et al.*, 2002).

Si la plupart des mutations survenant dans l'ADNmt sont délétères (Rand, 2001), elles sont éliminées au cours de l'évolution pour préserver le métabolisme mitochondrial, lui-même essentiel pour le bon fonctionnement de la cellule. Les mutations doivent donc principalement être neutres ou presque neutres pour être maintenues dans le génome, comme le soutient la théorie neutre de l'évolution (Kimura, 1968; Kimura, 1980; Kimura,

1983; Kimura, 1990; Kimura, 1991). Ainsi, même dans une perspective neutre ou quasi-neutre, les substitutions doivent respecter les limites de changements moléculaires permettant de maintenir le métabolisme aérobie. Étant donné que la structure d'un peptide est étroitement liée à sa fonction (Petsko & Ringe, 2004), les caractéristiques structurales des peptides mitochondriaux définiraient ces limites de changements moléculaires préservant ETS et OXPHOS. Dans le métabolisme aérobie, chaque peptide mitochondrial a conservé sa fonction générale au cours de l'évolution des vertébrés. Conséquemment, son importance pour la production d'énergie dans la cellule devrait favoriser des changements évolutifs contribuant à la stabilité structurale et fonctionnelle des peptides mitochondriaux, et ce, avec une certaine indépendance face à la phylogénie des espèces, c'est-à-dire par le biais d'une évolution convergente des propriétés fonctionnelles de ces peptides, lorsque nécessaire.

Dans de rares cas, si les propriétés physico-chimiques ne sont pas conservées, il serait possible d'observer le processus d'adaptation moléculaire dans des taxons partageant des caractéristiques physiologiques pouvant interagir avec les propriétés physico-chimiques de la cellule (Hochachka & Somero, 2002). La convergence des acides aminés du génome mitochondrial devrait alors répondre à divers facteurs biologiques et environnementaux incluant la taille et la masse corporelle, le taux métabolique, le temps de génération, la taille de la population effective, les températures environnementale et corporelle, etc. (Martin & Palumbi, 1993; Rand, 1994; Martin, 1995; Templeton, 1996; Yang, Nielsen *et al.*, 1998;

Rand, 2001; Angilletta, Niewiarowski *et al.*, 2002; Somero, 2002; Somero, 2005), lesquels seront également analysés au chapitre 2.

Par exemple, vu la grande sensibilité de la mitochondrie à la température (Blier & Lemieux, 2001) – variable écophysiologique majeure imposant des limites strictes aux organismes (Angilletta, Niewiarowski *et al.*, 2002; Hochachka & Somero, 2002) – de même que les impacts potentiels d'une simple substitution sur les propriétés fonctionnelles et la sensibilité thermique de ses enzymes (Holland, McFall-Ngai *et al.*, 1997), il est concevable que l'évolution du génome mitochondrial permette de s'ajuster aux différents régimes thermiques auxquels les espèces sont confrontées, et ce, malgré l'existence des nombreuses autres variables déjà mentionnées (Martin & Palumbi, 1993; Rand, 1994; Martin, 1995; Templeton, 1996).

La nature physico-chimique des acides aminés, liée à la sensibilité thermique des propriétés enzymatiques des protéines de l'ETS, influencerait les substitutions en acides aminés (Yang, Nielsen *et al.*, 1998). Ces modifications pourraient se manifester par une sélection des acides aminés dont les propriétés balancent la charge électrique totale de la protéine (Schmidt, Wu *et al.*, 2001) selon la température, lesquels influencent directement les interactions physico-chimiques tant avec la membrane lipidique qu'avec les enzymes qui y sont intégrées. La température influence d'une part les réactions biologiques car la réactivité des molécules dépend de leur énergie cinétique, étroitement liée à la température, de même qu'à l'action des enzymes, lesquelles sont thermiquement affectées par le biais de

changements de conformation nécessaire à leur fixation au substrat (Eisenberg & Crothers, 1979; McQuarrie & Rock, 1991; Hochachka & Somero, 2002). Les processus chimiques sont par conséquent facilités par un accroissement de température. D'autre part, la fluidité membranaire est à la fois influencée par la structure des lipides amphipathiques ainsi que par les changements de température (Lucu & Towle, 2003), affectant probablement les protéines mobiles de l'ETS, notamment le cytochrome *c*.

Cela laisse supposer qu'en température froide, l'efficacité catalytique des enzymes membranaires est maintenue par une flexibilité accrue des enzymes, permise par une diminution des liaisons faibles (Hochachka & Somero, 2002), de même que de l'ensemble des types de liaisons dont la force augmente en général avec une diminution de température. L'énergie cinétique moyenne des molécules est corrélée avec la température, consolidant ainsi les liens moléculaires à basse température (Eisenberg & Crothers, 1979; McQuarrie & Rock, 1991). Toutefois, la force des liaisons hydrophobes a la particularité de s'amenuiser avec une diminution de la température. Les liens hydrophobes impliquent l'action d'un solvant, l'eau en l'occurrence, duquel les molécules non-polaires s'abritent. Comme l'entropie de l'eau diminue avec la température, la plus grande organisation des molécules d'eau contribue à diminuer la force des liens hydrophobes (Eisenberg & Crothers, 1979; Hochachka & Somero, 2002). Par conséquent, la sélection positive pourrait jouer un rôle dans l'ajustement des propriétés enzymatiques des protéines de l'ETS à la contrainte environnementale qu'est la température.

1.3 ADAPTATION OU NEUTRALITÉ?

Les peptides codés par l'ADNmt sont sans contredit essentiels à la survie de tout métazoaire dépendant du métabolisme aérobie. Chacun des 13 gènes mitochondriaux codant pour des protéines ont leurs caractéristiques propres, leur taux d'évolution particulier. Bien qu'on puisse s'entendre relativement bien sur les patrons évolutifs directement observables, l'inférence des processus évolutifs sous-jacents à ces patrons est souvent controversé (de Ricqlès, 1997). Est-ce que les contraintes structurales contribuent à la détermination du taux d'évolution des gènes mitochondriaux chez les vertébrés? Plus largement, est-ce que l'adaptation génotypique et phénotypique sont des déterminants important de l'évolution de l'ADNmt? Sinon, comment le hasard des changements moléculaires, circonscrits par les différentes contraintes existantes, peut-il rendre compte de l'évolution? Et puis, est-ce que le hasard ne brouille tout simplement pas les directions adaptatives que pourrait suivre les espèces, dépendantes des contraintes biologiques et environnementales? Ce qui expliquerait notre difficulté à identifier clairement dans les gènes le processus d'adaptation. L'évolution laisse-t-elle vraiment un signal moléculaire dans les gènes, ou bien l'évolution phénotypique repose-t-elle sur le développement et la croissance des organismes, sans forcément impliquer un changement inscrit dans les gènes codant pour des protéines?

Évidemment, pour observer directement les processus de l'évolution et enfin clore ces discussions, nous avons malheureusement un sérieux problème de temps. Toutefois, malgré

cette incapacité à confirmer complètement les modèles conceptuels élaborés, ceux-ci rendent compte d'une vision de la nature qui peut constamment être renforcée, améliorée, nuancée, raffinée. Il semble de plus en plus clair que la connaissance des gènes et de leurs fonctions ne suffit pas pour comprendre la vie. Malgré tout, sans vouloir réduire la vie à l'étude de ses différentes composantes, l'étude de l'ADN et des interactions biochimiques sous-jacentes à la construction des organismes vivants (Fox Keller, 2004) nous permet de se rapprocher toujours un peu plus de cette idée de l'étude intégrée des systèmes biologiques par la biologie intégrative, dont le paradigme est quotidiennement enrichi par les nouveaux outils bioinformatiques. Alors, plutôt que de rester muets devant notre fascination pour l'évolution, bien des questions restent à poser...

1.4 OBJECTIFS DE RECHERCHE

Afin d'évaluer si les propriétés physico-chimiques assurent l'intégrité fonctionnelle des peptides mitochondriaux, vérifier si les changements évolutifs de l'ADNmt contribuent à la convergence des propriétés fonctionnelles selon les facteurs biologiques et environnementaux susceptibles d'interférer avec les caractéristiques physiologiques.

OBJECTIFS SPÉCIFIQUES

1. Définir le patron évolutif des propriétés physico-chimiques des 13 peptides codées par l'ADNmt. Pour ce faire ;
    a. Définir les propriétés physico-chimiques des peptides codées par l'ADNmt par l'indice hydropathique lequel assigne à chaque acide aminé une valeur reflétant leur hydrophilie ou leur hydrophobie relative (Kyte & Doolitle, 1982). L'indice hydropathique total dépend ainsi de la composition en acides aminés des peptides.
    b. À partir de ces indices, évaluer les distances physico-chimiques séparant chacune des espèces pour chacun des peptides codés par l'ADNmt.
    c. Quantifier la divergence évolutive totale, neutre et non-neutre, nucléotidique et en acides aminés des 13 gènes codant pour un peptide.
    d. Vérifier si les distances physico-chimiques sont corrélés avec la divergence évolutive.

Hypothèse 1 : L'évolution des propriétés physico-chimiques des peptides codés par l'ADNmt est neutre (ou suit la phylogénie de façon stricte). Prédiction centrale : une corrélation parfaite entre les distances physico-chimiques et les divergences évolutives indique une évolution neutre.

2. Vérifier si les propriétés physico-chimiques des peptides codées par l'ADNmt se sont ajustées au régime thermique de même qu'aux autres facteurs biologiques et environnementaux susceptibles d'interférer avec les caractéristiques physiologiques des vertébrés. Pour ce faire ;
    a. Élaborer une base de données des variables susceptibles d'interférer avec les caractéristiques physiologiques (incluses dans le tableau 1) des espèces de vertébrés sélectionnées (tableau 2).
    b. Comparer les propriétés physico-chimiques avec l'ensemble des variables choisies pour caractériser les espèces (tableau 1).
        i. Le cadrage multidimensionnel non-métrique (MDS) (Krebs, 1999) fera ressortir les variables les plus susceptibles d'expliquer les propriétés physico-chimiques des peptides codées par l'ADNmt.
        ii. La méthode des moindres carrés (Krebs, 1999) permettra d'établir les liens de corrélations entre les différentes variables.

Hypothèse 2 : La variable α [ex. régime thermique] explique l'évolution convergente des propriétés physico-chimiques.

Tableau 1. Variables proposées pour les analyses statistiques devant expliquer les propriétés physico-chimiques des peptides codées par l'ADNmt.

Variables à l'étude	**Type de variable**
Espèce et groupe taxonomique	Nominale
Gène	Nominale
Composition en acides aminés	Nominale
Caractéristiques physico-chimiques des acides aminés	Nominale
Indice d'hydropathie des acides aminés	Quantitative discrète
Distance évolutive (neutre/non-neutre/totale)	Quantitative aléatoire
Caractéristiques thermiques (homéothermie, endothermie, ectothermie, sténothermie, eurythermie)	Nominale
Fenêtres de tolérance et de préférence thermique	Quantitative aléatoire
Variations annuelles de température du milieu	Quantitative aléatoire
Taux d'activité métabolique indirect (activité de locomotion, fréquence de reproduction, croissance, système ventilatoire et circulatoire)	Nominale
Facteurs associées à la plasticité (différence entre la capacité métabolique maximale et *in vivo*, densité mitochondriale, etc.)	Quantitative aléatoire et nominale

CHAPITRE 2

LA VARIABILITÉ PHYSICO-CHIMIQUE PRÉDIT LE TAUX D'ÉVOLUTION DES PEPTIDES CODÉS PAR L'ADNmt CHEZ LES VERTÉBRÉS

RÉSUMÉ

Les changements physico-chimiques des peptides codés par l'ADNmt pouvant prédire l'évolution de leurs acides aminés ont été explorés chez 164 espèces de vertébrés. Les génomes mitochondriaux complets de ces vertébrés ont été échantillonnés sur le site Web de NCBI. À partir de ces séquences, nous avons aligné les 13 protéines codées par l'ADNmt à l'aide de ClustalX 1.83, et nous avons défini les distances physico-chimiques de ces protéines mitochondriales à partir des indices hydropathiques de Kyte & Doolittle. Nous avons ainsi généré un modèle des changements physico-chimiques en fonction des distances évolutives en acides aminés, lequel montre une forte corrélation positive (valeurs de probabilité entre P = 0,001 et P = 0,0005, avec des valeurs de corrélation "R" allant de 0,8200 à 0,9013, selon le taxon). La compilation de l'ensemble des valeurs en dehors de la prédictiondu modèle, i.e. excédant l'intervalle de confiance de 95%, montre que la sélection négative est clairement dominante dans l'ADNmt. De plus, notre modèle nous a permis d'identifier plusieurs patrons évolutifs, notamment l'augmentation de la variabilité physico-chimiques des peptides mitochondriaux chez les oiseaux, de même qu'une augmentation

générale de la sensibilité du complexe I (NADH déshydrogénase) aux variations physico-chimiques. Le modèle prédit la capacité des protéines à supporter des changements physico-chimiques sans affecter les propriétés fonctionnelles, suggérant que plus un gène tolère de grands changements physico-chimiques, plus il est susceptible d'évoluer rapidement. Étant donné qu'il n'y a pas d'évidence claire que la longueur du gène ou sa position par rapport à l'origine de réplication sont liées au taux d'évolution, la tolérance aux changements physico-chimiques serait le meilleur indice connu pour expliquer les différences de taux d'évolution entre les gènes codées par l'ADNmt.

PHYSICO-CHEMICAL VARIABILITY PREDICTS EVOLUTIONARY RATES OF mtDNA-ENCODED PEPTIDES IN VERTEBRATES

ABSTRACT

We explored the general predictability of physico-chemical changes in mtDNA-encoded peptides as a function of their amino acid evolution in 164 vertebrate mitochondrial genomes from NCBI website. We carried out multiple alignments of the 13 protein-coding mitochondrial genes with ClustalX 1.83, and defined physico-chemical distances of mitochondrial peptides from hydropathic indexes of Kyte & Doolittle. We generated a model of physico-chemical changes as a function of amino acid distances which shows strong positive correlation (probability values between $P = 0.001$ and $P = 0.0005$, with R correlation values ranging from 0.8200 to 0.9013, according to the vertebrate taxon). We compiled every deviation from the model's prediction and as expected, negative selection is clearly dominant in mtDNA. Moreover, we discuss many evolutionary patterns identified with our model, such as increased physico-chemical variability in birds, and a general sensitivity to physico-chemical variability in the complex I, the NADH dehydrogenase. The model predicts the capacity of proteins to suffer hydropathic changes without affecting functional properties, suggesting that more a gene tolerates physico-chemical variability, more it can suffer important number of substitutions, and consequently, evolve faster. Since there is no evidence that gene length and the localization to the origin of replication is linked to evolutionary rate, tolerance to physico-

chemical variability would be the best predictor known to explain rate variability among mtDNA-encoded peptides.

2.1 INTRODUCTION

According to the neutral theory (Kimura, 1968; Kimura, 1983), the great majority of substitutions, randomly fixed through sampling drift during evolution, are neutral or nearly neutral. The observed evolutionary distances between species' DNA sequences (i.e. substitutions) depends on the mutation rate and the probability of fixation (Kimura, 1983; Rand, 2001). This probability is gene and site dependant, ranging from a weak number of substitutions in conserved genes, to a rate of substitution close to that of pseudogenes (or "dead" genes) (Kimura, 1990; Kimura, 1991). Neutrality tests show that deleterious mutations are common in mitochondrial DNA (mtDNA)(Rand, 2001), suggesting that the mitochondrial metabolism need to be preserved to insure cellular functions. Thus, even in a neutral perspective, substitutions should respect the minimal requirements for the maintenance of protein function.

Since protein structure is intimately linked to its function (Petsko & Ringe, 2004), structural characteristics of mitochondrial peptides would define the minimal requirement, for the maintenance of protein function in the electron transport system (ETS) and the oxidative phosphorylation (OXPHOS). Protein structure is principally determined by amino acid sequences and their physico-chemical properties including hydrophobicity levels (Kyte & Doolitle, 1982; Uversky, Gillespie *et al.*, 2000). In the aerobic metabolism, each mtDNA-encoded peptide conserved its function through vertebrate evolution. Therefore, if physico-chemical properties, associated to the different levels of structure, strictly

determine protein functions in mtDNA, then amino acid evolutionary distances between vertebrate species should not be linked with structural or physico-chemical properties. In other words, the evolution of physico-chemical properties of proteins would not follow phylogenetical constraints, but it would rather evolve in a convergent fashion, principally to conserve physico-chemical characteristics of each polypeptide that has conserved similar functions. In rare cases, if physico-chemical properties are not conserved, it would be possible to observe adaptive processes in groups sharing physiological characteristics which could interact with physico-chemical properties in the cell (Hochachka & Somero, 2002). Convergence of amino acid evolution in the mitochondrial genome could then respond to diverse environmental and biological specificities, including body weight and height, metabolic rates, generation time, effective population size, environmental and body temperature, etc. (Martin & Palumbi, 1993; Rand, 1994; Martin, 1995; Templeton, 1996; Yang, Nielsen *et al.*, 1998; Rand, 2001; Angilletta, Niewiarowski *et al.*, 2002; Somero, 2002).

In order to shed light on the role of physico-chemical properties in the evolution of vertebrate mtDNA, we explore the general predictability of physico-chemical changes in mtDNA-encoded peptides as a function of their amino acid evolution. We show how a general pattern can be a useful tool to understand mtDNA evolutionary processes of particular species of vertebrates, including actinopterigii, amphibia, reptilia, aves, and mammalia, as well as specific patterns of evolution for each of the 13 mtDNA-encoded peptides.

2.2 METHODS

We chose 164 vertebrate mitochondrial genomes from GenBank (NCBI website: www.ncbi.nlm.nih.gov) with particular attention to the biodiversity of each vertebrate group (table 2). The number of species was a limiting factor for two groups, amphibians and reptiles, for which there was only 7 and 11 complete mitochondrial genome sequences available.

Table 2. Number accession of mitochondrial genomes used in the study.

Actinopterigii	
Acipenser transmontanus (NC_004743)	*Nansenia ardesiaca (NC_004596)*
Alepocephalus tenebrosus (NC_004590)	*Neocyttus rhomboidalis (NC_004399)*
Allocyttus niger (NC_004398)	*Oncorhynchus mykiss (NC_001717)*
Anoplogaster cornuta (NC_004391)	*Oncorhynchus tshawytscha (NC_002980)*
Aulopus japonicus (NC_002674)	*Opisthoproctus soleatus (NC_004600)*
Bathylagus ochotensis (NC_004591)	*Ostichthys japonicus (NC_004394)*
Beryx splendens (NC_003188)	*Parazen pacificus (NC_004396)*
Caelorinchus kishinouyei (NC_003169)	*Petroscirtes breviceps (NC_004411)*
Caranx melampygus (NC_004406)	*Physiculus japonicus (NC_004377)*
Carassius auratus (NC_002079)	*Platytroctes apus (NC_004597)*
Caulophryne pelagica (NC_004383)	*Polyodon spathula (NC_004419)*
Cetostoma regani (NC_004389)	*Poromitra oscitans (NC_003172)*
Chaunax abei (NC_004381)	*Pterocaesio tile (NC_004408)*
Chaunax tosaensis (NC_004382)	*Retropinna retropinna (NC_004598)*
Chlorophthalmus agassizi (NC_003160)	*Rondeletia loricata (NC_003186)*
Cobitis striata (NC_004695)	*Salangichthys microdon (NC_004599)*
Crossostoma lacustre (NC_001727)	*Salmo salar (NC_001960)*
Cyprinus carpio (NC_001606)	*Salvelinus fontinalis (NC_000860)*
Danacetichthys galathenus (NC_003185)	*Sarcocheilichthys variegatus microoculus*
Danio rerio (NC_002333)	*(NC_004694)*
Eutaeniophorus sp.033-Miya (NC_004390)	*Sargocentron rubrum (NC_004395)*
Euthynnus alleteratus (NC_004530)	*Saurida undosquamis (NC_003162)*
Gadus morhua (NC_002081)	*Scaphirhynchus cf. Albus (NC_004420)*
Glossanodon semifasciatus (NC_004595)	*Scopelogadus mizolepis (NC_003171)*
Harpadon microchir (NC_003161)	*Theragra chalcogramma (NC_004449)*
Lefua echigonia (NC_004696)	*Thunnus alalunga (NC_005317)*
Lophius americanus (NC_004380)	*Thunnus thynnus thynnus (NC_004901)*
Lota lota (NC_004379)	*Trachurus japonicus (NC_002813)*
Melanocetus murrayi (NC_004384)	*Zenion japonicus (NC_004397)*
Melanonus zugmayeri (NC_004378)	*Zenopsis nebulosus (NC_003173)*
Monocentris japonicus (NC_004392)	*Zeus faber (NC_003190)*

Amphibia	
Andrias davidianus (NC_004926)	*Ranodon sibiricus (NC_004021)*
Fervarya limnocharis (NC_005055)	*Typhlonectes natans (NC_002471)*
Mertensiella luschani (NC_002756)	*Xenopus laevis (NC_001573)*
Rana nigromaculata (NC_002805)	
Reptilia	
Alligator mississipiensis (NC_001922)	*Dogania subplana (NC_002780)*
Alligator sinensis (NC_004448)	*Eumeces egregius (NC_000888)*
Caiman crocodilus (NC_002744)	*Iguana iguana (NC_002793)*
Chelonia mydas (NC_000886)	*Pelomedusa subrufa (NC_001947)*
Chrysemys picta (NC_002073)	*Sphenodon punctatus (NC_004815)*
Dinodon semicarinatus (NC_001945)	
Mammalia	
Artibeus jamaicensis (NC_002009)	*Muntiacus reevesi (NC_004069)*
Balaenopterus musculus (NC_001601)	*Mus musculus (NC_005089)*
Balaenopterus physalus (NC_001321)	*Myoxus glis (NC_001892)*
Cebus albifrons (NC_002763)	*Ochotona collaris (NC_003033)*
Ceratotherium simum (NC_001808)	*Ornithorhynchus anatinus (NC_000891)*
Chalinolobus tuberculatus (NC_002626)	*Orycteropus afer (NC_002078)*
Cynocephalus variegatus (NC_004031)	*Oryctolagus cuniculus (NC_001913)*
Dasypus novemcinotus (NC_001821)	*Ovis aries (NC_001941)*
Didelphis virginiana (NC_001610)	*Papio hamadrias (NC_001992)*
Dugong dugon (NC_003314)	*Physeter catodon (NC_002503)*
Echinosorex gymnura (NC_002808)	*Pongo pygmaeus abelii (NC_002083)*
Elephas maximus (NC_005129)	*Procavia capensis (NC_004919)*
Equus asinus (NC_001788)	*Pteropus dasymallus (NC_002612)*
Equus caballus (NC_001640)	*Pteropus scapulatus (NC_002619)*
Erinaceus europaeus (NC_002080)	*Rattus norvegicus (NC_001665)*
Eumetopias jubatus (NC_004030)	*Rhinoceros unicornis (NC_001779)*
Felis catus (NC_001700)	*Sciurus vulgaris (NC_002369)*
Gorilla gorilla (NC_001645)	*Soriculus fumidus (NC_003040)*
Halichoerus grypus (NC_001602)	*Sus scrofa (NC_000845)*
Hemiechinus auritus (NC_005033)	*Tachyglossus aculeatus (NC_003321)*
Hippopotamus amphibius (NC_000889)	*Talpa europaea (NC_002391)*
Isoodon macrourus (NC_002746)	*Tamandua tetradactyla (NC_004032)*
Lemur catta (NC_004025)	*Tapirus terrestris (NC_005130)*
Lepus europaeus (NC_004028)	*Trichosurus vulpecula (NC_003321)*
Loxodonta africana (NC_000934)	*Ursus americanus (NC_003426)*
Macropus robustus (NC_001794)	*Ursus maritimus (NC_003428)*
Manis tetradactyla (NC_004027)	*Volemys kikuchii (NC_003041)*
Muntiacus criniformis (NC_004577)	*Vombatus ursinus (NC_003322)*
Aves	
Anomalopteryx didiformis (NC_002779)	*Dromaius novaehollandiae (NC_002784)*
Anser albifrons (NC_004539)	*Emeus crassus (NC_002673)*
Apterys haastii (NC_002782)	*Endromia elegans (NC_002772)*
Arenaria interpres (NC_003712)	*Eudyptula minor (NC_004538)*
Aythya americana (NC_000877)	*Falco peregrinus (NC_000878)*
Buteo buteo (NC_003128)	*Gallus gallus (NC_001323)*
Casuarius casuarius (NC_002778)	*Haematopus ater (NC_003713)*
Ciconia boyciana (NC_002196)	*Pterochemia pennata (NC_002783)*
Ciconia ciconia (NC_002197)	*Rhea americana (NC_000846)*
Corvus frugilegus (NC_002069)	*Smithornis sharpei (NC_000879)*

Aves	
Coturnix chinensis (NC_004575)	*Struthio camelus (NC_002785)*
Coturnix japonica (NC_003408)	*Tinamus major (NC_002781)*
Dinornis giganteus (NC_002672)	*Vidua chalybeata (NC_000880)*

We carried out multiple alignments of the 13 protein-coding mitochondrial genes for each vertebrate group with ClustalX 1.83 (Thompson, Gibson *et al.*, 1997). Manual adjustments were made when necessary. Since our analyses were carried out on a great diversity of species, we had to circumvent many problems, such as strong occurrences of multiple hits, biases in base composition and in rate of transversions and transitions, and evolutionary rate variations among sites (Li, 1997). We used amino acid sequences to limit these difficulties.

Distance analyses were conducted using MEGA 2.1 (Kumar, Tamura *et al.*, 2001). We used Kimura two-parameters (Kimura, 1980), recognized as a good tool for distance analyses of highly divergent sequences (Kumar, 1996) with a gamma correction of α = 0.65, corresponding to a general mean of rate variation among sites in mtDNA (Gu & Zhang, 1997). Although sites in these genes are probably evolving differently, large data set appear to outweigh concerns of model variations (Faith & Pollock, 2003). Since it was not possible to integrate every specificity of our extensive data base in the calculation of distances, this simple model allowed us to extract a general pattern of mitochondrial DNA evolution in vertebrates. We abandoned the idea of detailed predictability to concentrate on predictable statistical distribution of metabolic events, a systemic perspective resumed by Bak and Packuski (1995).

To define physico-chemical properties of mitochondrial peptides, we calculated hydropathic indexes (Kyte & Doolitle, 1982) for each mtDNA-encoded peptides for each species. We used hydropathic differences between pairs of species to generate physico-chemical distances (ΔHI) comparable to amino acid evolutionary distances. We generated a model from the comparison between mean amino acid distances and physico-chemical distances from hydropathic indexes of the 13 mtDNA-encoded peptides for actinopterigii, amphibia, reptilia, mammalia, and aves. The means of amino acid distances (Daa) and physico-chemical distances include every possible pairs of species within each vertebrate class. Then, the general model includes one point for each gene, for each vertebrate taxon (corresponding to 65 data). We validate the general model in generating particular model inside each vertebrate taxon.

Models have been generated with the reduced major axis used for two random variables; amino acid distances and physico-chemical distances. The reduced major axis method (Kermack & Haldane, 1950; Riggs, Guarnieri *et al.*, 1978; Webb, Howe *et al.*, 1981) also gives the standard deviation of slopes and "R" associated to each model. The model significance was obtained from *Table Curve 2D Windows 407* software. We calculated 95% confidence interval of slope (Jolicoeur, 1998) to identify cases that show a departure from the prediction of the model. Comparison in pairs of species exceeding the 95% confidence interval of hydropathic index difference are suspected to include at least one species having faced strong selective pressure during its evolutionary history.

2.3 RESULTS

NEUTRAL PREDICTIONS OF PHYSICO-CHEMICAL EVOLUTIVE CHANGES OF mtDNA-ENCODED PEPTIDES

The model shows a strong positive correlation between physico-chemical and amino acid distances (figure 2). This corresponds to an increase in physico-chemical variability of mitochondrial proteins when amino acid distances increase. Since the observations are independent because of a random sampling that covers major vertebrate groups, and since the correlation between physico-chemical properties and amino acid distances is highly significant (R=0.8389; P=0.0005), we can assume this correlation as the neutral or quasi-neutral prediction of random molecular evolution of these proteins (table 3). The model predicts the capacity of proteins to suffer hydropathic changes without affecting functional properties i.e. their tolerance to physico-chemical variability. Because adaptive processes are effective at small scales (i.e. inside species and populations), large evolutionary times clear every particular deviation from the neutral prediction of the model.

To link protein functions to their physico-chemical properties, comparisons between groups sharing physiological characteristics which could interact with physico-chemical properties of proteins in the cell, such as endothermy and ectothermy (Hochachka & Somero, 2002) was necessary. We therefore tested the model at smaller scales, inside monophyletic vertebrate class. It should be mentionned that the few amphibian mitochondrial genomes available in GenBank generate a less exhaustive model which

could drive to some deviation from the "real" neutral hypothesis. Reptiles have been excluded because they form a paraphyletic unity, which does not respect the phylogenetical constraint, that all descendants should come from the same common ancestor. However, even if they do not constitute a good model to use, they show a positive correlation, as do other vertebrates, which partly explains why the model including all vertebrates has a slope weaker than models in other vertebrate classes (table 3). Aves constitute a particular case. For every amino acid change, they suffer significantly greater hydropathic changes if compared with the general pattern present in vertebrates (table 3).

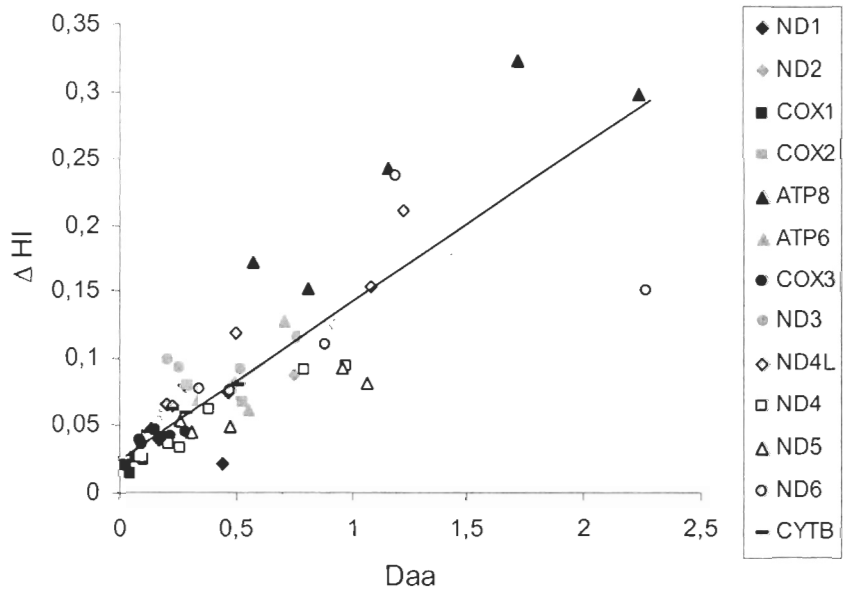


Figure 2. Neutral prediction of physico-chemical changes (ΔHI) as a function of amino acid distances (Daa) in five classes of vertebrates. Each point corresponds to mean physico-chemical and amino acid distances in one of the 13 mtDNA encoding genes in a specific vertebrate group.

Table 3. Neutral predictions (model) of physico-chemical changes in mitochondrial genes coding for proteins in vertebrates. Models have been generated with the reduced major axis used for two random variables with amino acid distances in "Daa" and physico-chemical distances in "ΔHI". The standard deviation of slopes and "R" associated to each model and model significance are also presented in the table. Aves show a significantly greater slope (in dark characters in the table) deduced from standard deviation values.

Vertebrate class	Model (Reduced major axis)	Slope SD	R	Model significance
Vertebrate	ΔHI = 0.1331Daa + 0.0172	0.0090	0.8389	P = 0.0005
Actinopterigii	ΔHI = 0.1731Daa + 0.0093	0.0223	0.8853	P = 0.0005
Amphibia	ΔHI = 0.1837Daa – 0.0230	0.0221	0.9013	P = 0.0005
Reptilia	ΔHI = 0.1093Daa + 0.0155	0.0169	0.8297	P = 0.0005
Mammalia	ΔHI = 0.1767Daa + 0.0088	0.0281	0.8200	P = 0.001
Aves	**ΔHI = 0.2817Daa – 0.0007**	0.0405	0.8549	P = 0.0005

To identify the genes subject to any kind of selection on their physico-chemical characteristics, we looked at different pairs of species. Comparison in pairs exceeding over or under the 95% confidence interval may reveal gene under natural selection. We compiled every deviation from the neutral prediction as a percentage of deviant couples on every possible comparison. We present here only deviant pairs of species exceeding at least by twice the 95% limits of the models in actinopterigii, mammalia, and aves (table 4).

As expected, negative selection is clearly dominant in mtDNA being 3 to 4 fold the value of positive deviation no matter the vertebrate class or the evolutionary distance (table 4). Therefore, physico-chemical properties are conserved in mtDNA-encoded peptides in ray-finned fishes (actinopterigii) as well as in endothermic classes including birds and mammals. An interesting exception is ND3. ND3 is the only mitochondrial gene

consistently out of the neutral prediction, clearly under either relaxed constraints or positive selection in every vertebrate group analysed (underlined data in the table 4).

Table 4. Deviant pairs of species exceeding at least by twice the 95% limits of the models in actinopterigii, mammalia, and aves. Each value corresponds to the percentage of deviant pairs of species on every possible comparison (1890 comparisons in actinopterigii, 1540 in mammals, and 325 in birds). Negative deviations indicate values below the 95% confidence interval and positive deviations, values over the 95% confidence interval. Underlined characters indicate genes exceeding the average percentage of positive deviation in the three taxa simultaneously, whereas dark characters indicate genes exceeding the average percentage of negative deviation in the three taxa simultaneously.

	Actinopterigii		Mammalia		Aves	
mt Gene	Negative deviation	Positive deviation	Negative deviation	Positive deviation	Negative deviation	Positive deviation
ND1	21.47	3.75	10.58	7.14	11.69	3.69
ND2	**27.23**	4.07	**34.22**	0.26	**41.54**	0.00
COX1	3.81	0.05	0.06	0.00	0.00	0.00
COX2	9.20	6.24	5.32	7.21	11.38	2.46
ATP8	20.47	10.68	16.56	12.79	19.38	8.31
ATP6	25.28	4.55	15.91	3.70	16.92	4.31
COX3	8.30	8.25	5.39	2.53	11.08	7.08
ND3	13.86	25.65	10.91	22.40	10.15	25.85
ND4L	13.22	16.82	15.97	8.96	25.54	7.38
ND4	**36.44**	0.05	**24.16**	0.91	**36.92**	0.00
ND5	**33.26**	0.79	**43.12**	0.06	**30.46**	0.31
ND6	**25.65**	6.24	**29.87**	3.38	**26.77**	1.85
CYTB	19.67	2.38	9.48	3.64	5.23	1.23
AVERAGE % of deviation	**19.83**	6.89	**17.04**	5.61	**19.01**	4.80

In our case study, mitochondrial gene evolution seems to be particularly associated with gene structural and functional properties rather than with animal lineage properties. Evolutionary patterns are gene dependant throughout vertebrates. There is great variability between genes whereas no such variability is observed if we compare percentage of variation by groups. Genes from the NADH dehydrogenase complex, the first complex of the electron transport chain, seem particularly and largely affected by physico-chemical

changes. ND2, ND4, ND5, and ND6 are exceeding the negative deviation mean simultaneously for the three studied taxa (in dark characters in the table 4). As mentioned above, ND3 variation comes from a great positive deviation, but in general, NADH dehydrogenase complex clearly suffers great evolutionary restrictions. Surprisingly, the three mitochondrial subunits of the cytochrome c oxydase, which are generally thought to be under great negative selection principally because of their slow evolutionary rates (Saccone, De Giorgi *et al.*, 1999; Saccone, Gissi *et al.*, 2000) and their physiological importance (Yoshikawa, Shinzawa-Itoh *et al.*, 1999), are rarely out of the model's prediction in our comparisons in pairs of species.

2.4 DISCUSSION

TOLERANCE TO PHYSICO-CHEMICAL VARIABILITY AND mtDNA EVOLUTION

Tracing the way between genes, phenotypes and their evolution is quite a difficult exercise, far beyond statistical evidence (Creevey & McInerney, 2002). Since most mutations are eliminated in the course of evolution (Li, 1997) and thus, clear every particular deviation from the neutral prediction, we suppose that adaptive processes are effective at little scales (i.e. inside species and populations), or at least, in groups sharing physiological characteristics which could interact with physico-chemical properties of the proteins in the cell (Hochachka & Somero, 2002). Therefore, a great number of factors may be implicated in the adaptive evolution of the mitochondrial genome. The evolution of nuclear genes coding for peptides interacting with mtDNA-encoded peptides is suspected to modify evolutionary rates expected from the neutral prediction in a co-adaptive fashion (Blier, Dufresne *et al.*, 2001; Willett & Burton, 2004). Moreover, body weight and height, metabolic rates, generation time, effective population size, as well as environmental and body temperature may intervene (Martin & Palumbi, 1993; Rand, 1994; Martin, 1995; Templeton, 1996; Rand, 2001).

Temperature, a major ecophysiological variable imposing strict limits to organisms (Angilletta, Niewiarowski *et al.*, 2002; Hochachka & Somero, 2002; Somero, 2002), directly affects mitochondrial metabolism (Blier & Lemieux, 2001). Thermoregulation strategies of species, whose differences are maximised between endothermy and

ectothermy, may thus surpass species clusters in evolutionary processes, and direct evolution independently of phylogenetic constraints. The physico-chemical nature of amino acids, linked to thermal adjustments of enzymatic properties, may influence amino acid substitutions (Yang, Nielsen *et al.*, 1998). This would contribute to the selection of amino acids with specific electric charges and physico-chemical properties (Schmidt, Wu *et al.*, 2001) according to temperature. These properties directly influence physico-chemical interactions in lipid membranes as well as of the enzymes integrated into them. Cellular membrane are indeed sensitive to amphipathic lipid structure and temperature changes (Lucu & Towle, 2003; Somero, 2005). Moreover, biological reactions are tightly linked to temperature, both through enzyme kinetics and enzyme conformation changes necessary to the fixation on the substrate (Hazel & Prosser, 1974; Hazel & Schuster, 1976; Eisenberg & Crothers, 1979; McQuarrie & Rock, 1991; Hochachka & Somero, 2002; Petsko & Ringe, 2004). Consequently, biochemical processes are facilitated by an increase in temperature.

We could then suppose that in cold temperatures, catalytic efficiency of membrane enzymes are maintained through an enhancement of their flexibility. This is possible if weak chemical links are diminished (Hochachka & Somero, 2002). The kinetic energy of molecules is correlated with temperature, where molecular links inside proteins are strengthened at low temperatures (Eisenberg & Crothers, 1979; McQuarrie & Rock, 1991; Kingsolver & Huey, 1998). Hydrophobic links imply the action of water, from which non-polar molecules will hide. Since entropy of water decreases with temperature lowering, a

greater organization of water molecule contributes to weaken hydrophobic links (Eisenberg & Crothers, 1979; Hochachka & Somero, 2002).

Therefore, hydrophobic amino acids would be favoured in colder species. Nevertheless, we did not discover any direct trend between body temperature and amino acid evolution, as already concluded for α-globin genes (Hamada, Horiike *et al.*, 2002). The hydropathic index appears to be a characteristic specific to each mitochondrial peptide (figure 3). In these peptides, hydrophobic amino acids are more likely conserved during evolution in order to maintain the structural foundation for evolving catalytic sites (Das, Miller *et al.*, 2004). Moreover, the effect of the local environment of amino acid site and protein location in relation to the mitochondrial membrane on the hydrophobic characteristics is well documented (Kumar, 1996; Tourasse & Li, 2000; Xia & Xie, 2002; Das, Miller *et al.*, 2004; Petsko & Ringe, 2004; Porto, Roman *et al.*, 2005). A surprising exception to the peptide-dependant hydrophobicity is ND4L in mammals, a small and variable mitochondrial gene that is overlapped by ND4 in the same way as ATP6 overlaps on ATP8 (Scheffler, 1999; Taanman, 1999).

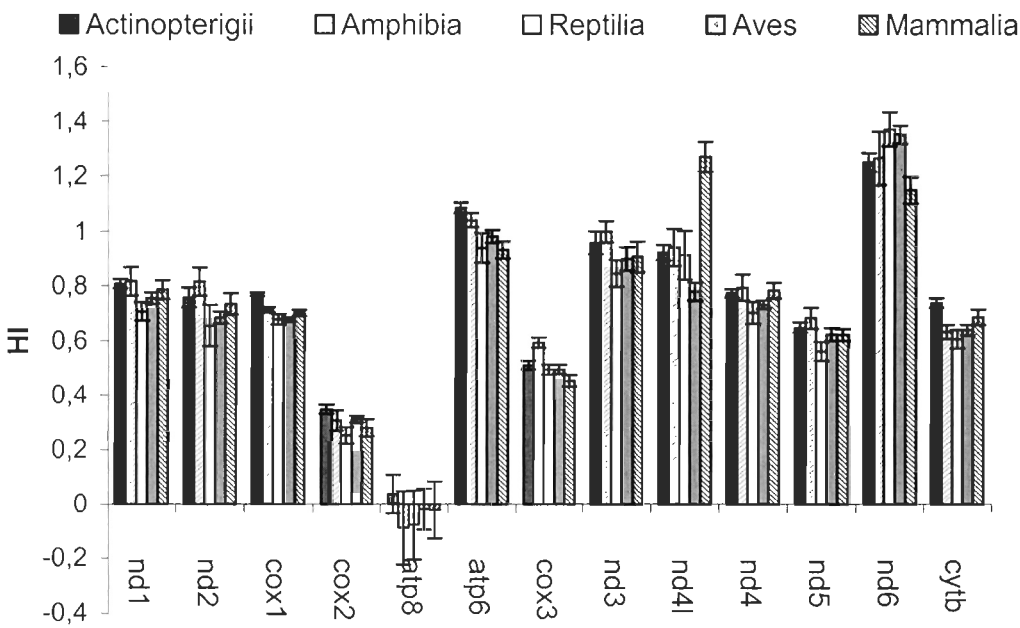


Figure 3. Hydropathic indexes (HI) compilation by vertebrate groups for each mtDNA-encoded peptide.

Because of the high significance of predictions in each of the studied vertebrate groups (table 3), our models confidently correspond to the possible physico-chemical variation inside the "biological solution", i.e. the metabolic pathway common to all vertebrates. These limits indicate the amplitude of changes generally accepted by species in a taxon, where physico-chemical properties determine if amino acid replacements are acceptable and thus conserved during evolution (Moilanen & Majamaa, 2003; Das, Miller *et al.*, 2004). As already hypothesized, a neutral prediction should predict evolutionary changes that follow the "biological solution". Therefore, more a gene shows a great tolerance to physico-chemical variability, more it could suffer an important number of substitutions, and consequently, evolve faster. Even if we observed the same trend between

physico-chemical variability (ΔHI) and amino acid distances (Daa) for each gene taken separately, there is a gene-dependant tolerance to ΔHI (table 5).

Table 5. Mean tolerance to physico-chemical variability (ΔHI) in mitochondrial genes and standard deviation (SD) associated to each value.

mtGene	ND1	ND2	COX1	COX2	ATP8	ATP6	COX3	ND3	ND4L	ND4	ND5	ND6	CYTB
ΔHI	0.071	0.104	0.024	0.055	0.237	0.072	0.042	0.106	0.123	0.063	0.064	0.131	0.057
SD	0.030	0.050	0.009	0.022	0.075	0.033	0.004	0.016	0.062	0.029	0.022	0.067	0.016

Since there is no evidence that gene length and the localization to the origin of replication is linked to evolutionary rate, tolerance to physico-chemical variability would be the best predictor known to explain rate variability among mtDNA-encoded peptides. Indeed, even if longer peptides might be more hydrophobic (Porto, Roman *et al.*, 2005) and thus indirectly interact with evolutionary rate, we do not find any such trend between evolutionary distances of mtDNA-encoded peptides and their length in our data. Finally, even if the hypothesis is compelling, the varying duration of time that mitochondrial genes spend in a single-strand state during replication, is not clearly linked to a variable probability of mutations (Gibson, Gowri-Shankar *et al.*, 2005). The localization of a gene on the circular genome would therefore not be responsible for the peptide-dependant rate of evolution in the mitochondrial genome.

The tolerance to physico-chemical variability defined by our model represents the general patterns observed in each of the five vertebrate taxa studied. At small scales, comparisons of pairs of species show that deviations from the neutral prediction are

common (table 4). As expected, negative selection is clearly dominant in mtDNA (Rand, 1994; Hasegawa, Cao *et al.*, 1998; Yang, Nielsen *et al.*, 2000; Gerber, Loggins *et al.*, 2001; Rand, 2001), indicating that fixed mutations are generally more conservative than what it is expected from neutral predictions. In these pairs of species, values of ΔHI as a function of Daa appear below the 95% limit of the model's prediction (table 3). Thus, the amount of physico-chemical changes defined by the model is higher than what would be expected as passive changes resulting from amino acid divergences. Physico-chemical properties might then be closely related to the catalytic function, and consequently, a slight modification in physico-chemical properties would be sufficient to disrupt physiological functions. Thus, fitness or survival depends on the conservation of physico-chemical properties. These highly constrained genes might allow us to identify physiological as well as environmental parameters that induce this strong pressure of evolutionary conservation on physico-chemical properties. These cases may help us link gene structure to metabolic function.

However, even if physico-chemical modifications could change metabolic properties through molecular interactions such as enzymatic affinity and kinetics (Eisenberg & Crothers, 1979; McQuarrie & Rock, 1991; Blier & Lemieux, 2001; Hochachka & Somero, 2002; Petsko & Ringe, 2004), new physico-chemical properties would not necessarily reveal the evolution of new metabolic properties (Kingsolver & Huey, 1998; Angilletta, Niewiarowski *et al.*, 2002). In the cases exceeding the expected value given by the model's prediction (table 3), a given gene in a pair of species suffers greater physico-chemical changes than expected for a given amino acid distance. To endure

this high level of physico-chemical changes in their mtDNA-encoded peptides, species might be apt to integrate these changes into slight variations of evolutionarily more labile traits, like physiological and behavioural traits (Angilletta, Niewiarowski *et al.*, 2002; Blomberg, Garland *et al.*, 2003) at higher functional levels. In other words, they would be under relaxed constraints.

Under relaxed constraints, a mutation driving to modifications in a proteins physico-chemical properties will have no effect on the general phenotype, or at least, will not disrupt physiological function. In fact, the high number of mtDNA molecules within the cell implies that mitochondrial mutations might accumulate gradually, without direct and immediate deleterious effects (Rand, 2001; Burger, Gray *et al.*, 2003). Fine tuning of enzyme activity may thus rely on more subtle mechanisms involving gene regulation, gene expression, and structure diversity (Scheffler, 1999; Carroll, 2000; Somero, 2002; Burger, Gray *et al.*, 2003; Somero, 2005), as well as on the capacity of mitochondria to adjust its concentration to a tissue's requirement (Scheffler, 1999). This fine tuning of enzyme activity would be accompanied by the use of a variety of integrated biochemical, physiological, and behavioural reactions in order to cope with environmental variations (Blomberg & Garland, 2002; Pörtner, 2002; Blomberg, Garland *et al.*, 2003; Somero, 2005).

This scenario seems more likely than positive selection where we could suspect evolution of new metabolic properties. It is worth noting that we analysed many variables

such as environmental and body temperature, body weight and height, nutritional habits, reproductive and behavioural characteristics, habitat and climatic characteristics with the MultiDimensional Scaling method (MDS statistics) and correlation methods (Krebs, 1999), but this yielded no significant result. Therefore, at great scales, the evolution of physico-chemical properties of an entire gene does not seem linked to local adaptation or environmental constraints. In others words, since molecular changes driving to a particular adaptation represent only rare events among neutral mutational changes (Golding & Dean, 1998), the evolution of metabolic pathways encoded by mtDNA of vertebrates may reflect a modern gene-dependant accumulation of neutral or quasi-neutral mutations marking quantitative differences between species.

GENE, SPECIES, AND SCALE SPECIFICITIES IN MITOCHONDRIAL GENOME EVOLUTION

Aves and physico-chemical changes

The case of aves shows that for every change in amino acid, they suffer significantly greater hydropathic changes when compared with the general vertebrate pattern (table 3: aves' model slope value is 0.2817 compared with 0.1331 for all vertebrates). The greater amplitude of physico-chemical changes inside the shared vertebrate mitochondrial metabolism could reveal changes in metabolic properties. However, since we can assume that positive selection represent rare mutational events at little scales, the probability to observe a general increase of physico-chemical variability driven by positive selection and leading to new metabolic properties, is unlikely. In a state of relaxed constraints, birds would rather be able to endure a few changes in their metabolism because they might be apt

to integrate these changes into slight variations of evolutionarily more labile traits, like physiological and behavioural traits (Angilletta, Niewiarowski *et al.*, 2002; Blomberg, Garland *et al.*, 2003) at higher functional levels.

For example, birds, with their cross-current gas exchange lungs, would be less susceptible to suffer oxygen limitation (Pörtner, 2002), a characteristic that could limit the consequences of mitochondrial gene variation. In fact, the oxygen affinity of the aerobic metabolism (ETS and OXPHOS) indicates that the maximum exercise limits is determined by the oxygen supply in mammals (Gnaiger, Lassnig *et al.*, 1998). If aerobic capacities are determined by oxygen availability in mammals, we could suspect the same limitation in groups having limiting respiratory systems like fishes, amphibians, and reptiles. Therefore, ETS and OXPHOS would have evolved to insure an optimal mitochondrial respiration in condition of weak oxygen supply in the cell. It is consequently reasonable to suppose that relaxed pressure in bird's mtDNA and physico-chemical properties of mtDNA-encoded peptides might be related to improved supply at the mitochondrial level which allowed a greater structural and functional plasticity.

NADH dehydrogenase and natural selection

Among the four complexes of the electron transport system (ETS), the complex I (NADH dehydrogenase) and the complex IV (cytochrome c oxydase) represent the sites that catalyze electron transfer associated to the highest potential (E°') changes (Scheffler, 1999), which is a indicator of reaction irreversibility. Therefore, their central role in the

reactions of the ETS lets us suspect a key physiological importance. In spite of the fact that the three mitochondrial subunits of the complex IV have not been identified as being under negative selection for their physico-chemical properties, their physico-chemical variations appear to be limited by their low rate of evolutive changes (Saccone, De Giorgi *et al.*, 1999; Saccone, Barome *et al.*, 2002). Moreover, their mean tolerance to physico-chemical variability is weak (table 5). For the complex I, the majority of mitochondrial subunits have been identified as genes under negative selection (table 4). Considering the great potential (E°') changes and considering that large changes in flux through the ETS pathway are sustained with minimal changes (Hochachka, McClelland *et al.*, 1998), we could suppose that these are key complexes in ETS. Even with a few differences in the way used, they conserved their properties to insure mitochondrial function.

Unfortunately, the structure of the complex I is not well known (Yoshikawa, Shinzawa-Itoh *et al.*, 1999), and thus, there are many difficulties suggesting mechanisms responsible for the observed pattern. For example, as noted earlier, ND4L has a hydropathic value in mammals significantly higher than in other vertebrates ($t = 19.079$; $P = 0.000$). Why did mammals conserve this particularly high hydrophobicity in ND4L? If specific lineage evolution contributes to generate a pattern whose existence does not disturb any other critical processes that previously worked, and whose presence do not too heavily affect fitness (Meléndez-Hevia, Waddell *et al.*, 1996; Das, Miller *et al.*, 2004), one could suspects that this "new" metabolic state, shared by every mammals, appeared by positive selection. It is tempting to suggest a key evolutionary pattern fundamental to mammals.

However, since we are in the presence of a overlapped, small, and quite variable mitochondrial gene, and since ND4L's hydrophobic amino acid composition could be maintain by genetic drift in the case of relaxed constraints, the question remains open. Even if mammal's ND4L composition is significantly different, it is worth noting that ND4L tolerance to physico-chemical variability is not different from that of other vertebrates or other NADH dehydrogenase subunits.

Since we evaluated both amino acid distances and physico-chemical properties (hydropathic indexes) from amino acid composition, in a strictly neutral perspective, we would not suspect any evolutionary processes. It would rather be interpreted as a simple platitude, i.e. the more evolutionarily variable a gene is, the more physico-chemically variable it will be. Deviations from the model's prediction would consequently not be possible. However, we observed that deviation from the model's prediction is quite usual, suggesting that species have to maintain mtDNA-encoded peptides in a structural state that is compatible with their function. Moreover, in a strict neutral perspective, it would be unlikely to envisage model's prediction differences, but the strong relationship between physico-chemical variability and amino acid distances is variable among taxa, like shown for birds. Negative selection would then determine in great part the tolerance to physico-chemical changes, thus, the model's prediction. Since strict neutrality is rare in biological systems, our model could therefore represent a quasi-neutralist view that supposes that, since most of the mutations are eliminated in the course of evolution through negative selection, the observed substitutions are maintained because they insure a structural state

compatible with mitochondrial respiration, and proper to each vertebrate group. However, it would be really interesting to compute a strictly neutral model of physico-chemical changes to make slope comparisons, and to finally and confidently conclude on the evolutionary nature of our model's predictions.

CHAPITRE 3

FUTURE INTERESTS

3.1 INTRODUCTION

Public gene banks allow scientists access to an amazing amount of genetic information. Coupled with our knowledge of the biology of species as well as the cellular machinery and its physico-chemical requirements, these banks contribute to study relationships between molecules through *systems biology*. Often as a first step, bioinformatics shed light on basic organizations of diverse molecular structures in the cell, of diverse patterns driven by mechanisms to explore further. This is a powerful way to clarify many hypothesises of genetic determinisms and paradigms in evolution, at large scales.

Thus, since we had to focus on physico-chemical parameters to build the model that predicts the capacity of proteins to suffer hydropathic changes without affecting functional properties (chapter 2), in this chapter, we present a few patterns discovered in the big data base we have generated from the NCBI GenBank during my master's degree. These briefly

presented results represent interesting ways to investigate and better understand mitochondrial genome and its evolution. However, the following results have to be explored more deeply and more consistently. They are current and future objects of reflection.

3.2 IS THE GC CONTENT A PREDICTOR OF AMINO ACID EVOLUTIONARY RATES IN MAMMALS?

We know that evolutionary rates deduced from amino acid distances greatly depend on the capacity of peptides to suffer modifications of their physico-chemical properties. From this observation one could ask what determines this tolerance to physico-chemical variability. There are many constraints that underly these physico-chemical characteristics of peptides. Amino acid composition is the first determinant. Indeed, physico-chemical properties were defined with hydropathic indexes (Kyte & Doolitle, 1982) which were calculated from amino acid composition (see Chapter 2). Therefore, what is responsible for a particular amino acid composition? The GC content, variable among taxa (Perna & Kocher, 1995) and associated with a transversion and transition rate bias, might interfere with the amino acid composition determining physico-chemical properties. If the variability between species is partly explained by compositional biases in the mitochondrial genome (Perna & Kocher, 1995), does the GC content of a mtDNA-encoded peptides correlate to their amino acid evolution?

There is a strong relationship between the GC content of peptides encoding mitochondrial genes and the amino acid evolutionary distance in mammals ($P=0.0005$; figure 4 and table 6), suggesting that the amino acid evolution of mtDNA-encoded peptides in mammals are at least in part determined by their GC content. This is not as clear for the other vertebrate taxa studied. Indeed, while the relationship is weak but existent in amphibians, reptiles and birds with R^2 ranging from 0.310 to 0.483 (table 6), it is inexistent

in actinopterians. Moreover, even if the correlations are statistically significant in other groups, the slopes determining the amplitude of GC content changes as a function of amino acid distances are relatively weak, particularly in amphibians and reptiles (table 6). This suggests that amino acid evolutionary distances would be firstly determined by physico-chemical properties (as argued in the chapter 2), rather than by DNA properties determined by GC content which contributes to change DNA stability through an increase of hydrogen links.

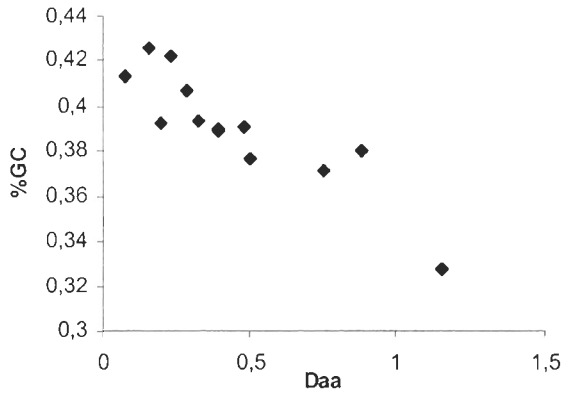


Figure 4. Percentage of GC content (%GC) as a function of mean amino acid evolutionary distances (Daa) in mammals. Each point corresponds to one of the 13 mtDNA-encoded peptide.

Table 6. Slope value compilation of the reduced major axis of GC content (%GC) as a function of amino acid distances (Daa) with their corresponding R^2 values. The model significances were obtained from *Table Curve 2D Windows 407* software.

Vertebrate taxa	Reduced major axis	R^2 value	Model significance
Actinopterigii	%GC = -0.077 Daa + 0.483	0.046	P > 0.05
Amphibia	%GC = -0.045 Daa + 0.427	0.483	P = 0.01
Reptilia	%GC = -0.018 Daa + 0.439	0.326	P = 0.05
Aves	%GC = -0.110 Daa + 0.484	0.310	P = 0.05
Mammalia	**%GC = -0.081 Daa + 0.427**	**0.803**	**P = 0.0005**

In the case of mammals, evolution of mtDNA-encoded peptides could be explained by both changes in DNA properties and tolerance to physico-chemical changes. Therefore, it seems that there is no "universal" solution to explain rate variability among vertebrate groups. The biochemical mechanisms of mutation, repair, replication, transcription of mtDNA (Perna & Kocher, 1995), as well as the RNAt pool (Scheffler, 1999; Taanman, 1999) and structural characteristics, must therefore all be taken into account as we search for the origin of the compositional bias and the mechanisms underlying mutations and molecular evolution.

3.3 MITOCHONDRIAL GENE LENGHT IS HIGHLY CONSERVED

When wondering about the effect of gene length on evolutionary rates, we decided to compare the number of nucleotides for each gene according to the vertebrate taxon. Surprisingly, gene length has been highly conserved during vertebrate evolution (figure 5). This suggest that the *race for replication* hypothesis of A.C. Wilson (Saccone, De Giorgi *et al.*, 1999) might not contribute to reduce mitochondrial genome size through an absolute gene length reduction. However, it is worth noting that genes in mtDNA can be overlapped, as ATP6 overlaps on ATP8, in the same way ND4 overlaps on ND4L (Scheffler, 1999; Taanman, 1999). Moreover, the number of RNAt molecules is reduced in the mitochondria when compared with the cytosol, containing respectively 22 RNAt and 30 to 32 RNAt (Scheffler, 1999; Taanman, 1999), as for the size of ribosomal sub-units 12S and 16S of the mitochondria which sequences are included in the 28S and 39S ribosomal sub-units of the cytosol (Rand, 2001).

It is possible that mitochondrial genes are as small as they can; they reached the smallest size possible to stay functional. Thus, does the *race for replication* refer only to a hypothetical moment when mtDNA contained more genes, more room between them, bigger ribosomes, and more RNAt molecules? Is this hypothetical process useful for the mitochondria whose products should be coordinated with nuclear products to insure aerobic metabolism? Does it represent a consistent hypothesis to explain mitochondrial genome size reduction?

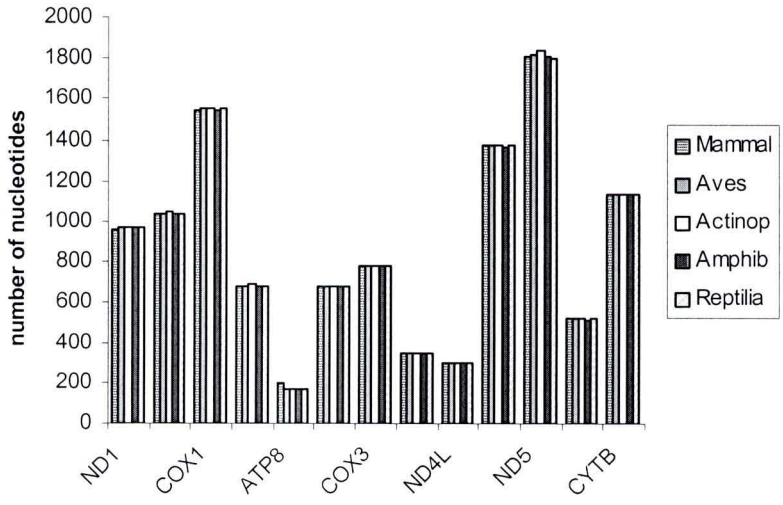


Figure 5. Gene lenght (number of nucleotides) compilation by vertebrate taxon for each mtDNA-encoded peptide.

At the moment of endosymbiosis, there might have been important redundancy between mitochondrial and nuclear genes. Consequently, many genes might have been lost (Scheffler, 1999; Rand, Haney *et al.*, 2004). The *race for replication* would then suggest why these genes were lost into the mitochondria rather than into the nucleus, at a moment where aerobic metabolism was not evolved (Rand, Haney *et al.*, 2004). However, many other insights suggest that the better capacity of the nucleus to regulate gene expression and the better protection against mutations through recombination and separation from free radicals (Saccone, Gissi *et al.*, 2000; Herrmann, 2003) might account for the mitochondrial gene transfer into the nucleus. Therefore, since the *race for replication* hypothesis seems to be used in many directions – lets say as a "luggage term" – we need to define it more precisely to know what we are talking about when using the concept.

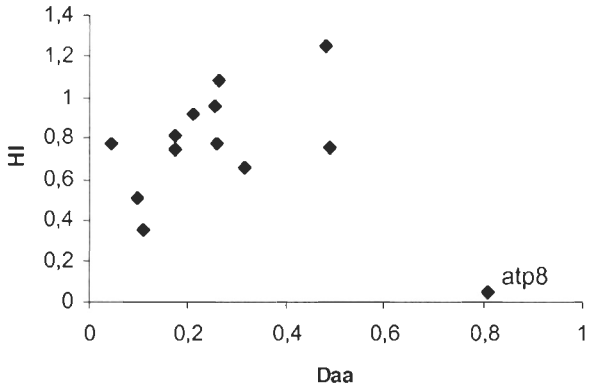
3.4 WOULD ATP8 BE ALREADY TRANSFERRED TO THE NUCLEUS?

Figure 6 (A-E) presents the mean values of the hydropathic indexes (Kyte & Doolitle, 1982) for each mtDNA-encoded peptide as a function of amino acid evolutionary distances in each of the five vertebrate taxa studied. Even if the R^2 values of the reduced major axis remain weak when excluding ATP8 (table 7), there is a general increase of hydrophobicity with evolutionary distance, statistically significant for amphibians, reptiles, and birds. This suggests that it would be easier to observe mutations driving to hydrophobic amino acids from either hydrophobic, neutral, or hydrophilic amino acids. This might be caused by a bias in the base composition of amino acids, increasing the likelihood of hydrophobic changes. Indeed, if we find that GC-rich codons encode for hydrophobic amino acids in a greater proportion than hydrophilic ones, the observed pattern would only be more likely. This hypothesis is supported by the fact that ATP8, excluded from the correlation, has the lowest GC content when compared to other mtDNA-encoded peptides in each of the five vertebrate groups studied. However, since the R^2 remains weak, this pattern might only account for a partial effect explaining amino acid evolutionary changes. Therefore, we cannot exclude the possibility of a structural constraint.

Table 7. Slope value compilation of the reduced major axis of hydropathic indexes (HI) as a function of amino acid distances (Daa) with their corresponding R^2 values. The model significances were obtained from *Table Curve 2D Windows 407* software. ATP8 was excluded from the analysis.

Vertebrate taxa	Reduced major axis	R^2 value	Model significance
Actinopterigii	HI = 1.728 Daa + 0.386	0.292	P > 0.05
Amphibia	HI = 0.655 Daa + 0.389	0.342	**P = 0.05**
Reptilia	HI = 0.478 Daa + 0.323	0.490	**P = 0.025**
Aves	HI = 1.138 Daa + 0.334	0.331	**P = 0.05**
Mammalia	HI = 2.746 Daa + 0.235	0.347	P > 0.05

A



B

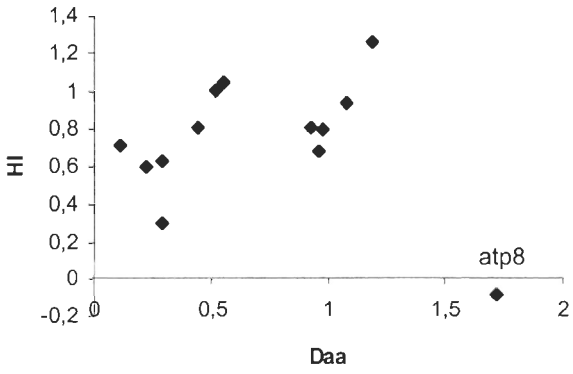
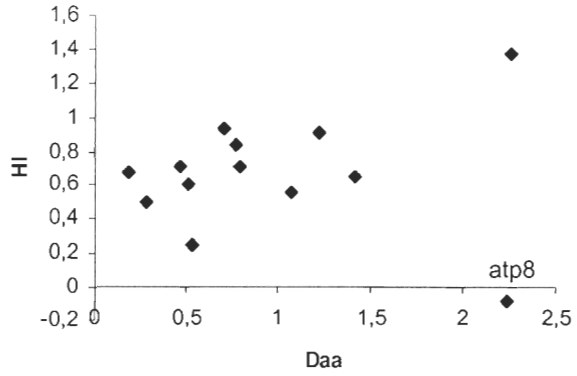
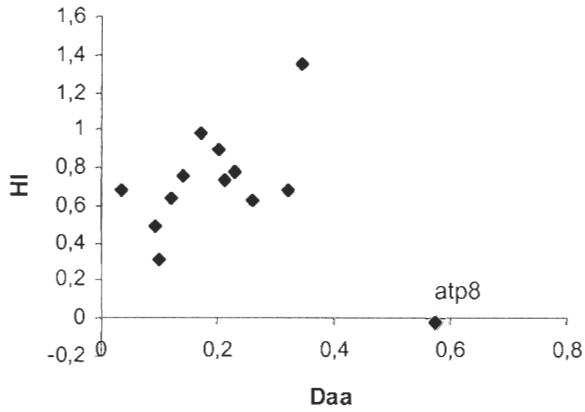


Figure 6 (A-B). Mean hydropathic indexes (HI) as a function of mean amino acid distances (Daa) in A) actinopterigii; B) amphibia. Each point represents one of the thirteen mitochondrial protein coding genes.

C



D



E

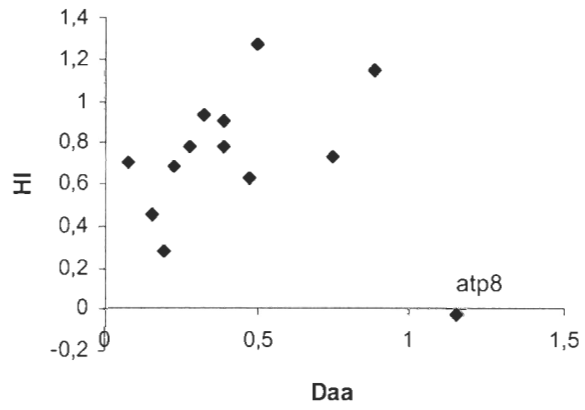


Figure 6 (C-E). Mean hydropathic indexes (HI) as a function of mean amino acid distances (Daa) in C) reptilia; D) aves; E) mammalia. Each point represents one of the thirteen mitochondrial protein coding genes.

Even if we could explain the singular physico-chemical properties of ATP8 by its low GC content, the low hydrophobicity of ATP8 remains surprising because it encodes a peptide that is part of the F_0 proton channel. This protein complex forms a transmembrane H^+ channel that participates in proton transduction across the mitochondrial inner membrane (Bianchet, Pedersen *et al.*, 1999; Souid & Penefsky, 1999). However, it is still unclear whether ATP8 is directly in contact with the mitochondrial membrane or in contact with other proteins of the F_0 channel. Moreover, it seems that the other mtDNA-encoded peptides have important catalytic functions (Papa, Guerrieri *et al.*, 1999). Therefore, while ATP8 would only constitute a part of the door through which protons are crossing back to the mitochondrial inner membrane to synthesize ATP, a role where the impact of mutations may be minimized, the catalytic role of other mitochondrial peptides encoded by mtDNA might limit their tolerance to physico-chemical changes.

A low content of hydrophobic amino acid residues is a significant factor in determining whether a protein is folded or unfolded (Uversky, Gillespie *et al.*, 2000). If ATP8 is natively unfolded, the lack of apparent structural order of the protein would explain this particular pattern. Moreover, a disorder-order transition induced in natively unfolded proteins, during the binding of specific targets *in vivo*, might represent a mechanism for the regulation of numerous cellular processes, thus suggesting an important role even without observable structure *in vitro* (Uversky, Gillespie *et al.*, 2000). Although a lot of descriptive work still needs to be done on structural characteristics of mitochondrial

peptides, would a natively unfolded and hydrophilic peptide be possible in the case of a structural peptide?

If ATP8 is not a natively unfolded peptide, since this small and variable gene is overlapped by ATP6, and since more hydrophilic mtDNA-encoded peptides would have been more susceptible to be transferred to the nucleus (Liò & Goldman, 2002; Herrmann, 2003), could we suspect ATP8 to be already transferred to the nucleus? To test this hypothesis, we tried to find the ATP8 of the *Homo sapiens* mitochondria in its nuclear genome with the GenBank *BLAST*, the Basic Local Alignment Search Tool. A few sequences producing significant alignments were found with significant E values (table 8). The Expect value (E or E value) is a parameter that describes the number of hits one can expect to see just by chance when searching a database of a particular size. The lower the E values, weaker are the chance to match with a similar sequence just by chance (http://www.ncbi.nlm.nih.gov/blast/blast_FAQs).

The problem is to identify the correct sequence corresponding to ATP8. Indeed, even if there are sequences with only a few differences from the mitochondrial ATP8, we could suspect greater differences between these two sequences because the nuclear one might be more conserved than the mitochondrial one. We should therefore analyse the nature of the sequences to identify the protein structure of gene sequences. In comparing secondary and tertiary structure of *in vivo* ATP8 peptides with the secondary and tertiary structure predicted from the sequence alone, we could perhaps eliminate the sequences that

do not correspond to the expected functional ATP8. Then, we might have tools to analyse and compare gene expression and product which could be useful to conclude on this hypothesis.

Table 8. *BLAST* results including the 10 most significant alignments and the ATP8 from the mitochondrial genome of *Homo sapiens*. E values < 0.01 are nearly identical to P values, thus near zero.

Sequence localization	**Sequence reference number**	**E value**
Mitochondria (ATP8)	ref|NC_001807.4|	2e-30
Chromosome 2	ref|NT_022184.14|Hs2_22340	3e-29
Chromosome 1	ref|NT_077913.3|Hs1_77982	1e-28
Chromosome 10	ref|NT_030059.12|Hs10_30314	1e-20
Chromosome 17	ref|NT_024862.13|Hs17_25018	5e-18
Chromosome 5	ref|NT_034772.5|Hs5_34934	1e-11
Chromosome 6	ref|NT_007299.12|Hs6_7456	2e-11
Chromosome 2	ref|NT_022135.14|Hs2_22291	5e-11
Chromosome 7	ref|NT_079592.1|Hs7_79657	1e-10
Chromosome 7	ref|NT_023629.12|Hs7_23785	1e-10
Chromosome 7	ref|NT_086709.1|Hs7_86393	5e-10

3.5 PHYLOGENETICAL TREES OF MAMMALS' EVOLUTION

In the linear array of amino acid sequence of a peptide, there are changes that provide a record of evolutionary history (Golding & Dean, 1998). The three-dimensional structure is a morphology directly related to function (Golding & Dean, 1998; Petsko & Ringe, 2004). Then, molecular changes that modify the three-dimensional structure correspond to phenotype changes eventually detectable as phylogenetic signals. However, a gene sequence itself sends a phylogenetic signal without any corresponding phenotypic effect, measurable in morphology.

Morphological and molecular phylogenetic signals are often conflicting messages. It is well known that the topologies of phylogenetic trees made from molecular data are highly dependant of chosen genes, thus highly variable (Liò & Goldman, 2002). This is indeed problematic when trying to resolve a congruent taxonomic pattern relating species to each other according to the "real" evolutionary history. Even with these difficulties, we should not ignore any kind of information. The degree to which a topology can be confirmed should depend on the use of character congruence in the search for the best-fitting hypothesis for all of the available characters (Eernisse & Kluge, 1993). In brief, it should follow the total evidence paradigm (Kluge, 1989; Eernisse & Kluge, 1993; Kluge, 1998; Kluge, 2004). However, it is worth noting that the phylogenetic data presented here corresponds to molecular data referring to physiological characters rather than morphological ones. Thus, the evolutionary lability of physiological characters and their

propensity to converge according to the species and environmental constraints is expected to send still another kind of phylogenetic message.

Therefore, this phylogenetic information would be used to question the processes underlying physiological patterns rather than to try to resolve a phylogeny that would be a good hypothesis of the real evolutionary history. Thus, if characters or genes relate species to each other in different ways, this simply suggests that evolution of such characters are in part determined by phylogenetic inertia – the linear history where changes are accumulating (Blomberg & Garland, 2002; Blomberg, Garland *et al.*, 2003) – and in part by the specific constraints lying on the features of the character which is confronted to different conditions depending on the species biogeography, biology, behaviour, etc. For example, why are Primates generally well grouped in the 13 trees from the 13 mtDNA-encoding genes but always excluding *Lemur Catta*? Why are Rondentia well grouped for COX3 but exclude *Myoxus glis* in COX1 and COX2 trees? Are these genes subject to different environmental or behavioural constraints that could explain these topologies, or do they correspond to inconsistencies of the classification into an order? [...] The trees was made from amino acid sequences of the 13 mtDNA-encoding genes in mammals, with *PAUP 4.0*, using the Neighbor Joining calculation (figure 7 et table 9).

Finally, the different topologies observed when comparing trees made from different genes might firstly come from the fact that the evolutionary rate is gene dependant (see chapter 2). Indeed, for a same couple of species, the sequence of a gene evolving

rapidly is not confronted to the same restrictions than a gene evolving slowly. The former might be subjected to multiple hits leading to an underestimation of the evolutionary distance, and the later, to a too weak phylogenetic signal. Moreover, the base composition and the transversion-transition rates being variable among genes and species, implies different model parameters that might not correspond to every species involved (Grant & Kluge, 2003). Since many species are analysed at the same time in a tree, each species taken separately may have different constraints, thus different topologies, complexifying the tree construction. Therefore, taking trees from molecular data to find interesting topologies that might hide interesting processes would help draw inconsistencies of certain characters that should not be used in phylogenetic studies.

Table 9. List of the 56 species and their order included in the presented phylogenetic trees, with their corresponding abbreviations. For the number accession of mitochondrial amino acid sequences used, see chapter 2, table 2.

ORDER	SPECIES	ABBREVIATION
Artiodactyla	Hippopotamus amphibius	Ar1
	Muntiacus crinifrons	Ar2
	Muntiacus reevesi	Ar3
	Ovis aries	Ar4
	Sus scrofa	Ar5
Carnivora	Eumetopias jubatus	Ca1
	Felis catus	Ca2
	Halichoerus grypus	Ca3
	Ursus americanus	Ca4
	Ursus maritimus	Ca5
Cetacea	Balaenoptera musculus	Ce1
	Balaenoptera physalus	Ce2
	Physeter catodon	Ce3
Chiroptera	Artibeus jamaicensis	Ch1
	Chalinolobus tuberculatus	Ch2
	Pteropus dasymallus	Ch3
	Pteropus scapulatus	Ch4
Dermoptera	Cynocephalus variegatus	De1
Didelphimorphia	Didelphis virginiana	Di1

Diprotodontia	Macropus robustus	Dp1
	Trichosurus vulpecula	Dp2
	Vombatus ursinus	Dp3
Edentata	Dasypus novemcinctus	Ed1
	Tamandua tetradactyla	Ed2
Hyracoidae	Procavia capensis	Hy1
Insectivora	Echinosorex gymnura	In1
	Erinaceus europaeus	In2
	Hemiechinus auritus	In3
	Soriculus fumidus	In4
	Talpa europaea	In5
Lagomorpha	Lepus europaeus	La1
	Ochotona collaris	La2
	Oryctolagus cuniculus	La3
Monotremata	Ornithorhynchus anatinus	Mo1
	Tachyglossus aculeatus	Mo2
Peramelemorphia	Isoodon macrourus	Pe1
Perissodactyla	Ceratotherium simum	Pr1
	Equus asinus	Pr2
	Equus caballus	Pr3
	Rhinoceros unicornis	Pr4
	Tapirus terrestris	Pr5
Pholidota	Manis tetradactyla	Ph1
Primates	Cebus albifrons	Pm1
	Gorilla gorilla	Pm2
	Lemur catta	Pm3
	Papio hamadryas	Pm4
	Pongo pygmaeus (abelii)	Pm5
Proboscidae	Elephas maximus	Po1
	Loxodonta africana	Po2
Rodentia	Mus musculus	Ro1
	Myoxus glis	Ro2
	Rattus norvegicus	Ro3
	Sciurus vulgaris	Ro4
	Volemys kikuchii	Ro5
Sirenia	Dugong dugon	Si1
Tubulidentata	Orycteropus afer	Tu1

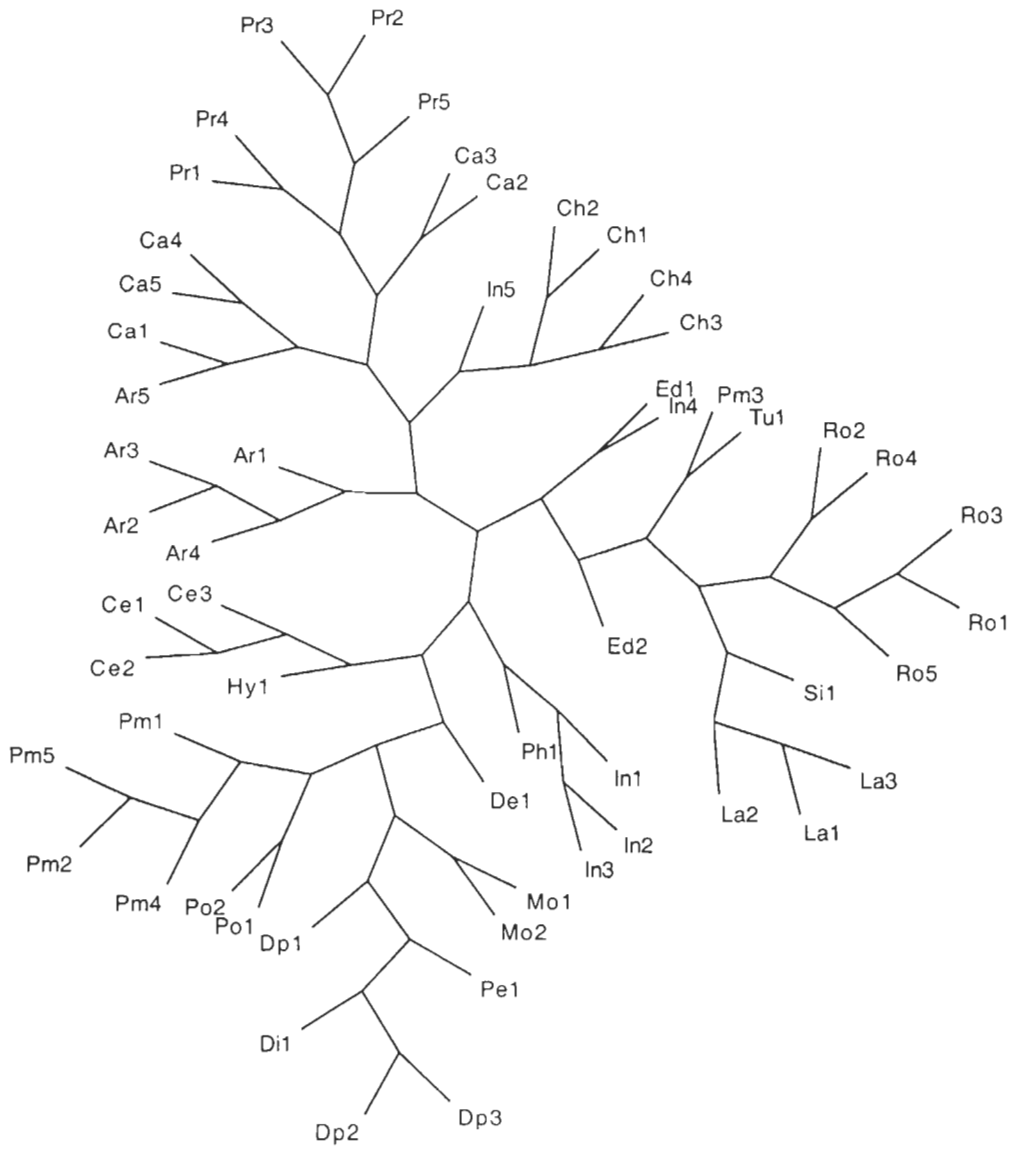


Figure 7a. Phylogenetic tree made from mammals' ATP6 amino acid sequence.

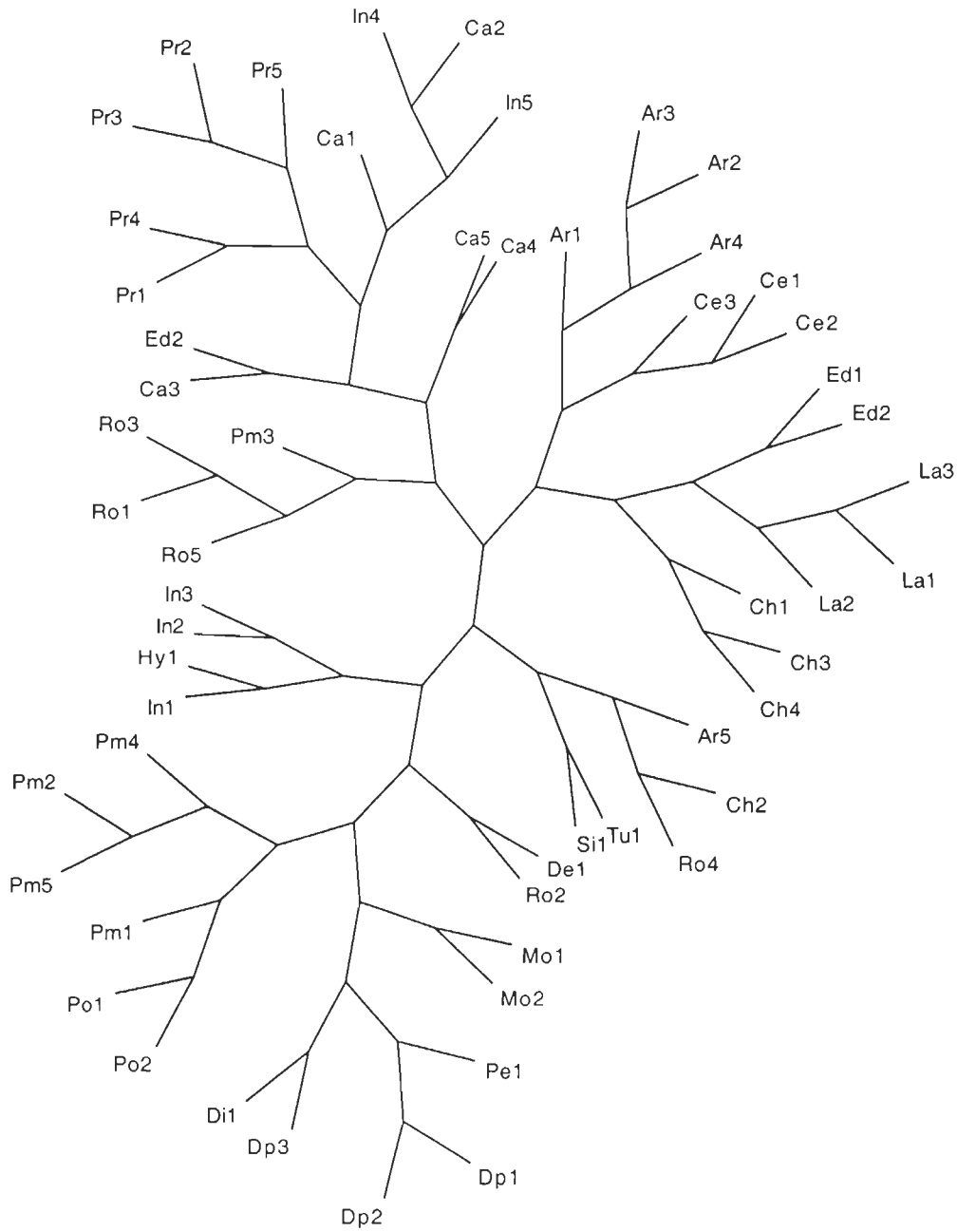


Figure 7b. Phylogenetic tree made from mammals' ATP8 amino acid sequence.

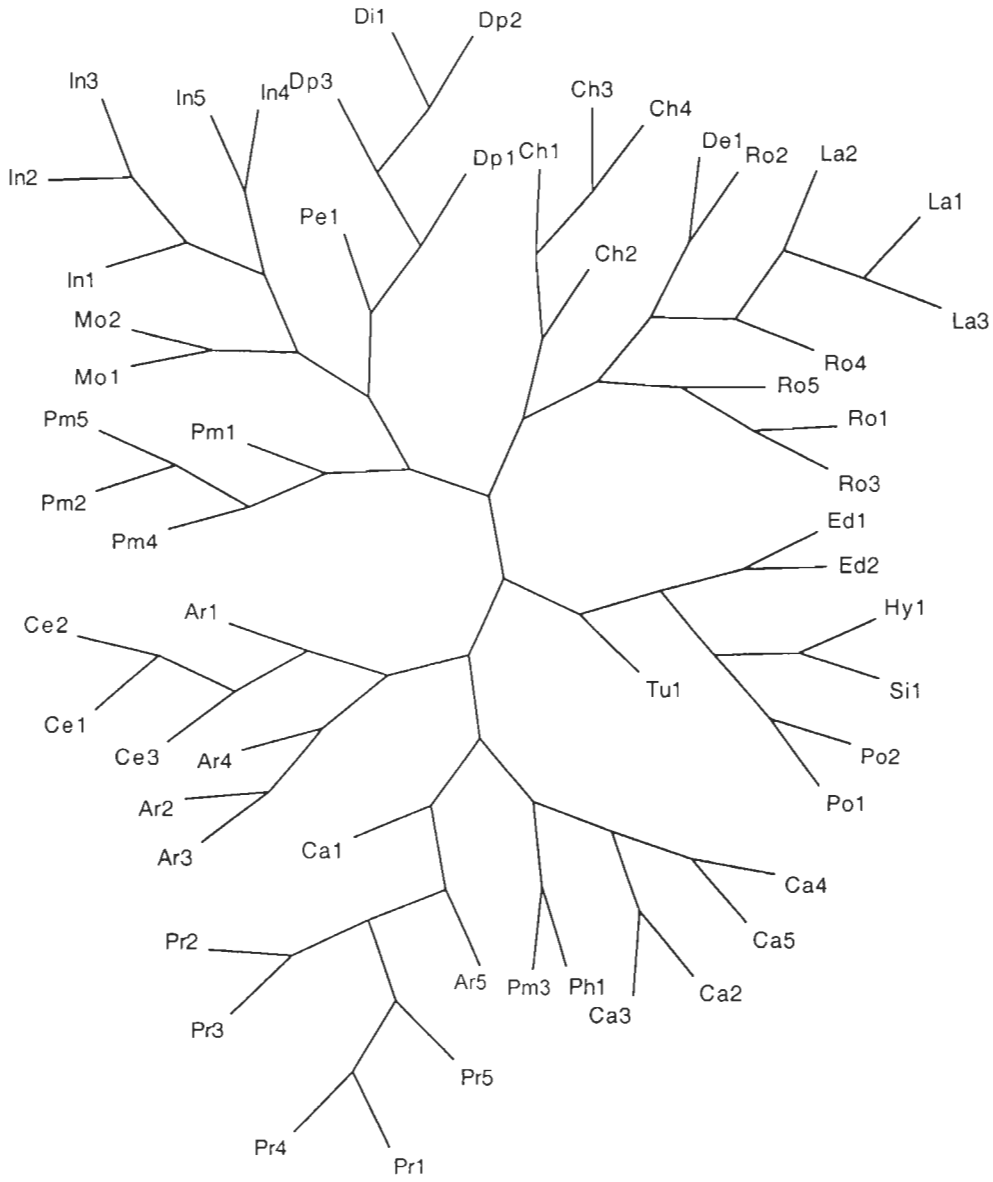


Figure 7c. Phylogenetic tree made from mammals' COX1 amino acid sequence.

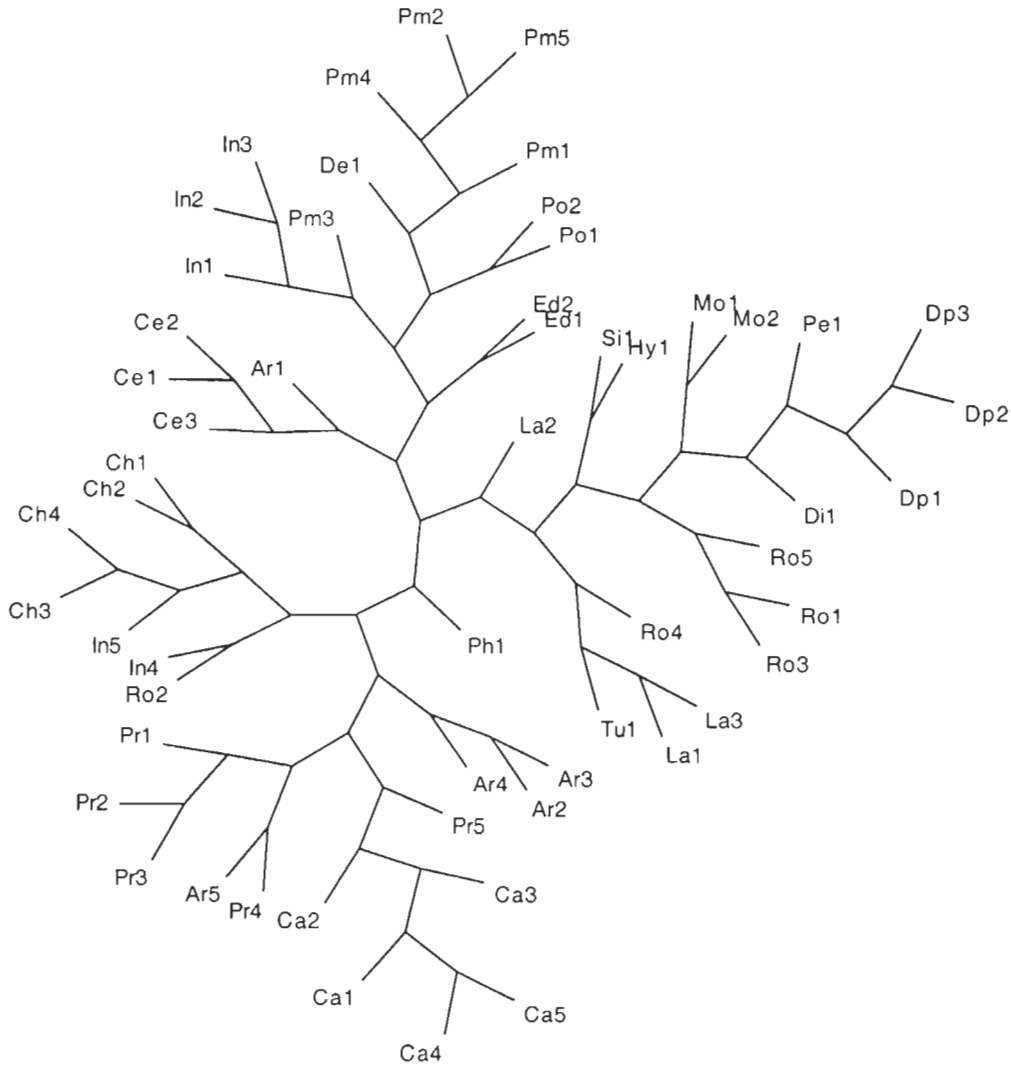


Figure 7d. Phylogenetic tree made from mammals' COX2 amino acid sequence.

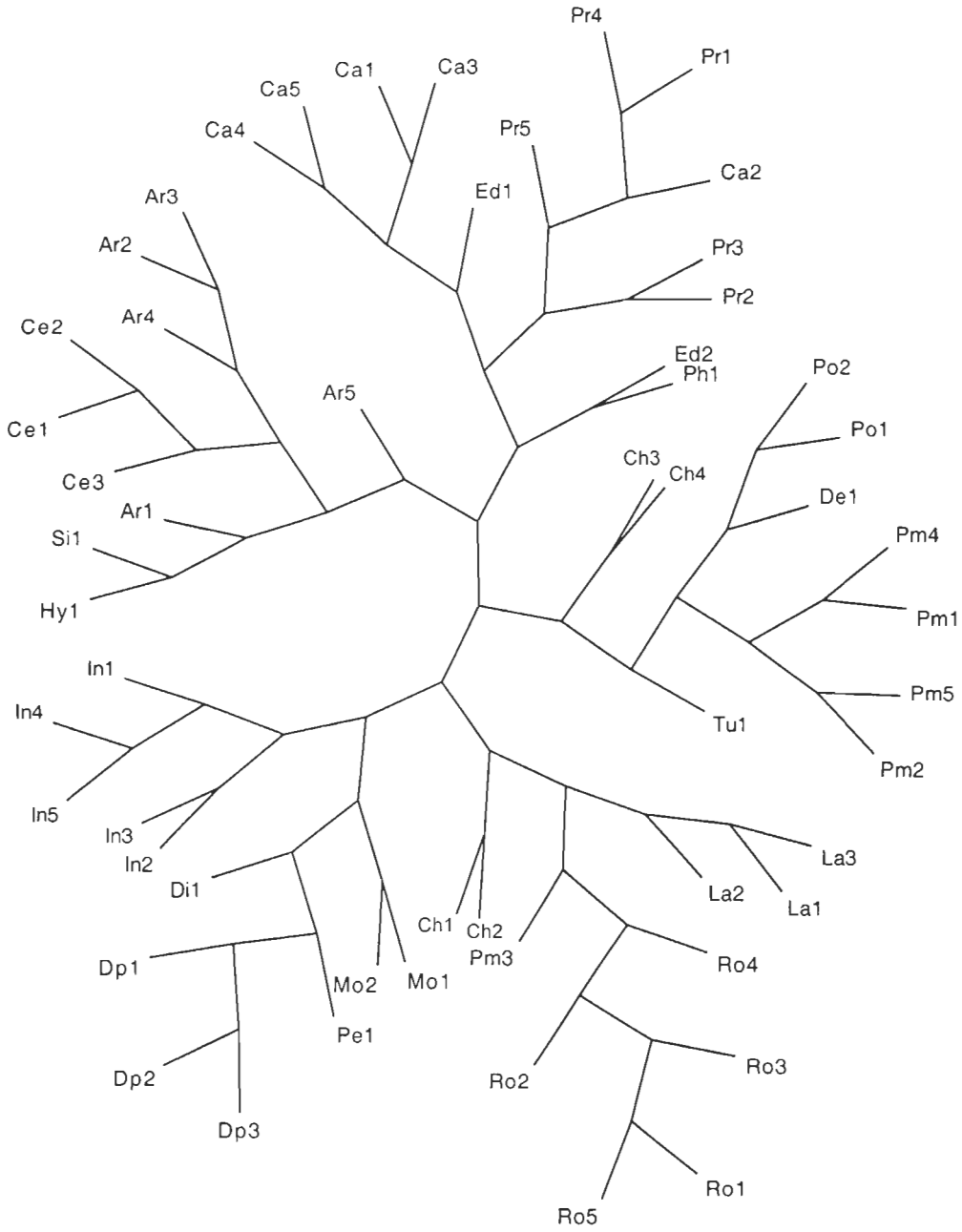


Figure 7e. Phylogenetic tree made from mammals' COX3 amino acid sequence.

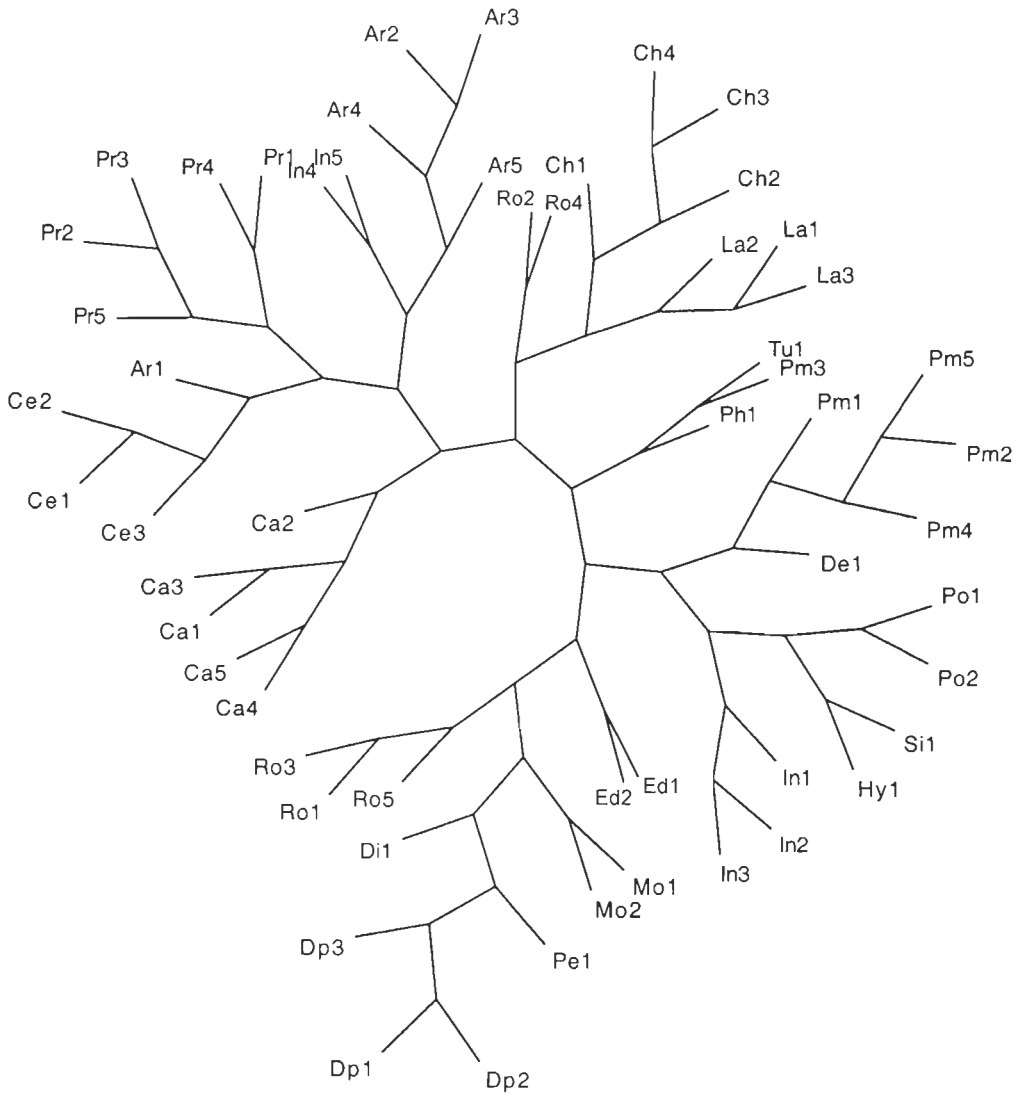


Figure 7f. Phylogenetic tree made from mammals' CYTB amino acid sequence.

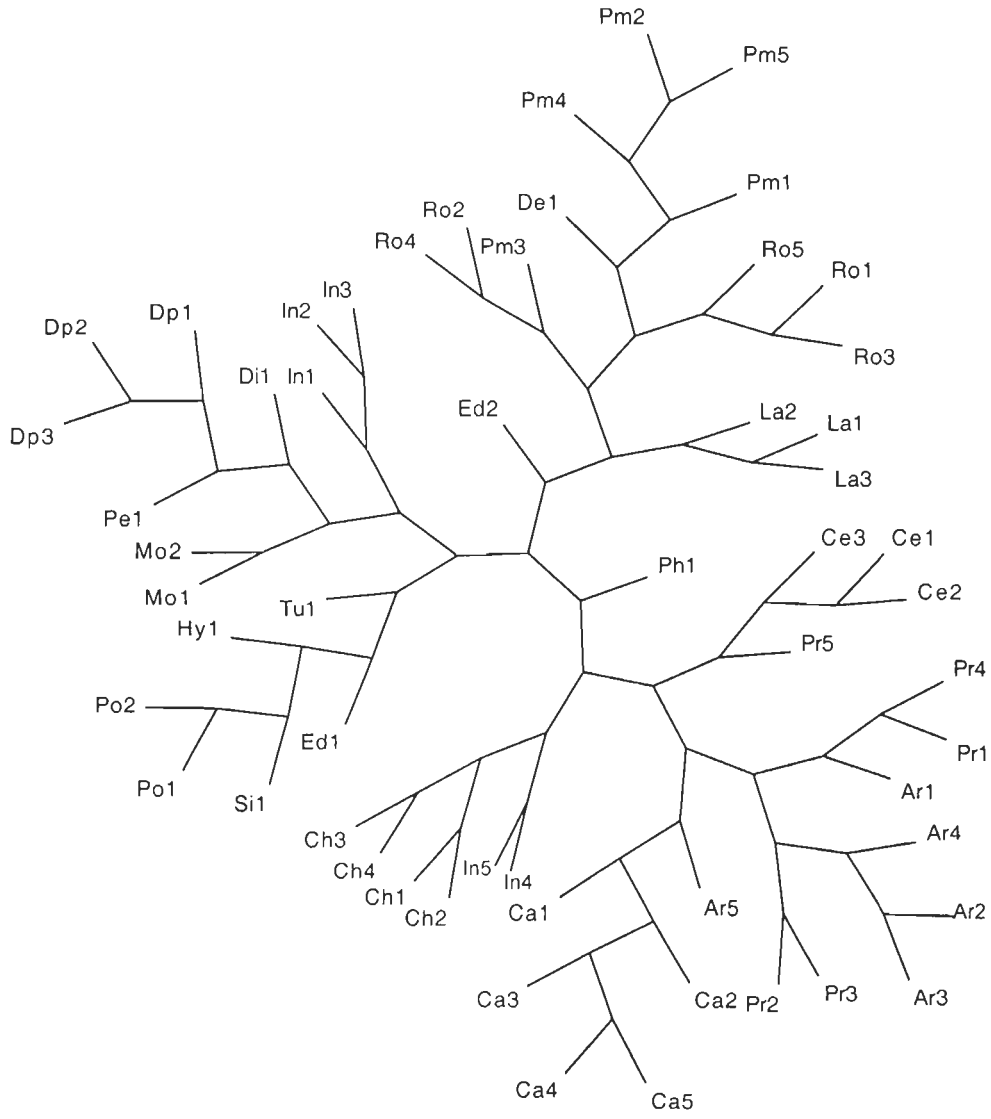


Figure 7g. Phylogenetic tree made from mammals' ND1 amino acid sequence.

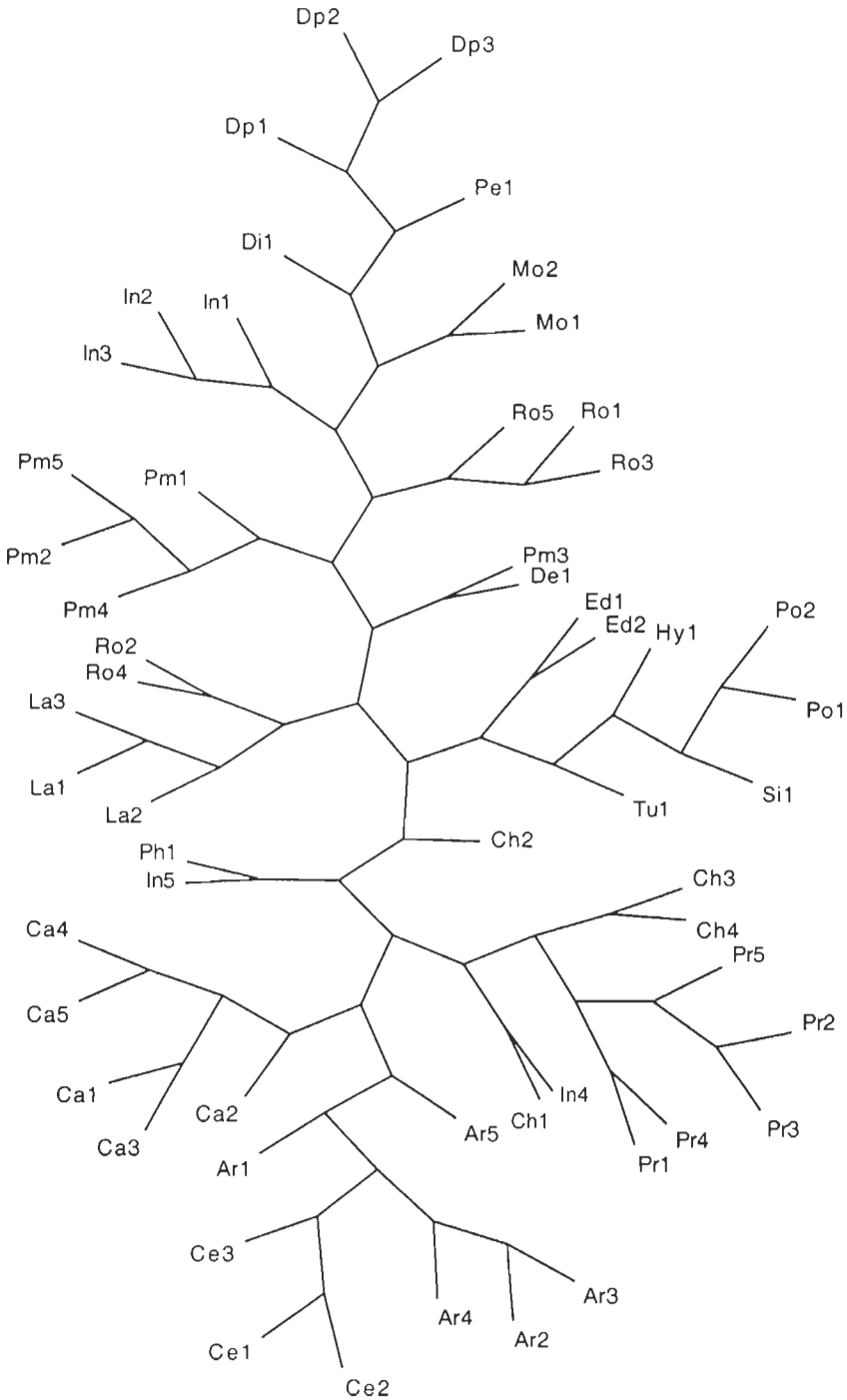


Figure 7h. Phylogenetic tree made from mammals' ND2 amino acid sequence.

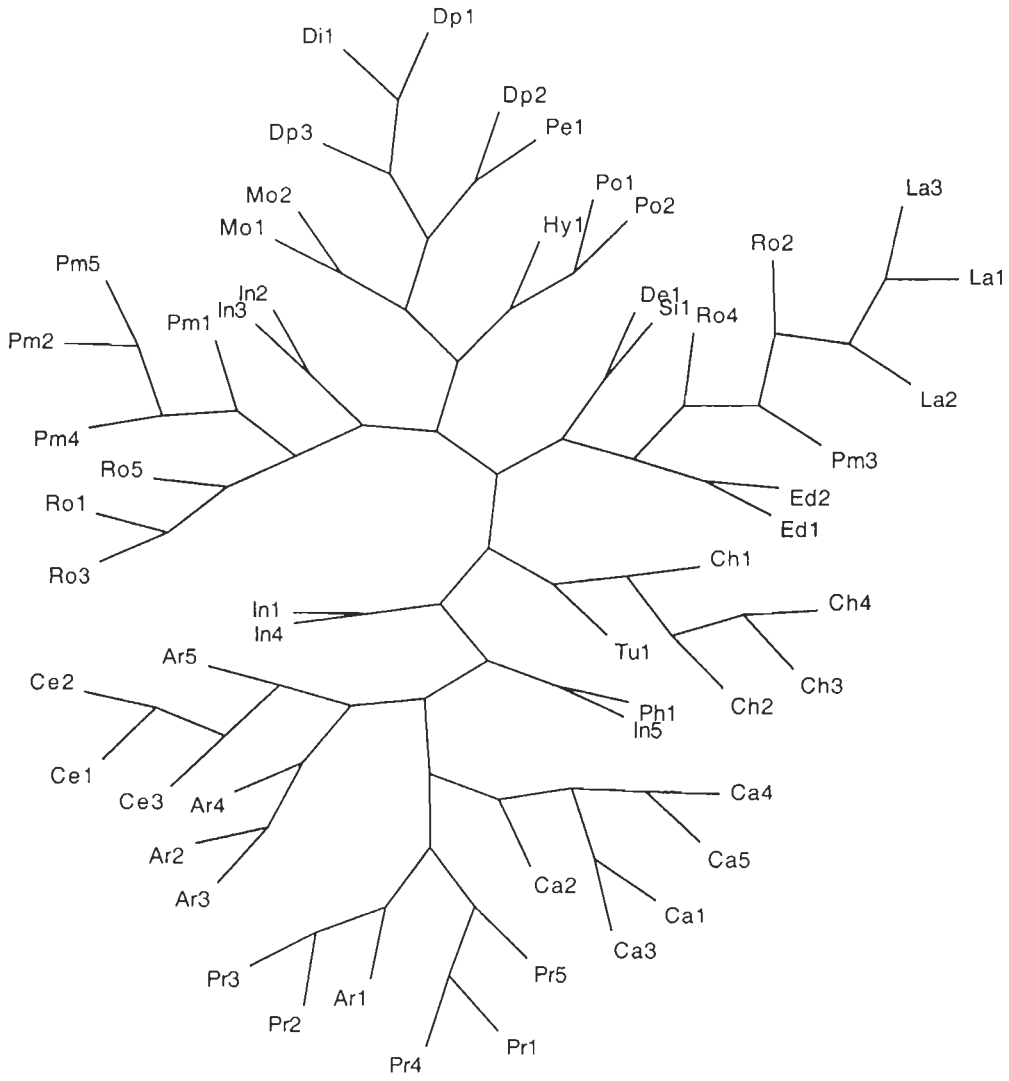


Figure 7i. Phylogenetic tree made from mammals' ND3 amino acid sequence.

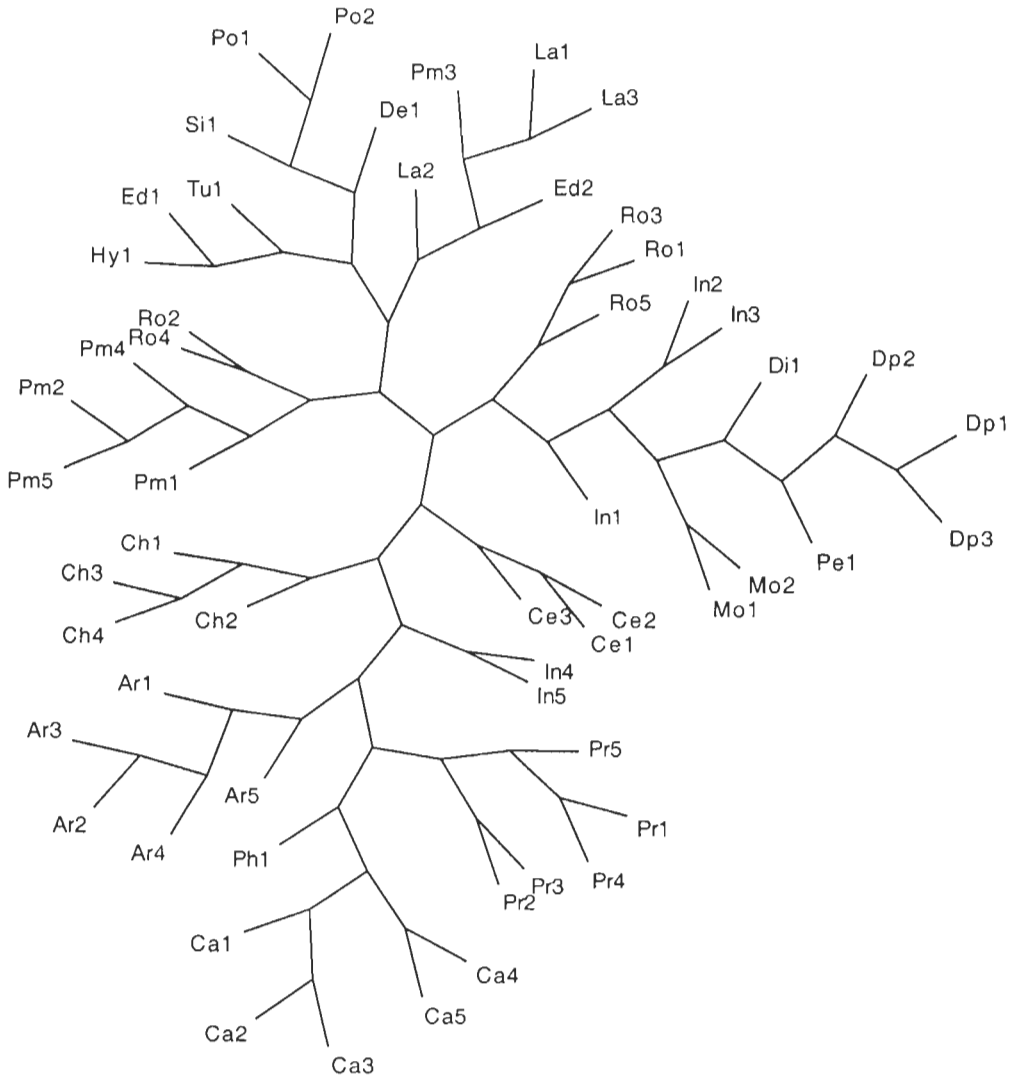


Figure 7j. Phylogenetic tree made from mammals' ND4L amino acid sequence.

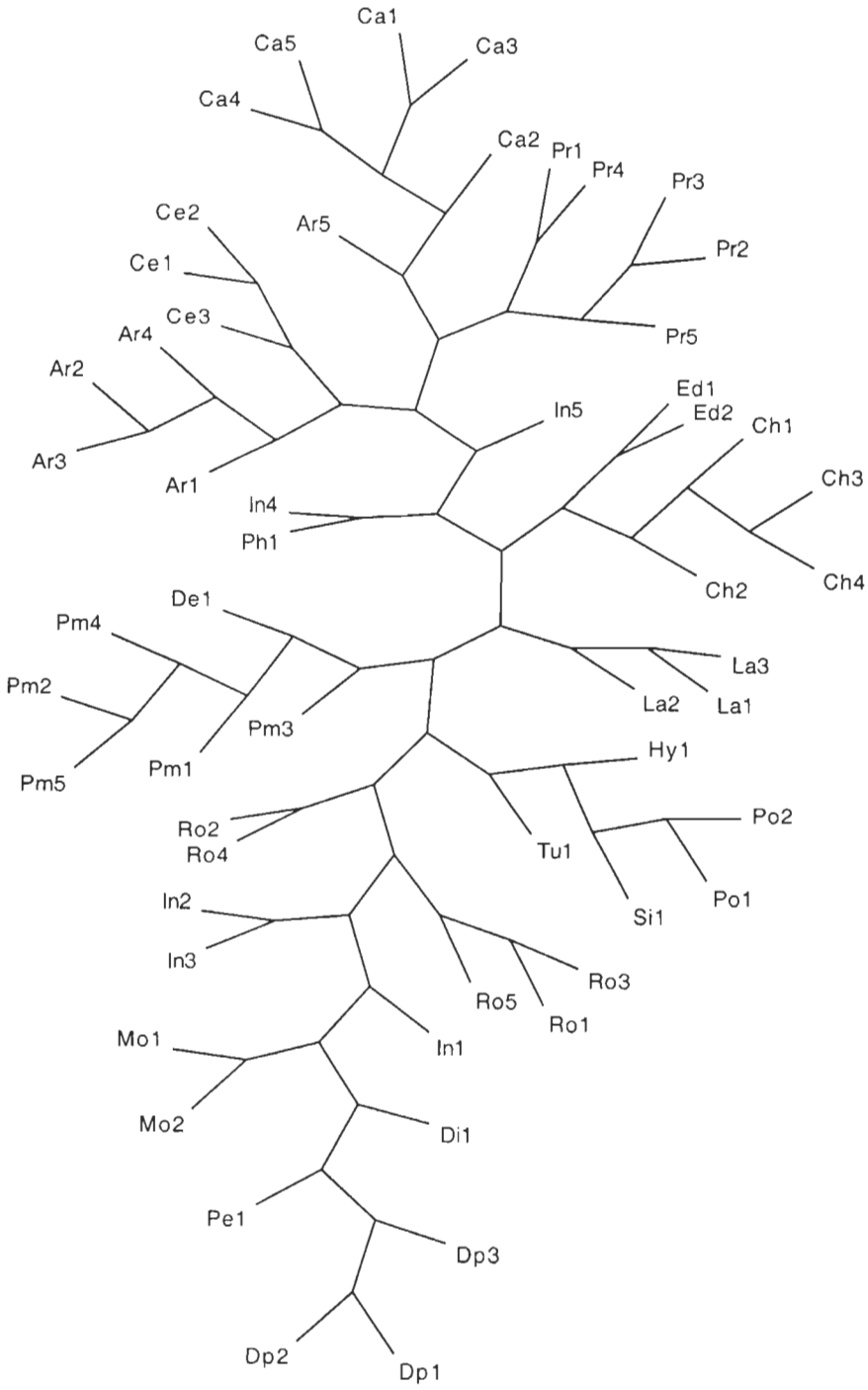


Figure 7k. Phylogenetic tree made from mammals' ND4 amino acid sequence.

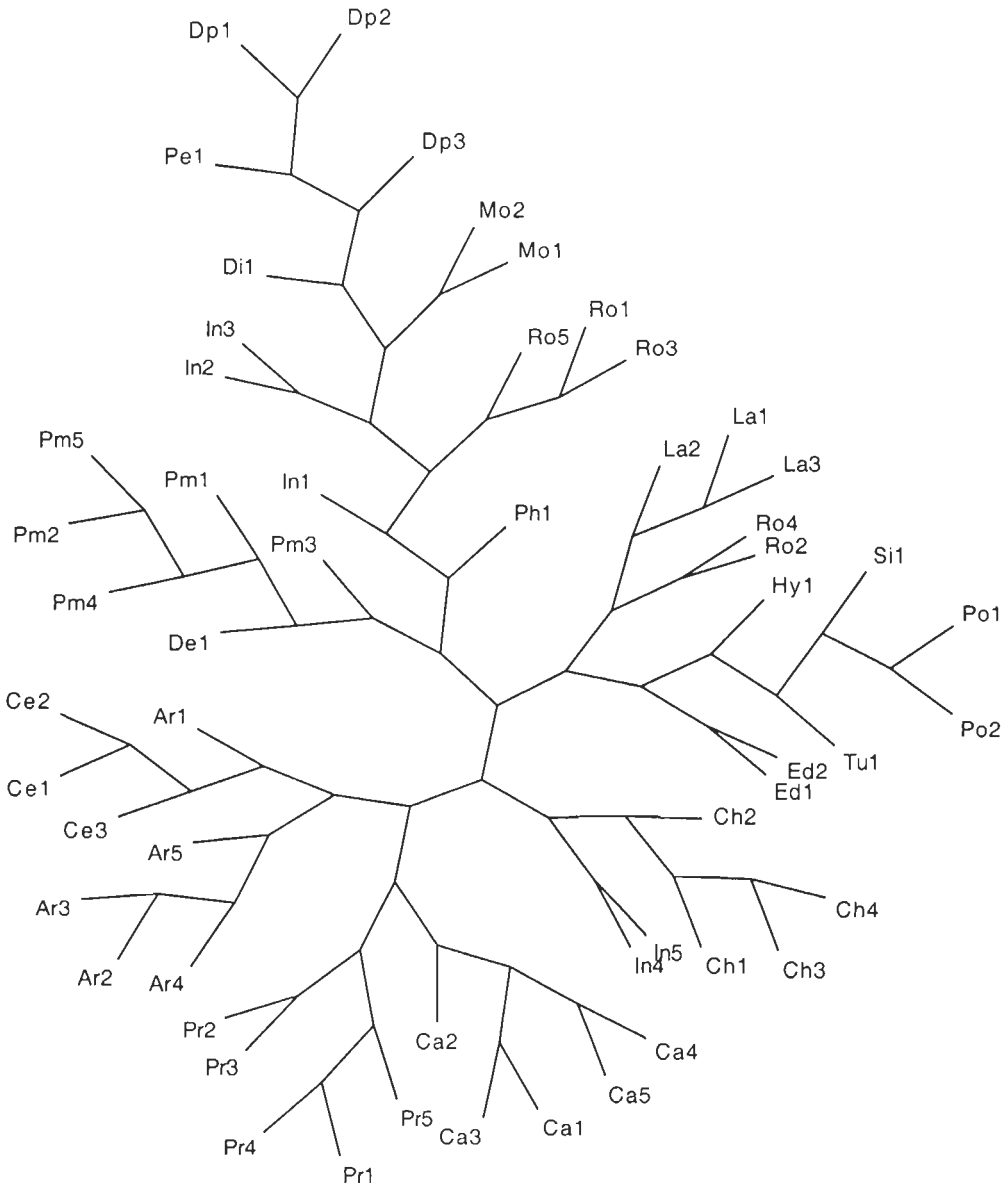


Figure 7l. Phylogenetic tree made from mammals' ND5 amino acid sequence.

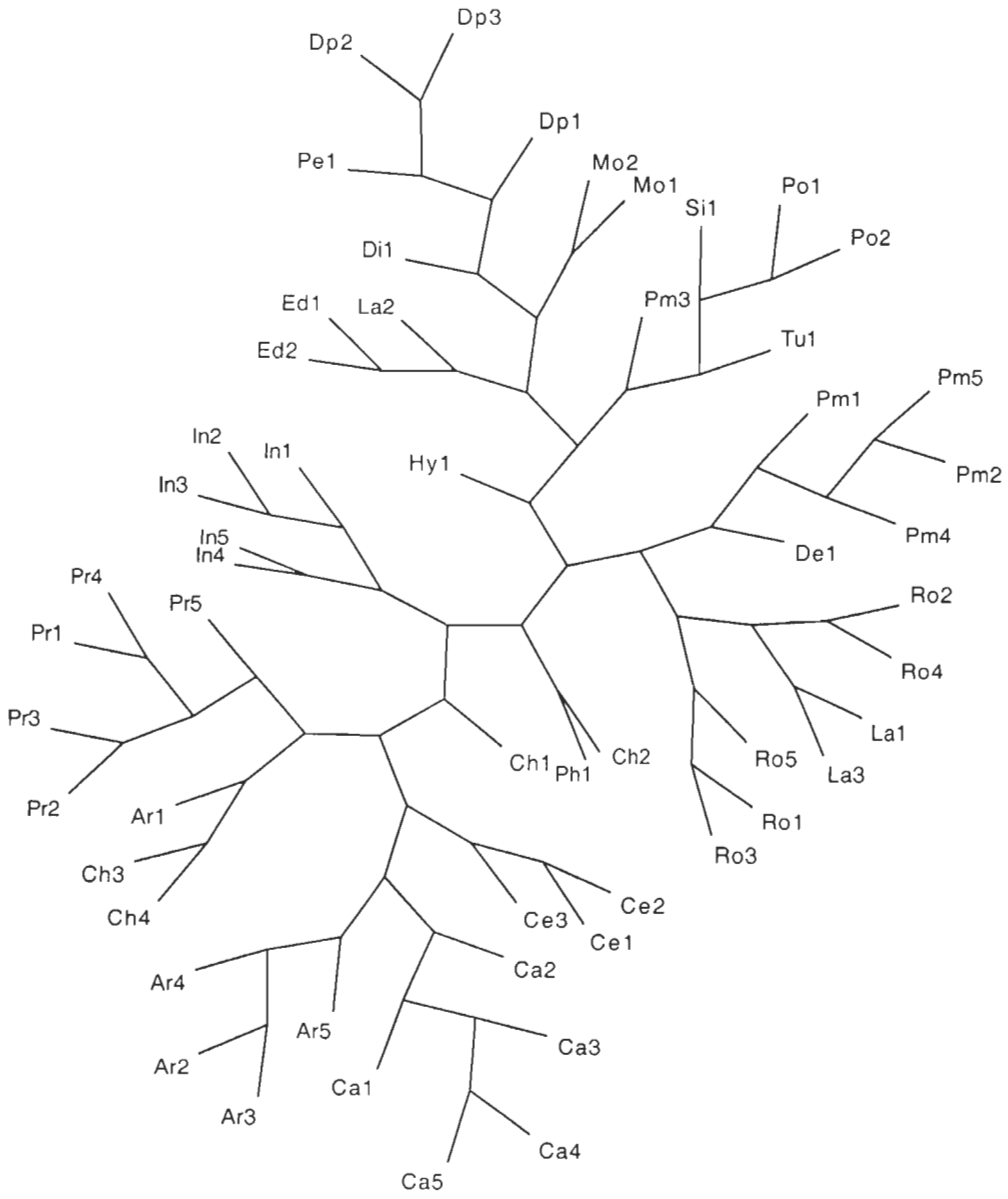


Figure 7m. Phylogenetic tree made from mammals' ND6 amino acid sequence.

Chapitre 4

Conclusion

L'étonnante diversité et quantité d'information génétique disponible dans les banques de gènes publiques ouvre l'avenir de l'étude du gène à l'étude du génome, de la génomique à la post-génomique (Fox Keller, 2004). Le génome est dorénavant un objet explicite d'investigation et de représentation (Mauron, 2001). Bien que l'intégrité des banques de gènes a pu être remise en cause dû à la présence d'erreurs de séquence impossibles à vérifier et corriger systématiquement (Harris, 2003), l'étude des grands patrons de l'évolution reste envisageable. En effet, les études à grandes échelles telle que celle présentée dans ce mémoire repose sur l'évolution de grands groupes taxonomiques où les différences entre espèces se compte souvent par centaines de mutations. Par conséquent, l'occurrence d'erreurs, bien que peu souhaitable, n'est responsable que d'une petite fraction des différences observées. De plus, une séquence fortement erronée est rapidement repérable, laquelle peut être éliminée de l'étude.

L'étude des grands patrons ordonnés des systèmes biologiques, mais aussi des inhomogénéités et singularités de ces systèmes est essentielle à leur compréhension. Elle

s'inscrit dans la post-génomique qui cherche, dans les grands patrons, questions et réponses dans l'intégration des différents facteurs composants un système. Cette vision ouverte et plus globale des gènes au sein des cellules, elles-mêmes composantes d'un organisme, est essentielle à une meilleure compréhension du monde vivant et de son évolution. De même, elle justifie tout le travail de séquençage fait jusqu'à maintenant qui, pour certain, n'a fait qu'accumuler les données sans développer une perspective de compréhension élargie du phénomène de la vie. Bref, c'est donner un sens à ce que nous nommons gène.

On peut douter, encore une fois, qu'une meilleure compréhension des structures moléculaires – ADN, protéines, etc. – et de leur évolution ne peut se faire par la recherche de modèles simples. En effet, de l'avis de certain, « l'essence d'un processus ne doit plus être recherchée dans des lois abstraites ou simples, mais dans la spécificité confuse d'adaptations particulières qui ont vu le jour au cours des processus désordonnés de l'évolution » (Fox Keller, 2004). La complexité des processus comme la vie repose certes dans les détails d'un foisonnement apparemment désorganisé de multiples éléments, toutefois, la science traite d'abord et avant tout de généralités (Bak & Paczuski, 1995). Le monde vivant est peut-être le domaine de la plus grande complexité, mais est-il pour autant moins intelligible que le monde étudié par les mathématiques ou la physique? Bien que cette question, notamment formulé par Einstein, est centrale pour l'épistémologiste : « l'éternellement incompréhensible à propos du monde est sa compréhensibilité », tenter de comprendre la nature, c'est en quelque sorte arriver à la comprimer en hypothèses simples, concises et élégantes (Chaitin, 2003).

Par conséquent, je crois qu'on ne construit pas un modèle lorsque nous avons compris le système, mais plutôt pour nous aider à comprendre la réalité que nous tentons de nommer. La modélisation nous aide à étoffer nos concepts et à les lier les uns aux autres, et c'est dans cette optique que j'ai construit cette recherche. L'idée largement partagée que la science doit commencer par le plus simple pour en arriver au plus complexe, que plus on complique un modèle, plus il sera susceptible de décrire la réalité telle quel est (Grant & Kluge, 2003), est toutefois risquée. Cette idée suggère que les obstacles vers un savoir complexe sont sous-estimés et disqualifiés, qu'ils deviennent simplement provisoires, que le progrès se chargera tout simplement de les contourner (Stengers, 1997). Toutefois, « un savoir intéressant commence toujours par le pertinent. Il commence toujours par la découverte des questions qui mettent en lumière la singularité de ce à quoi il s'adresse » (Stengers, 1997). Et en ce qui nous concerne, le pertinent, c'est comprendre l'évolution du génome mitochondrial dans toute sa complexité, et ce, par le biais de l'outil puissant qu'est l'informatique.

Si le modèle élaboré (chapitre 2) semble *a priori* simple, il rend compte de la singularité des espèces et des gènes et de leur façon de se détacher de la prédiction générale qui, dans ce contexte, décrit ce qui est considéré comme *normal*. Leur singularité n'est pas forcément prévisible ou uniforme, elle renvoie à différentes contraintes, différents contextes où baignent les espèces. C'est peut-être pour cela que, malgré mes efforts pour lier ces singularités génétiques à de multiples facteurs touchant tant la biologie des espèces que leur

comportement, leur écologie et leur taxonomie, je ne suis pas arrivée à dégager d'éléments clairs pouvant être la source de ces patrons moléculaires. Du moins, les patrons semblent dorénavant clairs; l'étude ne demande maintenant qu'à être raffinée pour trouver davantage d'indices sur les processus les ayant mis en place.

Si cette étude peut susciter des questions concernant notre façon de théoriser l'évolution du génome mitochondrial, mais aussi plus largement l'évolution biologique, si elle contribue à rassembler sous une loi des éléments apparemment épars, elle aura accompli ce par quoi un scientifique peut se réaliser; s'étonner, douter, questionner, construire ou reconstruire. Qu'il s'agissent des résultats présentés au chapitre 2 ou au chapitre 3, le processus de recherche par lequel ont émergé tous ces résultats aura généré davantage de questions que de réponses. Et c'est tant mieux! Ils représentent des voies à explorer, de nouveaux facteurs à intégrer. Ils sont devenus des outils de questionnement, des pavés dans la chaîne ou le réseau des causalités.

Le gène reste un élément déterminant dans la compréhension de l'évolution biologique. Longtemps considéré comme la base de la vie, il reprend peu à peu sa place au sein du génome et d'un ensemble de pratiques – notamment la biologie du développement – lesquelles sont toutes nécessaires pour l'avancement des connaissances dans ce domaine. Bien que la tentation est forte de considérer le génome comme l'essence du vivant, déterminant tant l'individualité que l'identité à l'espèce (Mauron, 2001; Mauron, 2002; Mauron, 2003), il est primordial de l'étudier sans lui accorder une primauté ontologique. Il

est fort peu probable que nous trouvions le nombre de mutations définissant l'appartenance à une espèce, de même qu'il est fort peu probable que le phénomène de la vie se réduise à une définition concise et intemporelle. Il reste que l'étude de l'évolution biologique répond à un besoin métaphysique en ce sens qu'elle contribue à nous situer, en tant qu'humain, dans l'enchevêtrement de ce qui constitue notre monde. L'homme a toujours cherché à se définir. On a cru pendant un moment que le gène seul pourrait répondre à cette question vertigineuse, mais en vain. Ce qui nous amène à penser que ce qui fait « l'humanité de l'homme n'est sans doute pas [exclusivement] dans sa matérialité. Si aujourd'hui des questions se posent sous un jour nouveau, ce sont des questions de toujours » (Mattei, 2005).

RÉFÉRENCES

Angilletta, J. Michael Jr., Peter H. Niewiarowski & Carlos A. Navas, "The evolution of thermal physiology in ectotherms". *Journal of thermal biology*, 2002, 27, p.249-268.

Arnold, Stevan J., Michael E. Pfender & Adam G. Jones, "The adaptive landscape as a conceptual bridge between micro- and macroevolution". *Genetica*, 2001, 112-113, p.9-32.

Audit, B., C. Vaillant, A. Arneodo, Y D'aubenton-Carafa & C. Thermes, "Wavelet analysis of DNA bending profiles reveals structural constraints on the evolution of genomic sequences". *Journal of biological physics*, 2004, 30(1), p.33-81.

Bak, Per & Maya Paczuski, "Complexity, contingency, and criticality". *Proceeding of the national academy of sciences USA*, 1995, 92(Colloquium paper), p.6689-6696.

Balon, Eugene, "A short historical prelude". *Unpublished*, 2005, Unpublished manuscript, p.1-71.

Bianchet, Mario A., Peter L. Pedersen & Mario Amzel, "Structure of F1-ATPase and the mechanism of ATP synthesis-hydrolysis". *Frontiers of cellular bioenergetics*, New York, Klumer academic/plenum publishers, 1999, p.361-376.

Blier, Pierre U., France Dufresne & Ronald S. Burton, "Natural selection and the evolution of mtDNA-encoded peptides: Evidence for intergenomic co-adaptation". *Trends in genetics*, 2001, 17(7), p.400-406.

Blier, Pierre U. & Hélène Lemieux, "The impact of the thermal sensitivity of cytochrome c oxidase on the respiration rate of Arctic charr muscle mitochondria". *Journal of comparative physiology*, 2001, 171, p.247-253.

Blomberg, Simon P. & Theodore Jr. Garland, "Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods". *Journal of molecular evolution*, 2002, 15, p.899-910.

Blomberg, Simon P., Theodore Jr. Garland & Anthony R. Ives, "Testing for phylogenetic signal in comparative date: behavioral traits are more labile". *Evolution*, 2003, 57(4), p.717-745.

Burger, Gertraud, Micheal W. Gray & Franz Lang, "Mitochondrial genomes: anything goes". *Trends in genetics*, 2003, 19(12), p.709-716.

Carroll, Robert, L., "Towards a new evolutionary synthesis". *Trends in ecology and evolution*, 2000, 15(1), p.27-32.

Chaitin, Gregory, "L'Univers est-il intelligible?" *La recherche*, 2003, 370, p.34-41.

Chong, Sy & E. Whitelaw, "Epigenetic germline inheritance". *Current opinion in genetics & development*, 2004, 14(6), p.692-696.

Creevey, Christopher J. & James O. Mcinerney, "An algorithm for detecting directional and non-directional positive selection, neutrality and negative selection in protein coding DNA sequences". *Gene*, 2002, 300, p.43-51.

Das, Jayatri, Stephen T. Miller & David L. Stern, "Comparison of diverse protein sequences of the nuclear-encoded subunits of cytochrome c oxidase suggests

conservation of structure underlies evolving functional sites". *Molecular biology and evolution*, 2004, 21(8), p.1572-1582.

De La Casa-Esperon, E. & C. Sapienza, "Natural selection and the evolution of genome imprinting". *Annual review of genetics*, 2003, 37, p.349-370.

De Ricqlès, Armand, "De la paléontologie évolutionniste à la paléontologie phylogénétique: avatars et permanence du Darwinisme". *Pour Darwin*. P. Tort, Paris, PUF, 1997, p.851-881.

Doiron, Sarah, Louis Bernatchez & Pierre U. Blier, "A comparative mitogenomic analysis of the potential adaptive value of arctic charr mtDNA introgression in brook charr populations (*Salvelinus fontinalis* Mitchill)". *Molecular biology and evolution*, 2002, 19(11), p.1902-1909.

Eernisse, Douglas J. & Arnold G. Kluge, "Taxonomic congruence versus total evidence, and Amniote phylogeny inferred from fossils, molecules, and morphology". *Molecular biology and evolution*, 1993, 10(6), p.1170-1195.

Eisenberg, David & Donald Crothers, *Physical chemistry with applications to the life sciences*. California, The Benjamin/Cummings publishing company, inc., 1979, 868 p.

Faith, Jeremiah J. & David D. Pollock, "Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomes". *Genetics*, 2003, 165, p.735-745.

Finta, Csaba & Peter G. Zaphiropoulos, "A statistical view of genome transcription?" *Journal of molecular evolution*, 2001, 53, p.160-162.

Fox Keller, Evelyn, *Le rôle des métaphores dans les progrès de la biologie*. Paris, Institut synthélabo pour le progrès de la connaissance, 1999, 157 p.

Fox Keller, Evelyn, "Génome, postgénome, quel avenir pour la biologie?" *La recherche*, 2004, 376, p.30-37.

Gerber, Anne S., Ronald Loggins, Sudhir Kumar & Thomas E. Dowling, "Does nonneutral evolution shape observed patterns of DNA variation in animal mitochondrial genomes?" *Annual review of genetics*, 2001, 35, p.539-566.

Gibson, Andrew, Vivek Gowri-Shankar, Paul G. Higgs & Magnus Rattray, "A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic". *Molecular biology and evolution*, 2005, 22(2), p.251-264.

Gnaiger, Erich, Barbara Lassnig, Andrey Kuznetsov, Gunde Rieger & Raimund Margreiter, "Mitochondrial oxygen affinity, respiratory flux control and excess capacity of cytochrome c oxydase". *Journal of experimental biology*, 1998, 201, p.1129-1139.

Golding, Brian & Antony M. Dean, "The structural basis of molecular adaptation". *Molecular biology and evolution*, 1998, 15(4), p.355-369.

Gould, S. J. & R. C. Lewontin, "The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme". *Proceeding of royal society London B*, 1979, 205, p.581-598.

Grant, Taran & Arnold G. Kluge, "Data exploration in phylogenetic inference: scientific, heuristic, or neither". *Cladistics*, 2003, 19, p.379-418.

Gu, Xun & Jianzhi Zhang, "A simple method for estimating the parameter of substitution rate variation among sites". *Molecular biology and evolution*, 1997, 14(11), p.1106-1113.

Hamada, Kazuo, Tokumasa Horiike, Shigehiko Kanaya, Hiroshi Nakamura, Hidetoshi Ota, Takayuki Yatogo, Kazuhisa Okada, Hiroshi Nakamura & Takao Shinozawa, "Changes in body temperature pattern in vertebrates do not influence the codon usages of a-globin genes". *Genes Genetic Systematic*, 2002, 77, p.197-207.

Harris, D. James, "Can you bank on GenBank?" *Trends in ecology and evolution*, 2003, 18(7), p.317-319.

Hasegawa, Masami, Ying Cao & Ziheng Yang, "Preponderance of slightly deleterious polymorphism in mitochondrial DNA: nonsynonymous/synonymous rate ratio is much higher within species than between species". *Molecular biology and evolution*, 1998, 15(11), p.1499-1505.

Hazel, Jr & Cl Prosser, "Molecular mechanisms of temperature compensation in poikilotherms". *Physiological reviews*, 1974, 54(3), p.620-677.

Hazel, Jr & Vl Schuster, "The effects of temperature and thermal acclimation upon osmotic properties and nonelectrolyte permeability of liver and gill mitochondria from rainbow trout (*Salmo gairdneri*)". *Journal of experimental zoology*, 1976, 195(3), p.425-438.

Hendry, Andrew P. & Michael T. Kinnison, "An introduction to microevolution: Rate, pattern, process". *Genetica*, 2001, 112-113, p.1-8.

Hennig, Wolfgang, "The revolution of the biology of the genome". *Cell research*, 2004, 14(1), p.1-7.

Herrmann, Johannes M., "Converting bacteria to organelles: evolution of mitochondrial protein sorting". *Trends in microbiology*, 2003, 11(2), p.74-79.

Hittel, D.S. & K.B. Storey, "Differential expression of mitochondria-encoded genes in a hibernating mammal". *Journal of experimental biology*, 2002, 205(11), p.1625-1631.

Hochachka, Peter, W., G. B. Mcclelland, G. P. Burness, J. F. Staples & R. K. Suarez, "Integrating metabolic pathway fluxes with gene-to-enzyme expression rates". *Comparative biochemistry and physiology*, 1998, 120, p.17-26.

Hochachka, Peter, W. & George N. Somero, *Biochemical adaptation: mecanism and progress in physiological evolution*. New York, Oxford University Press, 2002, 466 p.

Holland, Lz, M Mcfall-Ngai & George N. Somero, "Evolution of lactate dehydrogenase_A homologs of barracuda fishes (genus Sphyraena) from different thermal environments: Difference in kinetic properties and thermal stability are due to amino acid substitutions outside the active site". *Biochemistry*, 1997, 36(11), p.3207-3215.

Jolicoeur, Pierre, *Introduction à la biométrie*. Montréal, Décarie, 1998, 514 p.

Kermack, K.A. & J.B.S. Haldane, "Organic correlation and allometry". *Biometrika*, 1950, 37, p.30-41.

Kimura, Motoo, "Evolutionary rate at the molecular level". *Nature*, 1968, 217, p.624-626.

Kimura, Motoo, "A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences". *Journal of molecular evolution*, 1980, 16, p.111-120.

Kimura, Motoo, *The neutral theory of molecular evolution*. Cambridge, Cambridge University Press, 1983, p.

Kimura, Motoo, "Some models of neutral evolution, compensatory evolution, and the shifting balance process". *Theorical population theory*, 1990, 37, p.150-158.

Kimura, Motoo, "Recent development of the neutral theory viewed from the Wrightian tradition of theorical population genetics". *Proceeding of the national academy of sciences USA*, 1991, 88, p.5969-5973.

Kingsolver, Joel G. & Raymond B. Huey, "Evolutionary analyses of morphological and physiological plasticity in thermally variable environments". *American zoologist*, 1998, 38, p.545-560.

Kinnison, Michael T. & Andrew P. Hendry, "The pace of modern life II: From rates of comtemporary microevolution to pattern and process". *Genetica*, 2001, 112-113, p.145-164.

Kluge, Arnold G., "A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes)". *Systematic zoology*, 1989, 38(1), p.7-25.

Kluge, Arnold G., "Total evidence or taxonomic congruence: cladistics or consensus classification". *Cladistics*, 1998, 14, p.151-158.

Kluge, Arnold G., "On total evidence: for the record". *Cladistics*, 2004, 20, p.205-207.

Krebs, C.J., *Ecological methodology*. New York, Harper & Row, 1999, p.

Kumar, Sudhir, "Patterns of nucleotide substitution in mitochondrial protein coding genes of vertebrates". *Genetics*, 1996, 143, p.537-548.

Kumar, Sudhir, Koichiro Tamura, Ingrid B. Jacobsen & Masatoshi Nei, "MEGA 2: Molecular Evolutionary Genetics Analysis software, Arizona State University, Tempe, Arizona, USA." 2001.

Kunz, Wolfram S., "Different metabolic properties of mitochondrial oxidative phosphorylation in different cell types - important implications for mitochondrial cytopathies". *Experimental biology*, 2003, 88(1), p.149-154.

Kyte, Jack & Russel F. Doolitle, "A simple method for displaying the hydropathic character of a protein". *Journal of molecular biology*, 1982, 157, p.105-132.

Lang, B. Franz, Bruno Paquin & Gertraud Burger, "L'évolution moléculaire et la révolution génomique". *médecine/sciences*, 2000, 16, p.212-218.

Lavenda, Bernard, "Le mouvement brownien". *L'ordre du Chaos*, Paris, Pour la Science, 1997, p.8-21.

Li, Wen-Hsiung, *Molecular evolution*. Sunderland Massachusetts, Sinauer Associates, 1997, 487 p.

Liò, Pietro & Nick Goldman, "Modeling mitochondrial protein evolution using structural information". *Journal of molecular evolution*, 2002, 54, p.519-529.

Lucu, Cedomil & David W. Towle, "Na+K+ATPase in gills of aquatic crustacea". *Comparative biochemistry and physiology Part A*, 2003, (135), p.195-214.

Martin, Andrew P., "Metabolic rate and directional nucleotide substitution in animal mitochondrial DNA". *Molecular biology and evolution*, 1995, 12(6), p.1124-1131.

Martin, Andrew P. & Stephen R. Palumbi, "Body size, metabolic rate, generation time, and the molecular clock". *Proceeding of the national academy of sciences USA*, 1993, 90, p.4087-4090.

Mattei, Jean-François, "Mythologie, génétique et permanence de l'Homme". *La recherche en génétique et en génomique: droits et responsabilités*. S. Philips-Nootens, B. Godard, B. M. Knoppers & M.-H. Regnier, Montréal, les éditions Thémis, 2005, p.303-312.

Mauron, Alex, "Is the genome the secular equivalent of the soul?" *Science*, 2001, 291, p.831-832.

Mauron, Alex, "Genomic metaphysics". *Journal of Molecular Biology*, 2002, 319, p.957-962.

Mauron, Alex, "Renovating the house of being - genomes, souls, and selves". *Annals New York academy of sciences*, 2003, 1001, p.240-252.

Mcnairn, A.J. & D.M. Gilbert, "Epigenomic replication: linking epigenetics to DNA replication". *Bioessays*, 2003, 25(7), p.647-656.

Mcquarrie, Donald A. & Peter A. Rock, *General chemistry*. New York, W.H. Freeman and company, 1991, 1083 p.

Meléndez-Hevia, Enrique, Thomas G. Waddell & Marta Cascante, "The puzzle of the krebs citric acid cycle: assembling the pieces of chemically feasible reactions, and opportunism in the design of metabolic pathways during evolution". *Journal of molecular evolution*, 1996, 43, p.293-303.

Moilanen, Jukka S. & Kari Majamaa, "Phylogenetic network and physicochemical properties of nonsynonymous mutations in the protein-coding genes of human mitochondrial DNA". *Molecular biology and evolution*, 2003, 20(8), p.1195-1210.

Nicholls, David G. & Stuart J. Ferguson, *Bioenergetics 3*. London, Academic Press Inc., 2001, 288 p.

Nowotny, Helga, Peter Scott & Michael Gibbons, *Repenser la science*. Paris, Éditions Belin, 2001, 319 p.

Papa, Sergio, Ferrucio Guerrieri & Joseph M. Tager, *Frontiers of cellular bioenergetics*. New York, Klumer academic/plenum publishers, 1999, 798 p.

Perna, Nicole T. & Thomas Kocher, D., "Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes". *Journal of molecular evolution*, 1995, 41, p.353-358.

Petsko, Gregory A & Dagmar Ringe, *Protein structure and function*. London, New science press Ltd, 2004, 195 p.

Pörtner, H.O., "Climate variations and the physiologycal basis of temperature dependent biogeography: systemic to molecular hierarchy of thermal tolerance in animals". *Comparative biochemistry and physiology A*, 2002, 132, p.739-761.

Porto, Markus, H. Eduardo Roman, Michele Verdruscolo & Ugo Bastolla, "Prediction of site-specific amoni acid distributions and limits of divergent evolutionary changes in protein sequences". *Molecular biology and evolution*, 2005, 22(3), p.630-638.

Rand, David M., "Thermal habit, metabolic rate and the evolution of mitochondrial DNA". *Trends in ecology and evolution*, 1994, 9(4), p.125-131.

Rand, David M., "The units of selection on mitochondrial DNA". *Annual review in ecology and systematics*, 2001, 32, p.415-448.

Rand, David M., Robert A. Haney & Adam J. Fry, "Cytonuclear coevolution: the genomics of cooperation". *Trends in ecology and evolution*, 2004, 19(12), p.645-653.

Riggs, D.S., J.A. Guarnieri & S. Addelman, "Fitting straight lines when both variables are subject to error". *Life Science*, 1978, 22, p.1305-1360.

Rudden, D.M., L. Xiao, M.D. Garfinkel & X.Y. Lu, "Hsp90 and envionmental impacts on epigenetic states: a model for the trans-generational effects of diethylstibesterol on uterine development and cancer". *Human molecular genetics*, 2005, 14, p.R149-R155.

Saccone, Cecilia, Pierre-Olivier Barome, Anna Maria D'erchia, Ilenia D'errico, Graziano Pesole, Elisabetta Sbisà & Apollonia Tullo, "Molecular strategies in metazoan genomic evolution". *Gene*, 2002, 300, p.195-201.

Saccone, Cecilia, Carla De Giorgi, Carmela Gissi, Graziano Pesole & Aurelio Reyes, "Evolutionary genomics in metazoa: the mitochondrial DNA as a model system". *Gene*, 1999, 238, p.195-209.

Saccone, Cecilia, Carmela Gissi, Cecilia Lanave, Alessandra Larizza, Graziano Pesole & Aurelio Reyes, "Evolution of the mitochondrial genetic system: an overview". *Gene*, 2000, 261, p.153-159.

Saccone, Cecilia, Carmela Gissi, Aurelio Reyes, Alessandra Larizza, Elisabetta Sbisà & Graziano Pesole, "Mitochondrial DNA in metazoa: degree of freedom in a frozen event". *Gene*, 2002, 286, p.3-12.

Scheffler, Immo E., *Mitochondria*. New York, Wiley-Liss Inc., 1999, 367 p.

Schmidt, Timothy R., Morris Goodman & Lawrence I. Grossman, "Amino acid replacement is rapid in primates for the mature polypeptides of cox subunits, but not for their targeting presequences". *Gene*, 2002, 286, p.13-19.

Schmidt, Timothy R., Wei Wu, Morris Goodman & Lawrence I. Grossman, "Evolution of nuclear- and mitochondrial-encoded subunit interaction in cytochrome c oxidase". *Molecular biology and evolution*, 2001, 18(4), p.563-569.

Seidel-Rogol, B.L. & G.S. Shadel, "Modulation of mitochondrial transcription in response to mtDNA depletion and repletion in HeLa cells". *Nucleic acids research*, 2002, 30(9), p.1929-1934.

Selosse, Marc-André, "Quels ancêtres pour nos cellules?" *La recherche*, 2003, (362), p.48-52.

Selosse, Marc-André, Béatrice Albert & Bernard Godelle, "Reducing the genome size of organelle favours gene transfer to the nucleus". *Trends in ecology and evolution*, 2001, 16(3), p.135-141.

Somero, George N., "Thermal physiology and vertical zonation of intertidal animals: Optima, limits, and costs of living". *Integrative and comparative biology*, 2002, 42(4), p.780-789.

Somero, George N., "Linking biogeography to physiology: Evolutionary and acclimatory adjustments of thermal limits". *Frontiers in Zoology*, 2005, 2(1).

Souid, Abdul-Kader & Harvey S. Penefsky, "Mechanism of ATP synthesis by the mitochondrial ATP synthase". *Frontiers of cellular bioenergetics*, New York, Klumer academic/plenum publishers, 1999, p.377-398.

Stengers, Isabelle, *Sciences et pouvoir: la démocratie face à la technoscience*. Paris, Éditions La Découverte, 1997, 119 p.

Taanman, Jan-Willen, "The mitochondrial genome: structure, transcription, translation and replication". *Biochimica et biophysica acta*, 1999, 1410, p.103-123.

Templeton, Alan R., "Contingency tests of neutrality using intra/interspecific gene trees: The rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the hominid primates". *Genetics*, 1996, 144, p.1263-1270.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin & D. G. Higgins, "The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools". *Nucleic acids research*, 1997, 25, p.4876-4882.

Tourasse, Nicolas & Wen-Hsiung Li, "Selective constraints, amino acid composition, and the rate of protein evolution". *Molecular biology and evolution*, 2000, 17(4), p.656-664.

Uversky, Vladimir N., Joel R. Gillespie & Anthony L. Fink, "Why are "natively unfolded" proteins unstructured under physiologic conditions?" *Proteins: Structure, function, and genetics*, 2000, 41(3), p.415-427.

Villena, J.A., M.C. Carmona, M.R. De La Concepcion, M. Rossmeisl, O. Vinas, P. Mampel, R. Iglesias, M. Giralt & F. Villarroya, "Mitochondrial biogenesis in brown

adipose tissue is associated with differencial expression of transcription regulatory factors". *Cellular and molecular life sciences*, 2002, 59(11), p.1934-1944.

Webb, T., S.E. Howe, R.H.W. Bradshaw & K.M. Heide, "Estimating plant abundance from pollen percentages: the use of regression analysis". *Review of Palaeobotany and Palynology*, 1981, 34, p.269-300.

Willett, Christopher S. & Ronald S. Burton, "Evolution of interacting proteins in the mitochondrial electron transport system in a marine copepod". *Molecular biology and evolution*, 2004, 21(3), p.443-453.

Xia, Xuhua & Zheng Xie, "Protein structure, neighbor effect, and a index of amino acid dissimilarities". *Molecular biology and evolution*, 2002, 19(1), p.58-67.

Yang, Ziheng, Rasmus Nielsen, Nick Goldman & Anne-Mette Krabbe Pedersen, "Codon-substitution models for heterogeneous selection pressure at amino acid sites". *Genetics*, 2000, 155, p.431-449.

Yang, Ziheng, Rasmus Nielsen & Masami Hasegawa, "Models of amino acid substitution and applications to mitochondrial protein evolution". *Molecular biology and evolution*, 1998, 15(12), p.1600-1611.

Yoshikawa, Shinya, Kyoko Shinzawa-Itoh & Tomitake Tsukihara, "The structure of crystalline bovine heart cytochrome c oxydase". *Frontiers of cellular bioenegetics*, New York, Klumer academic/plenum publishers, 1999, p.131-156.