

UNIVERSITÉ MONTPELLIER II
SCIENCES ET TECHNIQUES DU LANGUEDOC

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ MONTPELLIER II

Discipline : Biologie des Populations et Écologie
Formation Doctorale : Biologie de l'Évolution et Écologie
École Doctorale : Systèmes Intégrés en Biologie, Agronomie, Géosciences,
Hydrosciences, Environnement

présentée et soutenue publiquement

par

Benoit Nabholz

le 17 novembre 2008

Titre :

**Dynamique évolutive de l'ADN mitochondrial des
oiseaux et des mammifères : Mutation, Sélection
et Taille des populations.**

JURY

Mme. EVELYNE HEYER, Professeur, MNHN, Paris	Rapporteur
M. ADAM EYRE-WALKER, Professeur, University of Sussex	Rapporteur
M. LAURENT DURET, Directeur de recherche, CNRS, Lyon	Examineur
M. YANNIS MICHALAKIS, Directeur de recherche, IRD, Montpellier	Examineur
M. NICOLAS GALTIER, Directeur de recherche, CNRS, Montpellier	Directeur
M. SYLVAIN GLÉMIN, Maître de conférence, Montpellier	Directeur

Remerciements

Me voilà arrivé à la fin de ce travail de thèse. Je dois dire que je n'aurais jamais imaginé dans ma scolarité ou même, lors de mon arrivée à la faculté comme étudiant, avoir l'honneur et le plaisir de réaliser une thèse. Lorsque je pense à mon parcours, je me rends compte de la chance que j'ai eu de baigner dans une ambiance familiale où ma curiosité et mon goût pour "la nature" ont toujours été encouragés et stimulés. Pour cela, je remercie tout particulièrement ma mère et ma grand-mère pour leurs encouragements sans faille, ainsi que mon père et mon grand-père pour m'avoir transmis leur curiosité du monde qui nous entoure.

L'évolution est difficile (impossible?) à prédire, les évolutionnistes parlent de la contingence pour expliquer la part de hasard dans les trajectoires évolutives prises par les espèces. Je pense que c'est également valable pour les parcours individuels. Dans mon parcours, j'ai eu la chance de rencontrer Nicolas Galtier et Sylvain Glémin. Cette rencontre m'a permis de découvrir de nombreuses problématiques fascinantes dans les sciences de l'évolution. Il ne fait pas de doute que cela influencera mon parcours futur. Leur encadrement a constitué pour moi une véritable formation scientifique. Pour cela, mais également pour avoir toujours répondu à mes sollicitations, m'avoir orienté et guidé scientifiquement, je les remercie tous les deux sincèrement.

Dans les deux équipes que j'ai fréquenté (GPIA puis Phylogénie Moléculaire), j'ai eu le plaisir de rencontrer des personnes intéressantes et au contact desquelles je me suis enrichi. Je voudrais saluer quelques personnes de l'ex-GPIA : Khalid Belkhir pour son aide et son efficacité, Alban Guillaumet pour le plaisir que j'ai pris à discuter avec lui de naturalisme (enfin surtout d'ornitho...), Jean-Baptiste Ferdy, Pierre Boursot et Pierre Caminade pour les discussions à la "pause café", fondamentales à l'équilibre d'une journée de travail! J'en profite également pour remercier Jean-François Mauffrey pour son aide et sa enthousiasme. Un grand merci à Eric Bazin et Julien Dutheil pour avoir supporté mes sollicitations de bioinformaticien "du dimanche" et pour leurs travaux scientifiques sans lesquels cette thèse n'aurait pas été la même.

Dans ma nouvelle équipe, je remercie : Emmanuel Douzery, Frédéric Delsuc, Vincent Ranwez et Marika Tilak-Jean pour leurs accueils mais également Andrew, Céline, Georgia et Benjamin qui ont rendu ma dernière année particulièrement agréable. L'acclimation a été rendu d'autant plus facile que l'équipe voisine, bien que n'ayant pas digérée la suprématie des datations moléculaires sur la paléontologie, c'est avéré particulièrement intéressante et agréable à côtoyer. Je pense tout particulièrement à Laurent, Rodolphe, Fabrice et Helder.

Table des matières

I	Introduction	3
1	Pourquoi s'intéresser à l'évolution moléculaire ?	5
1.1	L'évolution moléculaire : les grandes avancées conceptuelles	6
1.2	L'outil moléculaire	11
2	La Mitochondrie : un petit organite avec une grande histoire	17
2.1	La chaîne respiratoire	17
2.2	L'origine de la cellule eucaryote	19
2.3	Evolution du génome mitochondrial	21
2.3.1	Structure du génome mitochondrial	21
2.3.2	Génome mitochondrial et biodiversité	24
2.3.3	Hypermutableté	25
3	Objectifs de cette thèse	27
II	Quantification et Déterminisme des variations du taux de mutation mitochondrial	29
4	Etat des lieux des connaissances	33
4.1	Les hypothèses répliation-indépendantes	33
4.2	Taux d'évolution et théorie du vieillissement mitochondrial : une hypothèse répliation-dépendante	36
4.3	Les enjeux	38
5	Modèles biologiques et approche phylogénétique	39
5.1	Pourquoi se focaliser sur les oiseaux et les mammifères ?	39
5.2	Inférence du taux de mutation par une approche phylogénétique. . . .	41
5.2.1	Relation entre le taux de mutation et le taux de substitution	41

5.2.2	Inférence phylogénétique du taux d'évolution : le tour de force des méthodes Bayésiennes	42
5.2.3	Diviser pour mieux régner : une approche dé耦plée	46
6	Variation du taux de mutation et longévit�	51
6.1	Taux de mutation mitochondrial : Diff�rence de comportement entre les oiseaux et les mammif�res	51
6.1.1	Variations extr�mes du taux de mutation du cytochrome <i>b</i> chez les mammif�res	51
6.1.2	Taux de mutation du cytochrome <i>b</i> chez les oiseaux : moins rapide et moins variable	52
6.2	Confirmation de l'hypoth�se de long�vit� ?	55
6.2.1	R�sultats mitochondriaux	55
6.2.2	R�sultats nucl�aires	60
6.3	Robustesse et g�n�ralisation des r�sultats cytochrome <i>b</i>	61
6.4	Discussion et implication des r�sultats	66
6.4.1	Consid�rations sur la variation des taux d'�volution	66
6.4.2	Consid�rations sur le d�terminisme des taux d'�volution	68
III	Taille des populations et �volution mol�culaire mitochondriale	73
7	Soup�ons sur la neutralit� de l'ADN mitochondrial	75
7.1	Historique	75
7.2	S�lection naturelle et taille des populations	77
8	Taille des populations et �volution de l'ADNmt	81
8.1	Pas de traces de la s�lection positive chez les vert�br�s	81
8.2	Diversit� g�n�tique et taille des populations	84
8.2.1	Taux de mutation et polymorphisme mitochondrial	84
8.2.2	Taille des populations et polymorphisme mitochondrial	85
8.3	Contrainte �volutive et taille des populations	88
8.4	Interpr�tation des r�sultats : l'instabilit� populationnelle	91
8.5	Mod�lisation de l'�volution mol�culaire en population fluctuante	95
8.5.1	Le mod�le	95
8.5.2	Les simulations	97

IV Conclusion	101
9 Les apports de cette thèse	103
10 Traits d'histoire de vie et évolution moléculaire	107
V Annexes	131

Liste de Publications

- Nabholz B, Mauffrey J, Bazin E, Galtier N & Glemin S. 2008. Determination of mitochondrial genetic diversity in mammals. *Genetics* **178** : 351-361.
- Nabholz B, Glémin S & Galtier N. 2008. Strong variations of mitochondrial mutation rate across mammals : the longevity hypothesis. *Molecular Biology and Evolution* **25** : 120-130.
- Nabholz B, Glémin S & Galtier N. 2008. Mitochondrial DNA evolutionary dynamics : what can we learn from the bird/mammal comparison ? *BMC Evolutionary Biology* (en révision).
- Galtier N, Blier, P U & Nabholz B. 2008. Inverse relationship between evolutionary rate of mitochondrial proteins and longevity in mammals. *Mitochondrion* (en révision).
- Fabre P-H, Nabholz, B, Galtier, N, & Douzery, E J P. 2008. Assessing the congruence between mitochondrial DNA clocks and the vertebrate fossil record. *Molecular Biology and Evolution* (sousmis).

Première partie

Introduction

Chapitre 1

Pourquoi s'intéresser à l'évolution moléculaire ?

In fact, the DNA of every living organism is an accumulation of historical records. [...] The purpose of molecular evolution is to unravel these historical records, put the information in order, read it, and decipher its meaning.

When-Hsiung Li et Dan Graur, *Fundamentals of Molecular evolution*, 1991.

L'évolution moléculaire est la discipline qui s'attache à décrire l'évolution des génomes et de leurs constituants (ADN, ARN, gènes, protéines). Cette jeune discipline est née de la rencontre entre la biologie évolutive et la biologie moléculaire. Cette naissance est liée au développement des techniques de la biologie moléculaire telles que l'électrophorèse, au milieu des années 60, ou le séquençage par PCR (Réaction de Polymérisation en Chaîne) en 1983. La confrontation de cette nouvelle source de données avec les théories développées par la génétique des populations a permis d'aboutir à des avancées conceptuelles majeures (voir chapitre 1.1).

Les génomes sont des objets biologiques à part : leur statut de support de l'hérédité les place au centre des processus évolutifs, la sélection naturelle n'ayant pas d'action sur un caractère non héritable. Les génomes portent une quantité non négligeable d'informations qui nous renseignent sur l'histoire évolutive des organismes. Les processus évolutifs sont souvent des processus lents. La plupart des événements évolutifs, comme la spéciation, ne sont pas observables dans une échelle de temps humaine. Cependant, les génomes portent les empreintes de ces événements et permettent à qui sait les décrypter d'avoir accès à des millions d'années d'histoire évolutive. Depuis le début de l'histoire de l'évolution moléculaire, cette discipline s'articule autour de deux objectifs. Il y a d'abord l'évolution moléculaire que j'appellerai "pure", où l'objet d'étude est directement la macro-molécule. Les génomes et leurs constitutants sont des entités biologiques à part entière qui ont une histoire que le biologiste peut avoir à cœur de comprendre. Mais les molécules peuvent être également utilisées en tant qu'outils, pour reconstruire l'histoire évolutive des espèces et des populations. Ces deux corpus sont étroitement connectés : on ne reconstruit l'histoire évolutive que si l'on connaît les fonctionnements des marqueurs utilisés et on ne comprend l'évolution des génomes que si l'on connaît l'histoire phy-

logénétique des espèces. Je vais, dans cette introduction, insister sur cette dualité entre objet d'études et outils informatifs, car elle est présente dans ce travail de thèse mais également parce qu'elle reflète mon parcours personnel où, en tant que naturaliste, je me suis d'abord intéressé aux molécules comme des outils avant de me pencher spécifiquement à l'évolution de celles-ci. Le thème central de cette thèse porte sur l'évolution moléculaire mitochondriale. Les problématiques liées à l'étude de l'évolution de l'ADN mitochondrial découlent des nombreuses spécificités de ce génome que nous introduiront au chapitre 2.3, mais sont également en connexion avec les concepts généraux de l'évolution moléculaire (chapitre 1.1).

1.1 L'évolution moléculaire : les grandes avancées conceptuelles

La redécouverte des lois de l'hérédité de Mendel en 1900 par Hugo De Vries et Carl Correns ne profite d'abord pas au Darwinisme. Les généticiens, notamment William Bateson, ne sont alors pas favorables à une vision Darwinienne de l'évolution où la sélection naturelle a une place primordiale, mais considèrent plutôt que l'évolution se produit par des changements brusques et radicaux. La vision Darwinienne est alors défendue par les biométriciens tel que Karl Pearson. Il faudra attendre les années 30 pour que les généticiens des populations – tels que Ronald A. Fisher, John B. S. Haldane, Gustave Malécot puis Sewall Wright – mais également des généticiens comme Thomas H. Morgan, réconcilient génétique et sélection naturelle. Les généticiens des populations ont apporté un cadre théorique pour comprendre l'évolution des allèles au sein des populations. Cette vision populationnelle, qui se détache des études des Mendéliens classiques, rend compatible la génétique Mendélienne avec l'évolution Darwinienne (Bowler, 2003). Les premiers modèles d'évolution moléculaire font jouer un rôle prépondérant à la sélection naturelle et permettent de faire un certain nombre de prédictions sur les patrons moléculaires. Par exemple, le polymorphisme doit toujours rester faible car les mutants favorables sont rapidement fixés dans la population. Concernant le taux d'évolution, Haldane (1957) (*in* Kimura, 1983) prédit que le *taux de substitution*, c'est à dire le nombre de mutations fixées par génération, devrait être d'environ $1/300$ pour un mammifère classique. Cette prédiction, calculée pour un génome de mammifère moyen, est basée sur le coût de la sélection naturelle, ou fardeau de sélection, qui considère la perte des individus portant un allèle défavorable lors du remplacement par un allèle favorable dans une population. La *fixation* d'un allèle sélectionné positivement a, en effet, un coût pour la population. Le nombre de substitutions se trouve ainsi limité, par ce coût, à une valeur "supportable" pour une taille de population donnée.

Ces premiers modèles vont rapidement être mis à l'épreuve des données par l'obtention des premières séquences protéiques. De cette confrontation naîtra l'une des plus intéressantes controverse de la biologie : le débat entre neutralistes et sélectionnistes (Kimura, 1983; Gillespie, 1991). La comparaison des séquences homologues entre espèces permet de constater la présence de différences, on parle de *substitutions*. En extrapolant aux génomes de mammifères le taux de substitution

de quelques protéines, Motoo Kimura constate que ce taux est bien supérieur à celui prédit théoriquement par Haldane (Kimura, 1968a). Concluant que ce taux n'est pas compatible avec l'action de la sélection naturelle, Kimura introduit en 1968 le modèle *neutre* qui postule que la majorité des différences fixées correspondent à des mutations sans effet sur la valeur sélective des individus qui les portent (Kimura, 1983). Pour confirmer sa théorie, Kimura s'appuie également sur la comparaison des divergences entre les gènes de la familles des hémoglobines chez les vertébrés. Des divergences globalement similaires entre paralogues indiquent que les gènes ont évolué à un taux constant, en accord avec la jeune théorie de l'*horloge moléculaire* (Zukerkandl & Pauling, 1965). Ce taux d'évolution constant est incompatible avec une évolution majoritairement sous l'influence de la sélection naturelle, forcément erratique au gré des événements d'adaptation. Kimura développe également des arguments basés sur la quantité de polymorphisme (Kimura, 1968b). En effet, les méthodes électrophorétiques, développées dans les années 60, montrent que le niveau de variabilité génétique de nombreux locus est bien supérieur à celui prédit par l'hypothèse sélectionniste (Kimura, 1983). Le travail de Kimura, qui s'accompagne d'une modélisation raffinée du processus d'évolution neutre, a une influence décisive sur la discipline. Parallèlement, King et Jukes ont également développé des idées neutralistes, en s'appuyant, cette fois, sur des arguments plus biochimiques (King & Jukes (1969) *in* Ohta & Gillespie (1996)). Ils constatent une relation négative entre l'importance fonctionnelle d'une protéine, ou d'un site à l'intérieur d'une protéine, et son taux d'évolution. Ils sont les premiers à introduire la notion de *contraintes évolutives* à l'intérieur des génomes. Ainsi, les mutations qui se fixent sont majoritairement celles qui se produisent dans des régions ou des protéines dont la séquence en amino-acides affecte peu la survie de l'organisme qui les porte. Au contraire, celles qui se produisent dans des régions contraintes sont éliminées par la sélection.

La théorie neutraliste propose une nouvelle conception des forces évolutives qui influencent l'évolution moléculaire en plaçant la *dérive génétique* au premier plan. Classiquement cette force était considérée comme trop lente pour avoir une influence prépondérante dans l'évolution, contrairement à la sélection naturelle, qui a une action beaucoup plus rapide. La variable prépondérante qui influence l'intensité de la dérive génétique est la *taille efficace*, N_e . La taille efficace correspond à l'effectif d'une population "idéale" de type Wright-Fisher qui présenterait les caractéristiques génétiques de la population réelle observée. Une population de Wright-Fisher correspond à une population de taille constante, non structurée, sans sélection et à générations non-chevauchantes. Dans le cas d'une population idéale, la taille efficace correspond au nombre moyen d'individus qui se reproduisent à chaque génération. Cependant, de nombreux paramètres peuvent provoquer un décalage entre la taille de la population réelle et la taille efficace, par exemple un sex-ratio biaisé, une structuration spatiale, une fluctuation des populations dans le temps. Ainsi, dans une population, une mutation est considérée comme neutre si $s \ll 1/N_e$, où s est la valeur sélective de la mutation. Selon la théorie neutraliste, les mutations avantageuses sont considérées comme extrêmement rares et, de ce fait, elles ne participent pas à la divergence de manière visible.

Autour des années 1970, Kimura et Ohta (Ohta & Kimura, 1971; Ohta, 1973)

clarifient la notion de *mutation presque neutre*. Cette notion a d'abord été employée pour désigner des mutations dont le coefficient de sélection est très faible et qui, de fait, se comportent comme des mutations neutres. La nouvelle conception associée au terme mutation presque neutre se focalise sur la classe de mutations avec un coefficient de sélection quasi équivalent à la dérive génétique, $s \approx 1/N_e$. Cette classe de mutation peut basculer dans une dynamique de type mutation sélectionnée ou neutre en fonction de l'effectif de la population considérée (Figure 1.1). La théorie presque neutre rajoute de la flexibilité à la théorie neutre. Observant un impact fort du temps de génération sur le taux de d'évolution neutre mais faible sur le taux d'évolution protéique, Ohta & Kimura (1971) proposent que les espèces à temps de génération long ont en moyenne une taille efficace plus faible, et fixent donc plus de mutations délétères, atténuant l'effet mutationnel. La théorie presque-neutre est, également, utilisée pour expliquer la faible variation de la diversité allozymique entre plusieurs espèces alors qu'une grande variation était attendue sous l'hypothèse neutre (Nei & Graur, 1984). Si la proportion des mutations presque neutres est importante alors les patrons de divergence et de polymorphisme attendus sont différents du modèle neutre strict (Figure 1.2). L'une des différences les plus notables est la relation négative attendue entre le taux d'évolution protéique et l'effectif des populations. La relation entre la taille des populations et le taux d'évolution est différente selon les différentes conceptions de l'évolution moléculaire : (i) si la majorité des substitutions sont sélectionnées positivement (modèle sélectionniste) la relation est positive, (ii) si les substitutions sont neutres (modèle neutraliste) la relation est plate, enfin (iii) si les substitutions sont délétères (modèle presque-neutre) la relation est négative (Figure 1.3). Selon le modèle neutre, le taux de substitution est indépendant de la taille des populations : le nombre de mutations qui apparaissent dans une population est égal au produit de la taille des populations N_e (soit $2N_e$ allèles pour une population diploïde) par le taux de mutation μ , et la probabilité de fixation de chaque mutation est égale à sa fréquence initiale dans la population, soit $1/2N_e$. Ainsi, le taux de substitution neutre est égal au produit $2N_e\mu t \times (1/2N_e)$, soit simplement μt . Selon le modèle sélectionniste, la probabilité de fixation d'une mutation avantageuse ($2s$) est indépendante de la taille des populations, et le taux de substitution, $4N_e\mu s$, est proportionnel à N_e . Enfin, selon le modèle presque-neutre, le taux d'évolution est négativement corrélé à la taille des populations car l'efficacité de la sélection ($N_e s$) augmente avec la taille efficace, diminuant la proportion des substitutions délétères (Figure 1.2).

Si le débat est toujours actif quant à la proportion de substitutions fixées sous l'effet de la sélection naturelle dans l'évolution des génomes, le modèle neutraliste a eu l'immense mérite de mettre en place une hypothèse nulle qu'il faut rejeter avant de pouvoir discuter en termes adaptatifs les patrons de variation observés dans les génomes (voir (Eyre-Walker, 2006) pour un revue). Bien que l'évolution moléculaire n'en soit évidemment pas à l'abri, l'héritage de Kimura confère à cette discipline, plus qu'à d'autres, une méfiance vis à vis d'éventuelles dérives adaptationnistes (parfois qualifiées de Panglossiennes, Gould & Lewontin (1979)). Actuellement, certains modèles se basent entièrement sur des forces non-adaptatives pour décrire l'évolution de la structure des génomes (Lynch & Conery, 2003; Lynch *et al.*, 2006)

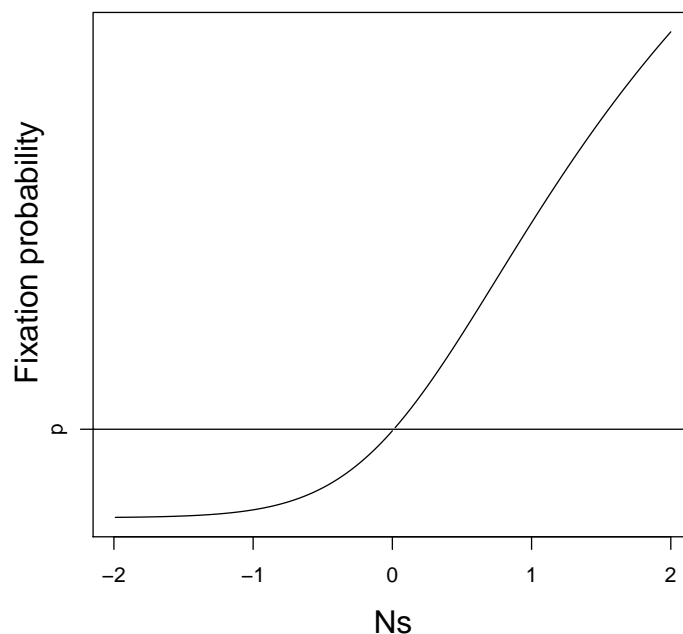


FIG. 1.1 – La probabilité de fixation d'un mutant en fonction du produit de l'effectif de la population et du coefficient de sélection, Ns , relativement au cas d'un allèle strictement neutre où la probabilité de fixation est égale à la fréquence initiale, p . Modifié d'après Ohta (1992)

Théorie Neutralistes

Délétères		Neutres	— Avantages
Sélectionnées	Presque neutres	Neutres	

Théorie Presque Neutre

FIG. 1.2 – Diagramme des différentes classes de mutations selon la théorie neutre classique ou la théorie presque neutre. La classe sélectionnées représente à la fois les mutations délétères ($N_e s \ll 1$) et avantageuses ($N_e s \gg 1$). La classe presque neutre contient des mutations neutres ou délétères selon l'effectif de la population considérée. Tiré de Ohta (1992)

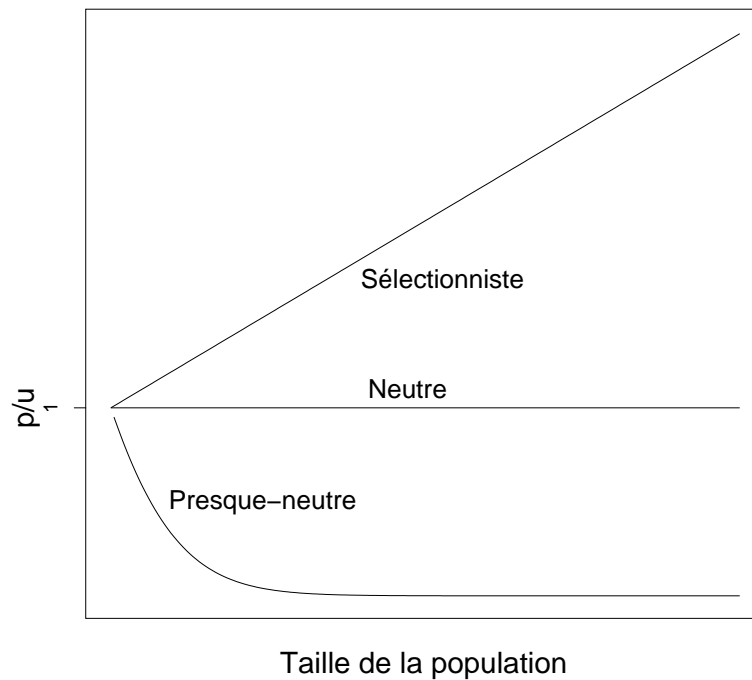


FIG. 1.3 – Le taux d'évolution divisé par le taux de mutation en fonction des trois modèles de l'évolution moléculaire. Modifié d'après Gillespie (2004)

et montrent l'impact de la logique initiée par Kimura. L'importance du neutralisme s'est, par ailleurs, propagée bien au-delà du champ de l'évolution moléculaire et de la génétique des populations, et on retrouve, par exemple, sa trace dans les modèles d'écologie des communautés (Hubbell, 1997; Bell, 2001). Dès les premières études directement rattachées à l'évolution moléculaire, les concepts de neutralisme, d'horloge moléculaire et de contrainte évolutive sont introduits. Le développement de la théorie neutre de l'évolution montre également comment la confrontation entre les données et la théorie est particulièrement fertile en avancées conceptuelles. Ce type de confrontation est facilité dans le domaine de l'évolution moléculaire par le fait que (i) suite aux pionniers des années 30, le développement théorique est toujours resté très actif, et (ii) la quantité de données issues des techniques de séquençage est toujours en augmentation exponentielle.

Outre les avancées conceptuelles en génétique des populations, les progrès dans l'inférence statistique à partir des séquences d'ADN ou d'acides aminés ont eu un impact décisif dans le domaine de l'évolution moléculaire. Contrairement aux autres caractères présentés par un organisme (morphologiques, comportementaux etc...), les caractères moléculaires sont composés de répétitions de fragments en nombre limité et bien définis. Cette singularité se prête tout particulièrement à une modélisation des processus évolutifs par des méthodes probabilistes telles que les modèles de Markov. Des modèles de plus en plus complexes vont être développés pour décrire l'évolution entre séquences (Felsenstien, 2004; Yang, 2007). Ces modèles vont être repris dans des méthodes statistiques élaborées telles que le maximum de vraisemblance et les méthodes Bayésiennes, utilisées pour inférer des phylogénies,

mais également d'autres paramètres ayant une signification biologique telles que les distances synonymes et non-synonymes (chapitre 8.3) ou des dates de divergence (chapitre 5.2).

1.2 L'outil moléculaire : la machine à remonter dans le temps

Les problématiques touchant à l'évolution moléculaire peuvent être abordées avec deux points de vues : celui du biologiste moléculaire, qui à travers des concepts évolutionnistes comprend mieux le fonctionnement des génomes et de leurs constituants ; ou celui du systématicien et de l'écologiste qui utilisent des outils moléculaires afin de dévoiler des facettes cachées ou d'apporter une nouvelle perspective de recherche sur un modèle d'étude traditionnel.

L'une des disciplines qui a le plus profité de l'apport des données moléculaires est la phylogénie. Les relations de parenté classiquement acceptées entre les organismes ont été totalement revisitées. La phylogénie moléculaire présente l'avantage, par rapport aux approches classiques basées sur la morphologie, d'utiliser des milliers de caractères homologues. De plus, certaines molécules assurant des fonctions élémentaires de la cellule sont partagées par tous les êtres vivants et leur utilisation permet de tenter de reconstruire l'arbre du vivant. J'ai choisi d'illustrer les apports des phylogénies moléculaires en montrant comment les relations de parenté à l'intérieur des deux grands groupes de vertébrés étudiés dans cette thèse ont évolué aux cours des dix dernières années. Chez les mammifères, les principaux bouleversements ont eu lieu au niveau des relations supra-ordinales des mammifères placentaires. Les caractères morphologiques se sont révélés très homoplasiques, du fait des convergences évolutives liées à des modes de vie particuliers (Murphy *et al.*, 2004), ou tout simplement inexistantes tant les morphologies peuvent avoir dérivé par rapport à une forme ancestrale. Chacun des quatre super-ordres définis par la phylogénie moléculaire regroupe des organismes morphologiquement très différents et révèle comment l'adaptation peut "gommer" des ressemblances ancestrales (Madsen *et al.*, 2001; Murphy *et al.*, 2001), Figure 1.4. Parmi les résultats les plus intéressants, citons la découverte du super-ordre des Afrotheria, dont les espèces actuelles ne sont caractérisées par aucune synapomorphie morphologique, et qui regroupe des organismes aussi hétéroclites que des grands herbivores terrestres (proboscidiens - éléphants) ou aquatiques (siréniens - lamantins), des petits insectivores écologiquement similaires à nos hérissons, musaraignes et taupes (Afrosoricidae, Tenrecidae et Macroscelidae) ou encore une espèce myrmécophage (l'oryctérope, Tubulidentata) (Tabuce *et al.*, 2007).

Concernant les oiseaux, une classification moléculaire basée sur la technique d'hybridation ADN-ADN a été mise en place dès le début des années 1990 par Sibley & Ahlquist (1990). Suite à ce travail, tout à fait remarquable par sa date de publication et la quantité d'espèces représentées, l'augmentation de la quantité de données et l'emploi de méthodes d'inférence phylogénétique plus efficaces ont permis de clarifier un certain nombre de relations de parenté (Barker *et al.*, 2004; Ericson *et al.*,

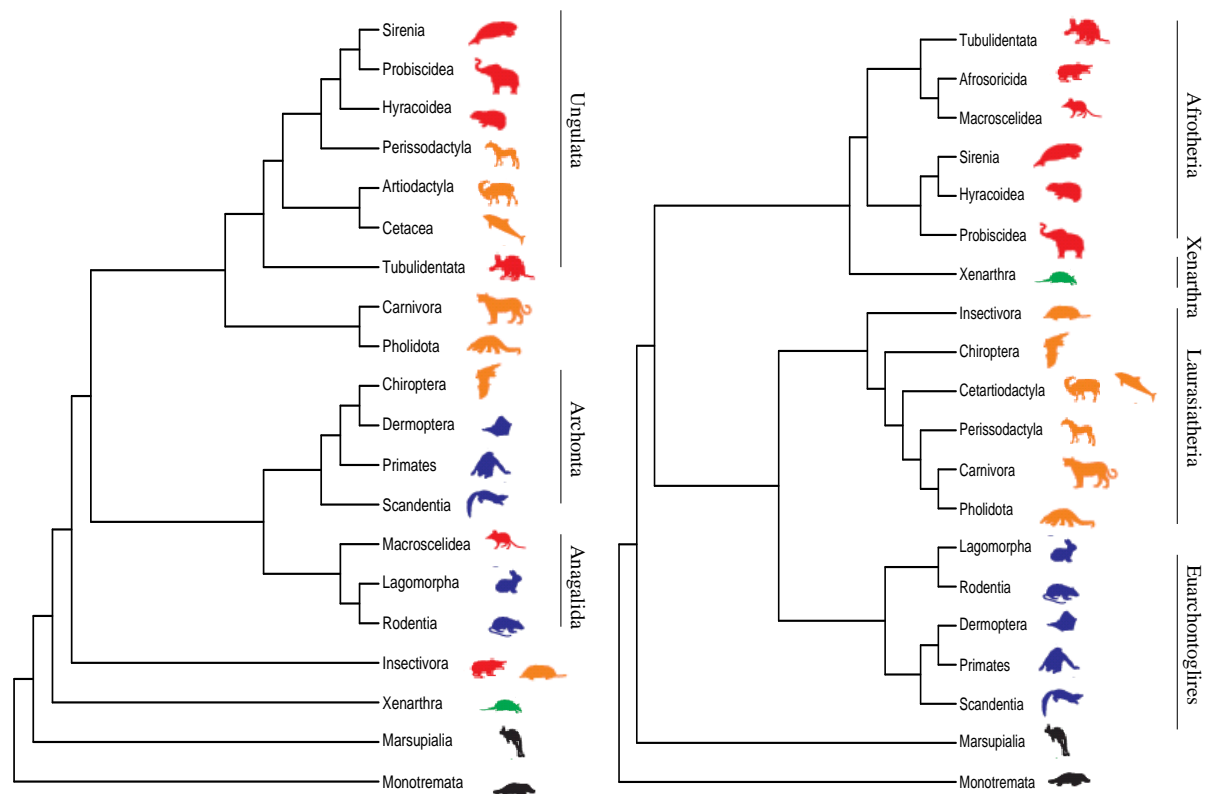


FIG. 1.4 – Comparaison entre l'arbre phylogénétique construit à partir de caractères morphologiques (gauche) ou de caractères moléculaires (droite). Les longueurs de branches sont arbitraires. La racine des mammifères placentaires, pour la topologie moléculaire, respecte le regroupement des Xénarthres avec les Afrothériens (Atlantogenata) et des Euarchotheries avec les Laurasiathériens (Boreoeutheria) conformément aux analyses phylogénomiques récentes (Prasad *et al.*, 2008; Ranwez *et al.*, 2007). Modifié d'après Murphy *et al.* (2004).

2006), (Figure 1.5). Il est intéressant de noter que même avec un jeu de données aux dimensions génomiques, certaines relations basales du groupe des Neoaves restent irrésolues, probablement sous l'effet de spéciations extrêmement rapides (Hackett *et al.*, 2008). La diversification des oiseaux, comme celle des mammifères, s'est accompagnée de nombreux événements de convergence évolutive qui brouillent les inférences basées sur des caractères morphologiques. Citons, par exemple, l'apparition de la morphologie "rapace" trois fois indépendamment dans le clade des rapaces nocturnes (Strigiformes, chouettes et hiboux), des faucons (Falconidae) et des autres rapaces diurnes (Accipitridae + Cathartidae, buses, vautours et aigles).

A la fin des années 80, le terme de phylogéographie est introduit dans la littérature pour désigner les études qui s'intéressent spécifiquement à décrire et interpréter la répartition géographique des généalogies (Avice, 1998). Cette discipline, à l'interface entre la phylogénie moléculaire, la génétique des populations et la biogéographie, a connu un grand succès avec la démocratisation des techniques de séquençage dans les années 90. La multiplication des études de phylogéographie,

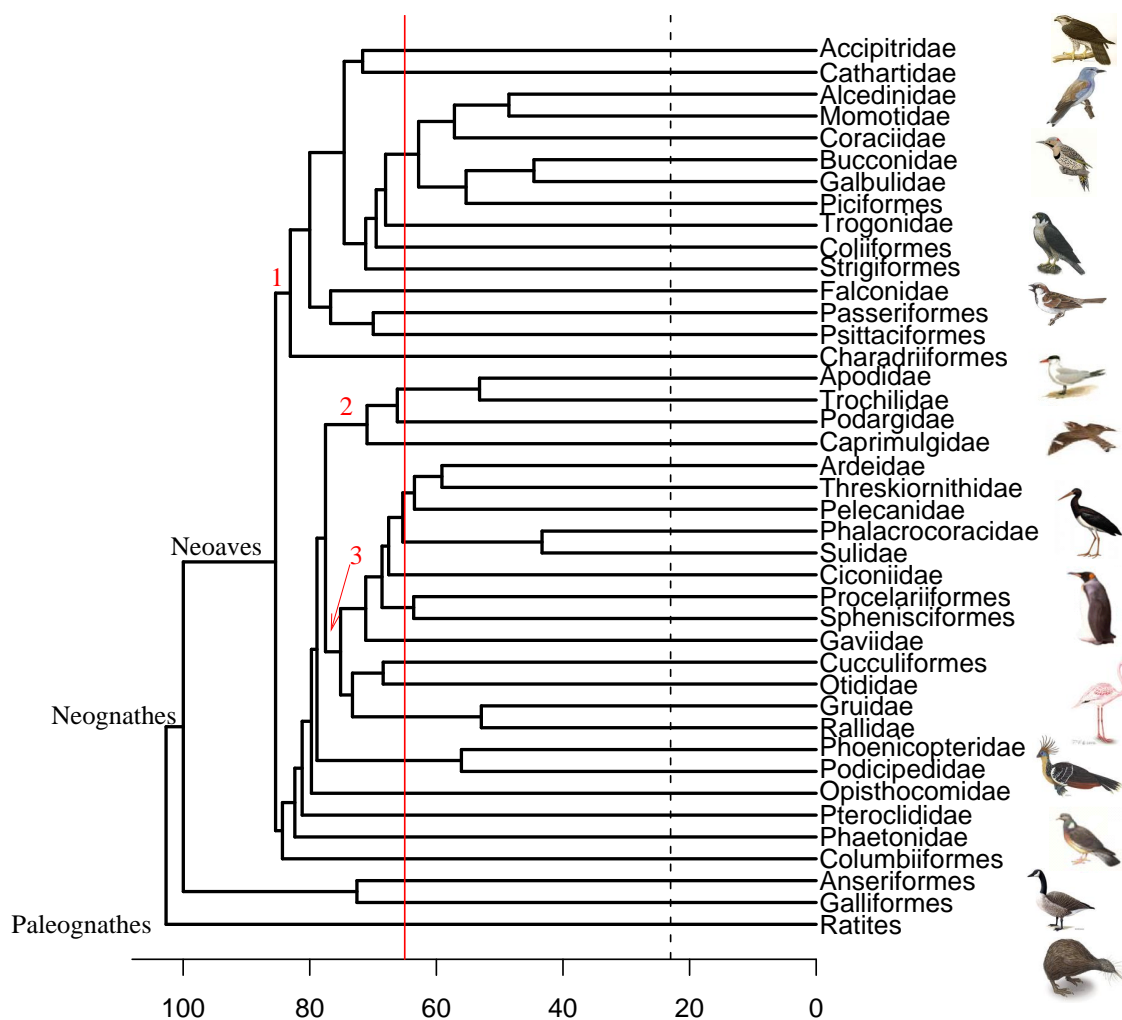


FIG. 1.5 – Phylogénie et datation moléculaire des principaux groupes d’Oiseaux. La datation est obtenue avec 4 gènes nucléaires représentant 4700 nucléotides et deux points de calibrations : Procellariiformes/Sphenisciformes entre 65 et 63 Millions d’années et galliformes/anseriformes à plus de 66 Millions d’années. La ligne verticale rouge indique la limite Crétacé/Cénozoïque et la ligne verticale pointillée la limite Paléogène/Néogène. La topologie respecte Hackett *et al.* (2008). Noter les spéciations très rapides à la base des Néoaves. Les positions des Phoenicopteridae + Podicipedidae, Opisthocomidae, Pteroclididae, Phaetonidae et Columbiformes restent incertaines, ainsi que celles des grands clades notés 1, 2 et 3.

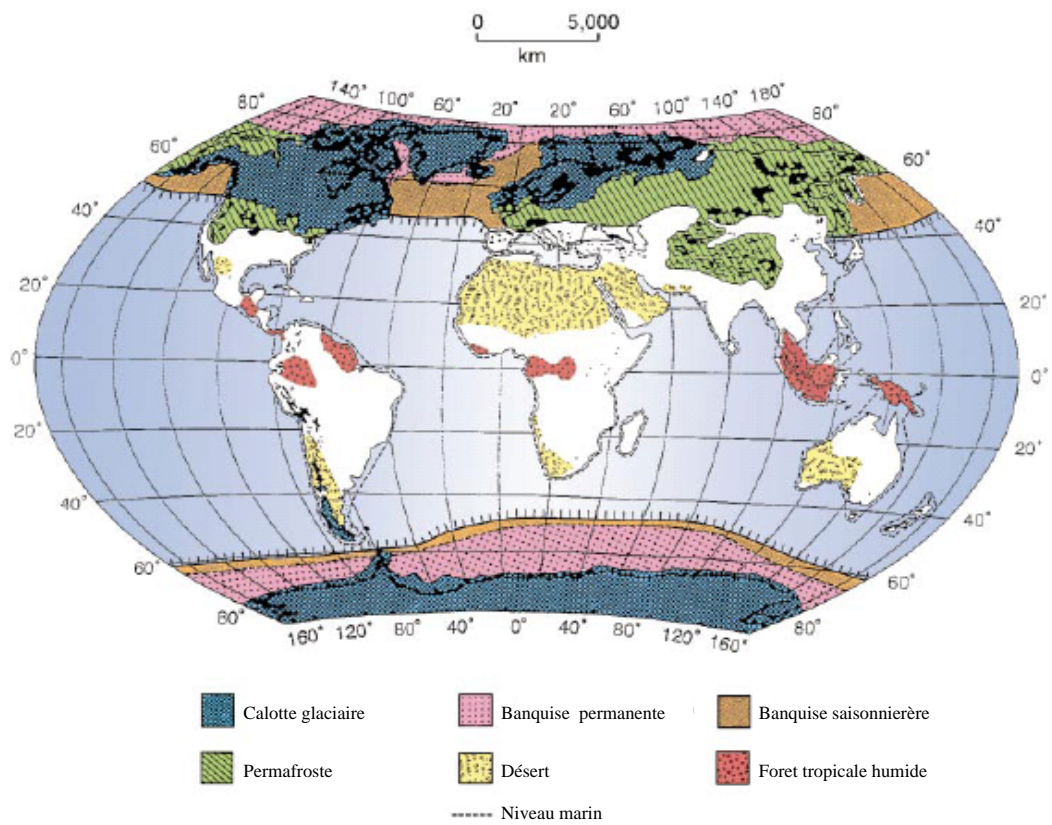


FIG. 1.6 – Climat lors du maximum de la dernière période glaciaire, il y a 20 000 ans. Tiré de Hewitt (2000)

principalement sur des espèces européennes et nord américaines, a permis de retracer l'histoire récente des espèces et, notamment, de faire ressortir des patrons similaires entre espèces prouvant l'influence prépondérante des fluctuations climatiques pléistocènes.

L'ère Quaternaire (2,4 Millions d'années) se caractérise par une alternance climatique de périodes chaudes et de périodes glaciaires. Au cours du dernier million d'années (Pléistocène), la périodicité de ces alternances est d'environ 100 000 ans, décomposés en périodes glaciaires ($\sim 80\,000$ ans) et périodes chaudes (ou interglaciaires, $\sim 20\,000$ ans). Ces alternances climatiques sont sous le contrôle de modifications de la trajectoire orbitale de la Terre connues sous le nom de cycle de Croll-Milankovitch. Actuellement, la Terre se trouve dans une période inter-glaciaire chaude mais des mesures paléoclimatiques comme les analyses isotopiques et palynologiques permettent de se représenter le climat du dernier maximum glaciaire (Hewitt (2000), Figure 1.6). Les principales conséquences écologiques de la dernière période glaciaire sont une migration (et une fragmentation) des écosystèmes vers des latitudes plus faibles ainsi qu'une baisse générale du niveau marin d'environ 120m créant des liaisons terrestres entre régions isolées pendant la période inter-glaciaire. Ces changements environnementaux ont façonné la diversité et la généalogie des espèces actuelles. En Europe, la fragmentation des écosystèmes tempérés dans les 3 grandes péninsules méditerranéennes (Ibérique, Ligurienne et Balkanique) a créé des refuges pour les espèces qui occupaient ces écosystèmes. Ces refuges sont marqués par une diversité élevée et un fort niveau d'endémisme. La recolonisation, pendant les périodes inter-glaciaires, s'accompagne de contacts secondaires entre les populations qui ont divergées dans ces péninsules pouvant éventuellement créer des zones hybrides (Hewitt, 2000). La présence de refuges glaciaires est aussi avérée dans d'autres parties du monde comme en Amérique du nord. Même si l'importance et le positionnement géographique des refuges glaciaires restent débattus (Weir & Schluter, 2004; Rowe *et al.*, 2004), la majorité des groupes taxonomiques étudiés par la phylogéographie porte les traces de cette fluctuation climatique pléistocène. Ainsi les aires de répartition actuelles des espèces ne reflètent probablement pas ce qu'elles étaient il y a moins de 20 000 ans.

Un grand nombre des résultats fondamentaux sur les relations de parenté et l'histoire évolutive des espèces ont été obtenus grâce à l'apport des données moléculaires. Si on s'intéresse plus précisément au type de marqueurs utilisés, on peut constater que la majorité de ces résultats provient de l'utilisation de l'ADN mitochondrial. On peut expliquer ce succès par un certain nombre de caractéristiques propres à cette molécule. L'ADN mitochondrial présente un taux d'évolution particulièrement rapide, qui constitue un critère de choix important (peut-être le plus important) car il garantit à l'utilisateur un bon pouvoir résolutif. L'ADN mitochondrial possède également d'autres qualités, comme sa nature non-recombinante et sa facilité d'extraction, induite par son abondance dans les tissus, qui l'ont rapidement fait devenir le marqueur privilégié pour la phylogéographie et la phylogénie à bas niveau taxonomique (Avise, 1998; Harrison, 1989, voir chapitre 2.3.2). Ce génome a, également, une histoire évolutive fascinante, notamment à travers son origine endosymbiotique (chapitre 2). Cette histoire singulière associée à sa très grande popularité comme marqueur justifie que l'on s'intéresse de près à l'évolution de ce

génomé.

Chapitre 2

La Mitochondrie : un petit organite avec une grande histoire

It seems to me that the fusion of two whole genomes to create the first eukaryotic cell is better viewed as a macro-mutation to create a *hopeful monster* than purely as a succession of small genetic changes.

Nick Lane, *Power, Sex, Suicide*, 2005.

La mitochondrie est un organite caractéristique des cellules eucaryotes. De la taille d'une bactérie (1 à 4 microns de longueur et 0,5 de diamètre), elle est formée de deux membranes phospholipidiques, dites externe et interne, et possède son propre génome. Ce petit organite, le plus souvent hérité maternellement, est présent en grande quantité dans nos cellules, entre quelques centaines et plusieurs milliers pour les ovocytes. La mitochondrie occupe une fonction centrale dans la cellule eucaryote qui consiste à produire de l'énergie via la respiration.

2.1 La chaîne respiratoire

A l'intérieur de la cellule eucaryote, la mitochondrie occupe la fonction de générateur d'ATP via la chaîne respiratoire. Ce processus se décompose en deux étapes dites de *chaîne de transport d'électrons* et de *phosphorylation oxydative*. Ces deux processus sont *couplés* et sont réalisés par cinq complexes protéiques encastés dans la membrane interne de la mitochondrie (Figure 2.1). Tout commence par la chaîne de transport d'électrons qui récupère les électrons produits par le cycle de Krebs pour produire une cascade de réactions d'oxydo-réduction. L'électron va passer d'un élément de la chaîne à un autre, chaque passage étant exoénergétique. Finalement, l'électron est transmis, dans la cytochrome oxydase (complexe IV), à l'accepteur final de la chaîne, un oxygène, pour former de l'eau (Figure 2.1). Cette succession de réactions d'oxydo-réductions produit de l'énergie qui va être utilisée pour le transport actif de protons, également issus du cycle de Krebs, de la matrice vers l'espace intermembranaire. Trois des quatre complexes de la chaîne de transport d'électrons font office de pompe à protons. La membrane de phospholipides étant

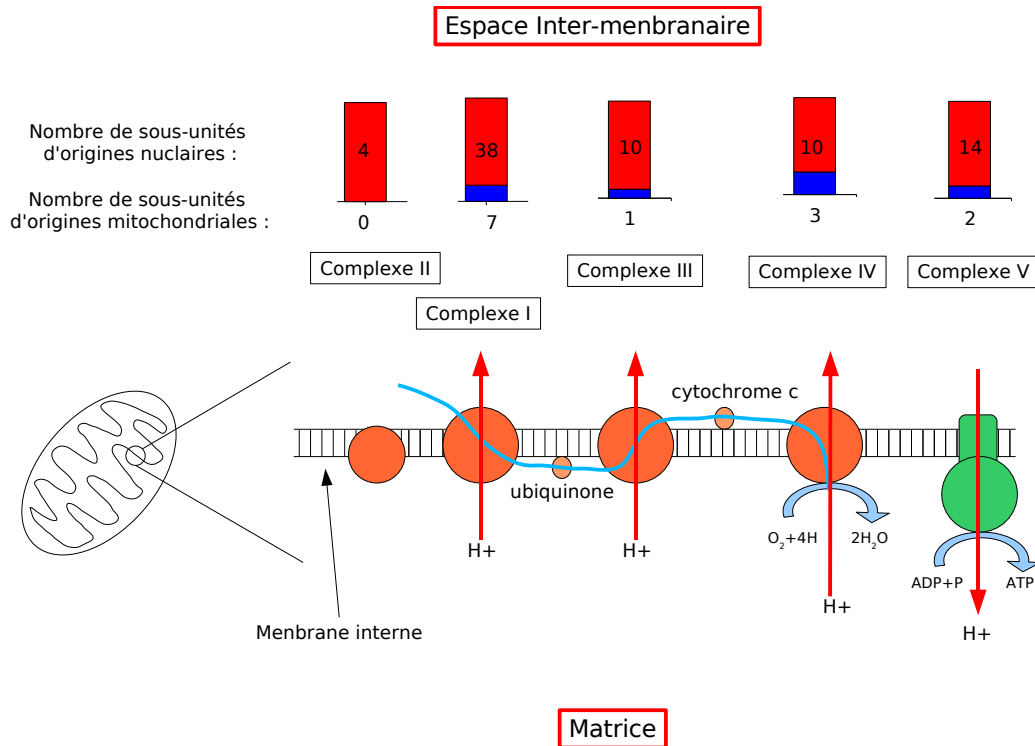


FIG. 2.1 – Représentation simplifiée de la chaîne respiratoire avec la chaîne de transport d'électron (e^-) en orange et de l'ATP synthétase en vert. Le trajet de l' e^- est représenté par le trait bleu. l' e^- pénètre dans la chaîne à la fois par l'NADH déshydrogénase (complexe I) ou la succinate déshydrogénase (complexe II), puis est transporté vers le complexe III (contenant le cytochrome b) par l'ubiquinone (ou Coenzyme Q). L'énergie libérée par le transport d' e^- est utilisée pour le transport actif de protons vers l'espace inter-membranaire (flèche rouge), créant un gradient de protons (ΔH^+). L'énergie potentielle est utilisée pour condenser l'ADP + P en ATP via l'ATP synthétase (complexe V).

imperméable aux protons, un gradient de potentiel et de concentration (ΔH^+) va se former de part et d'autre de la membrane. Ce gradient va être utilisé par l'ATP synthétase (complexe V) pour générer de l'ATP à partir de ADP + Pi.

De manière surprenante, les complexes de la chaîne de transport d'électrons et de l'ATP synthétase sont des panachés de protéines issues à la fois du génome nucléaire et du génome mitochondrial. La proportion de protéines issues du génome mitochondrial est faible (Figure 2.1), mais ces peptides remplissent les fonctions catalytiques majeures au sein des complexes (Race *et al.*, 1999).

2.2 L'origine de la cellule eucaryote

Lynn Margulis propose en 1967 (Margulis, 1967, 1981) que les organites des eucaryotes étaient, à l'origine, des procaryotes libres. D'abord fortement controversée, cette hypothèse a été remise sur le devant de la scène avec l'avènement de la phylogénie moléculaire. La phylogénie du vivant réalisée avec la petite sous-unité de l'ARN ribosomique montre bien l'apparentement du génome mitochondrial avec les eubactéries (Woese & Fox, 1977; Pace *et al.*, 1986) (Figure 2.2). La cellule eucaryote serait issue de la fusion entre une cellule "proto-eucaryote" et une bactérie de type α -protéobactérie il y a près de 2 milliards d'années. Si l'origine de la mitochondrie est bien identifiée comme étant apparentée aux α -protéobactéries, la nature de la cellule proto-eucaryote reste débattue, et particulièrement le fait de savoir si la cellule eucaryote possédait déjà un noyau et un cytosquelette avant l'acquisition d'une mitochondrie ou si ces caractéristiques ont été acquises après la fusion. Les partisans de l'apparition d'un noyau avant la fusion avec la mitochondrie s'appuient sur la présence de cellules eucaryotes actuelles dépourvues de mitochondries positionnées à la base des eucaryotes. Ce groupe d'eucaryotes est appelé "archezoa" (Cavalier-Smith, 1987). Cependant cette hypothèse a été battue en brèche par la découverte de traces de gènes mitochondriaux dans le génome des Archezoa, prouvant que ces eucaryotes ont perdu secondairement leurs mitochondries (Clark & Roger, 1995; Keeling, 1998). Les archezoa sont, en effet, pour la plupart des parasites qui utilisent les fonctions métaboliques de leurs hôtes.

Une série d'analyses visant à enraciner les trois grands domaines du vivant, a constaté des similarités dans la machinerie cellulaire des eucaryotes et Archaeobactéries. De la même façon, des traces de génomes Archae dans le génome eucaryote tendent à prouver que le génome eucaryote est une chimère entre Eubactérie et Archae (Lang *et al.*, 1999; Rivera & Lake, 2004). L'hypothèse de la fusion entre une Archae et une eubactérie est connue sous le nom d'*hypothèse hydrogène* (Martin & Muller, 1998). Cette hypothèse propose que la fusion s'est faite à la faveur d'un rapprochement symbiotique entre une Archae autotrophe (hydrogène dépendante) jouant le rôle de l'hôte et une eubactérie polyvalente capable de respirer mais produisant de l'hydrogène comme produit final d'un métabolisme anaérobie. La cellule proto-eucaryote doit donc avoir évolué dans des conditions anaérobies avant de passer à un métabolisme aérobie (Martin & Muller, 1998).



FIG. 2.2 – Arbre phylogénétique réalisé avec la petite sous-unité de l'ARN ribosomique d'après Pace *et al.* (1986). En noir, gris foncé et gris clair, les trois grands domaines du vivants respectivement : les Archebactéries, les eucaryotes et les Eubactéries.

2.3 Evolution du génome mitochondrial

2.3.1 Structure du génome mitochondrial

Par de nombreux aspects, le génome mitochondrial apparaît dans son organisation, différent du génome nucléaire. Le génome mitochondrial est généralement circulaire et de taille très inférieure au génome nucléaire. La recombinaison est très rare, voire totalement absente, et il a une hérédité maternelle (Dawid & Blackler, 1972; Hutchiso *et al.*, 1974). Une dichotomie claire existe entre le génome mitochondrial des végétaux et des animaux (Lang *et al.*, 1999; Lynch *et al.*, 2006), alors qu'elle se retrouve plutôt entre les procaryotes et les eucaryotes en ce qui concerne le génome nucléaire. La taille du génome mitochondrial des végétaux est généralement élevée et fortement variable, entre ~ 100 et ~ 400 kb. Ces variations de taille sont majoritairement dues à l'expansion de régions non-codantes et à la présence d'introns (Lynch *et al.*, 2006). Au contraire, le génome mitochondrial des animaux est remarquablement compact, avec très peu de régions inter-géniques. Le génome mitochondrial des animaux mesure typiquement entre ~ 14 et ~ 20 kb (Gissi *et al.*, 2008). Il en est de même pour le contenu en gènes : 13 gènes protéiques, 2 ARN ribosomiques et 22 ARN de transferts (Figure 2.3). Les déviations de ce modèle canonique sont rarissimes chez les métazoaires, particulièrement en ce qui concerne les gènes protéiques. Certains nématodes, bivalves et les placozoaires ont perdu leurs ATP8 (Kim *et al.*, 2006; Milbury & Gaffney, 2005; Signorovitch *et al.*, 2007), le gène ND5 manque apparemment à certaines grenouille arboricoles du genre *Polypedates* (Zhang *et al.*, 2005), enfin les extravagants téléostéens antarctiques du groupe des Notothenioidei ont perdu leur ND6 (Papetti *et al.*, 2007). Cette stabilité du génome mitochondrial est présente jusque dans la synténie (ordre des gènes) qui reste relativement conservée à l'échelle des vertébrés. Ainsi, la comparaison de la synténie entre les mammifères et les oiseaux révèle très peu de changements pour deux groupes qui ont pourtant divergé il y a près de 320 millions d'années (Figure 2.3).

Cette stabilité du contenu en gènes du génome mitochondrial des animaux est d'autant plus étonnante que le début de l'histoire évolutive de la mitochondrie s'est accompagné d'une réduction drastique de son contenu. Cette réduction s'est effectuée à la faveur d'un transfert massif vers le génome nucléaire. Pour retracer l'histoire évolutive du génome mitochondrial, il est important d'identifier à quoi ressemblait exactement son ancêtre libre. En effet, les particularités écologiques des bactéries influencent la quantité de gènes contenus dans leurs génomes. Par exemple, le génome, pauvre en gènes, du parasite *Rickettsia prowazekii* apparaît phylogénétiquement proche de la mitochondrie. Néanmoins, cet apparemment pourrait être artificiellement provoqué par le mode vie parasitaire entraînant la simplification de *Rickettsia* (Andersson *et al.*, 1998; Müller & Martin, 1999). Au contraire, l'hypothèse hydrogène, invoquée dans le chapitre précédent, implique que l'ancêtre de la mitochondrie devait être polyvalent au niveau de son spectre métabolique, et notamment capable de métabolisme anaérobie avec production d'hydrogène. Une telle versatilité dans les fonctions métaboliques existe chez les α -protéobactéries, on la trouve notamment chez des espèces telles que *Rhodobacter*. Ainsi en considérant que le génome mitochondrial ancestral ressemble à celui de *Rhodobacter*, la réduction de

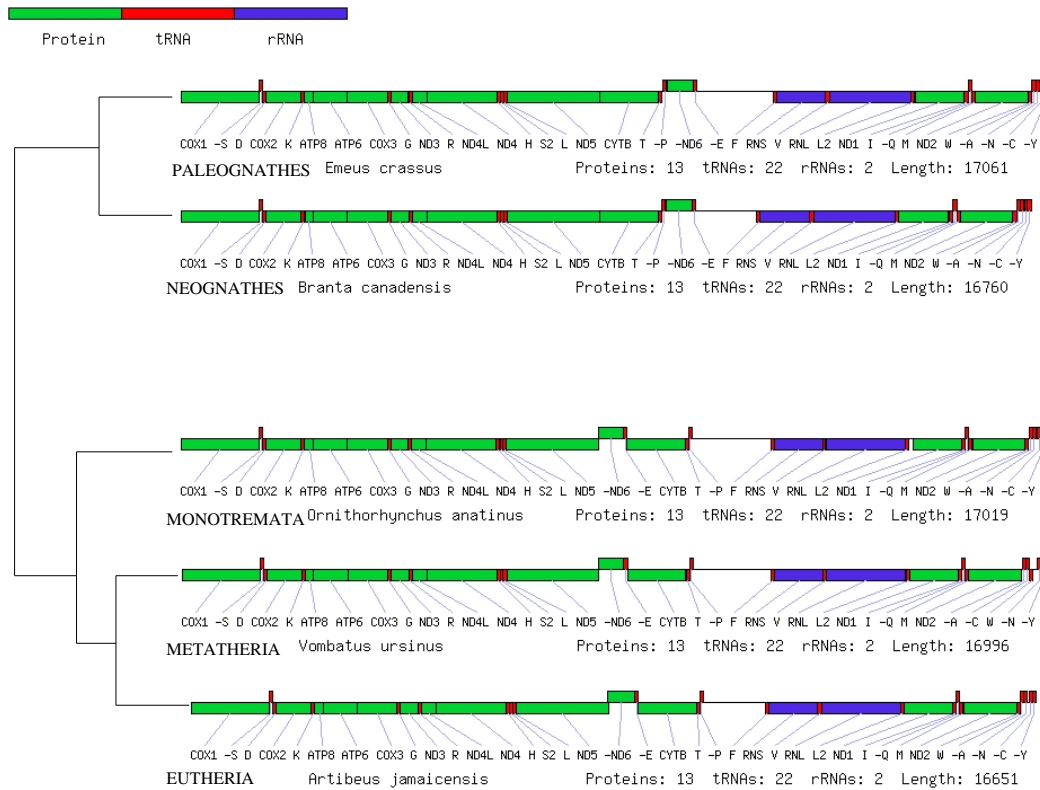


FIG. 2.3 – Représentation schématique du génome des grands clades de mammifères et d’oiseaux. Les traits à gauche des génomes représentent les relations phylogénétiques. La longueur des rectangles représentant les gènes est proportionnelle à la longueur du gène. ATP correspond aux ATP synthétases, NADH aux NADH déhydrogénases, CYTB au Cytochrome *b* et COX aux Cytochrome oxydases. Figure réalisée à partir du site web OGRE (<http://drake.physics.mcmaster.ca/ogre/>)

taille serait de l'ordre de 99,5%.

De manière surprenante, le transfert massif des gènes mitochondriaux vers le noyau ne s'est pas poursuivi jusqu'à la disparition totale du génome mitochondrial. Cette constatation est quelque peu troublante étant donné qu'il y a de bonnes raisons pour réduire la taille du génome mitochondrial, et que certaines mitochondries ont totalement perdu leur génome (il s'agit des *hydrogénosomes* qui existent chez certains eucaryotes unicellulaires (Embley *et al.*, 1997). Parmi les hypothèses invoquées pour expliquer la réduction de taille du génome, citons une sélection pour une réplication rapide à travers une compétition inter-mitochondriale (Selosse *et al.*, 2001). On peut également ajouter à cette sélection au niveau mitochondrial, une sélection au niveau de l'organisme liée au coût de la réplication du génome mitochondrial. En effet, les cellules contiennent généralement des centaines de mitochondries qui contiennent elles-même des dizaines de copies de leur génome. Le coût de la réplication est d'autant plus faible que le génome mitochondrial est compact. Une hypothèse explique la différence de taille du génome mitochondrial entre les plantes et les animaux en invoquant le taux de mutation élevé de l'ADN mitochondrial des animaux. Ce fort taux de mutation pourrait limiter la prolifération des introns, à cause du fardeau de mutation supplémentaire qu'impliquerait la mise en place de sites d'épissages (Lynch *et al.*, 2006). De plus, les gènes mitochondriaux transférés dans le noyau ont l'avantage de récupérer le faible taux de mutation nucléaire, réduisant ainsi les risques de mutations délétères (Allen & Raven, 1996).

S'il existe de bonnes raisons pour transférer les gènes mitochondriaux vers le noyau, pourquoi le processus s'est-il arrêté si proche de l'élimination complète du génome ? On sait, par ailleurs, que des fragments d'ADN mitochondrial sont continuellement transférés vers le noyau (Martin, 2003). Les occasions de réduire encore la taille du génome mitochondrial ne manquent donc pas. Il est difficile de trouver des hypothèses explicatives convaincantes et la solution reste peut-être à découvrir. Mais nous pouvons proposer deux explications :

- Premièrement, le génome mitochondrial a un code génétique particulier, qui varie même entre les végétaux, les vertébrés et les arthropodes. On peut imaginer que ce changement de code génétique peut bloquer la réussite d'éventuels transferts vers le noyau, du fait de l'impossibilité de la traduction de la protéine mitochondriale fraîchement transférée. Selon cette hypothèse, le code génétique mitochondrial aurait changé, quelle qu'en soit la raison, entraînant un arrêt du transfert des gènes mitochondriaux vers le noyau. Le contenu en gènes actuel des mitochondries serait, en quelque sorte, une photographie de l'état du génome à la date du changement de code génétique.
- Deuxièmement, par le biais de la chaîne respiratoire, la mitochondrie peut modifier rapidement le potentiel d'oxydo-réduction (redox) de la cellule. Des ajustements dans la quantité de protéines sont nécessaires pour conserver un potentiel optimal imposé par les changements d'activité physique et physiologique de l'organisme. Les gènes protéiques du génome mitochondrial seraient conservés afin de réguler l'activité redox de manière fine. Chaque mitochondrie peut, en fonction de son potentiel redox, synthétiser quelques

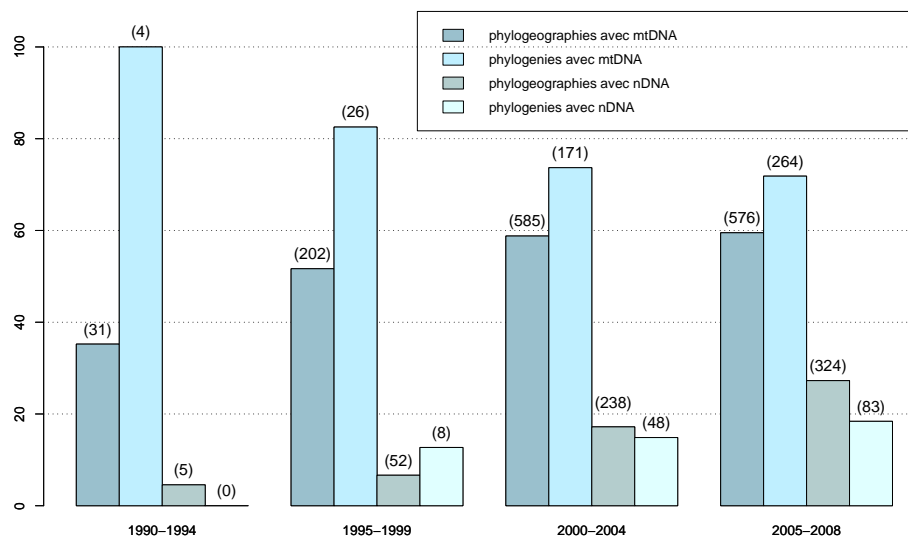


FIG. 2.4 – Pourcentage de publications en phylogénie et en phylogéographie réalisées à l'aide de marqueurs mitochondriaux (mtDNA) ou nucléaires (nDNA) par périodes. Les chiffres entre parenthèses représentent le nombre de publications correspondantes. Recherche bibliographique réalisée à partir de PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>). La recherche de publications en phylogénie s'est limitée au journal *Molecular Phylogenetics and Evolution*.

gènes clés qui, par leur positionnement dans la membrane phospholipidique, serviront de signal pour la formation de chaînes respiratoires via l'importation de gènes mitochondriaux d'origine nucléaire (Allen, 1993; Race *et al.*, 1999). Cette théorie est confirmée par la position centrale, encastrés dans la membrane phospholipidique, des gènes protéiques conservés dans le génome mitochondrial.

2.3.2 Génome mitochondrial et biodiversité

Comme je l'ai rapidement évoqué au chapitre 1.2, l'ADN mitochondrial est sans conteste le génome le plus souvent utilisé pour étudier la biodiversité. La majorité des résultats de phylogéographie et une grande partie des phylogénies ont été obtenus avec l'ADN mitochondrial (Figure 2.4). Récemment, la technique du code barre génétique (DNA barcoding) propose d'employer une petite portion standardisée de 648 bases d'un gène protéique de l'ADN mitochondrial (Cytochrome oxydase I) dans le but d'identifier toutes les espèces animales. Cette taxonomie moléculaire, même si elle reprend largement des concepts initiés par la phylogéographie, connaît actuellement un grand succès, comme en témoigne la progression du projet Barcoding Of Life Data system (BOLD) avec plus de 46 000 espèces séquencées (<http://www.barcodinglife.org>, Ratnasingham & Hebert (2007)).

Cet engouement remet ainsi une nouvelle fois l'ADN mitochondrial au centre des problématiques liées à la biodiversité.

On peut raisonnablement se poser la question des raisons d'un tel succès. Comme je l'ai évoqué au chapitre précédent, le génome mitochondrial des animaux est compact, petit et peu variable chromosomiquement parlant, ce qui fait des gènes mitochondriaux des homologues clairement identifiés, même entre espèces assez divergentes. L'obtention d'un alignement d'une fraction de quelques centaines de bases d'ADN mitochondrial ne pose généralement pas de problème particulier, facilitée par l'abondance de l'ADN mitochondrial dans les tissus. Le marqueur mitochondrial est, par exemple, privilégié dans les analyses à partir d'animaux éteints (*e.g.* Rohland *et al.*, 2007). Enfin, le fait que ce génome soit haploïde évite également la phase de clonage et permet un séquençage direct.

Au delà de ses qualités pratiques pour l'extraction et l'amplification, l'ADN mitochondrial a aussi des propriétés favorables qui découlent de sa biologie. Il a une hérédité clonale avec une transmission uniquement maternelle (Dawid & Blackler, 1972; Hutchins *et al.*, 1974), les événements de recombinaison sont très rares (mais voir Piganeau *et al.*, 2004), facilitant ainsi l'interprétation des patrons généalogiques (Moritz *et al.*, 1987; Harrison, 1989; Avise, 1994). De plus, les fonctions primordiales, liées à la respiration, codées par le génome mitochondrial, laissent supposer que la majorité des variations qui touchent ce génome sont neutres et que les événements de sélection positive sont rares (Moritz *et al.*, 1987). Enfin, et c'est peut-être la raison majeure du succès de l'ADN mitochondrial, ce marqueur présente un taux d'évolution particulièrement élevé. Cette rapidité évolutive associée à une taille efficace, dans le cas d'une espèce avec un sex ratio équilibré (Birky *et al.*, 1983), quatre fois plus faible qu'un marqueur autosomal, permet d'augmenter les chances de pouvoir reconstruire les patrons et le tempo de l'histoire récente des populations. Dans le chapitre suivant, j'exposerai les hypothèses qui ont été proposées pour expliquer le taux d'évolution élevé.

2.3.3 Hypermutableté

Une des caractéristiques les plus évidentes de l'ADN mitochondrial des animaux est son fort taux de mutation - 10 à 25 fois plus important que celui de l'ADN nucléaire (Ballard & Whitlock, 2004; Lynch, 2006). Cette *hypermutableté* mitochondriale est peut-être la caractéristique décisive qui explique le succès de ce marqueur pour les études de génétique des populations, de phylogéographie et de phylogénie. Ici encore, les végétaux se démarquent des animaux par leurs taux de mutation extrêmement faibles, en moyenne 50 à 100 fois moins vite que celui de l'ADN mitochondrial des animaux (Brown *et al.*, 1979; Wolfe *et al.*, 1987). Quelques rares exceptions extravagantes existent cependant. Les genres *Geranium*, *Silene* et *Plantago* ont acquis indépendamment une accélération exceptionnelle de l'évolution de leurs génomes mitochondriaux (Cho *et al.*, 2004; Mower *et al.*, 2007). Les raisons qui expliquent ces accélérations sont d'autant plus difficiles à trouver que ces phénomènes sont rarissimes dans la phylogénie des Embryophytes et que ces événements sont apparemment temporaires. Par exemple, chez *Plantago*, certaines espèces ont retrouvé un taux de mutation comparable à la moyenne des végétaux

(Cho *et al.*, 2004).

La dichotomie du taux de mutation entre végétaux et animaux nous montre que l'hypermutableté mitochondriale n'est pas une généralité du monde vivant et que certains organismes "savent" réduire leur taux de mutation. D'autant que chez les animaux, il existe également des exceptions chez les anthozoaires (coraux, gorgones et anémones de mer, appartenant à la classe des cnidaires) qui possèdent apparemment tous un taux de mutation faible (Hellberg, 2006). On peut donc se poser la question de savoir pourquoi les animaux présentent un taux de mutation mitochondrial aussi élevé. Deux hypothèses proximales peuvent être proposées : (i) les protéines de la réplication de l'ADN mitochondrial sont moins fidèles que celles de l'ADN nucléaire. L'ADN polymérase mitochondriale (polymérase γ) est différente de la principale ADN polymérase nucléaire (polymérase α). In vitro, elle n'apparaît pourtant pas moins fidèle que la polymérase nucléaire (voir Loeb & Monnat (2008) pour une revue). Cependant, la différence de fidélité pourrait provenir des mécanismes de réparation, moins efficaces dans le cas de l'ADN mitochondrial que pour l'ADN nucléaire (Mason & Lightowers, 2003). (ii) La deuxième cause possible provient de la production de *radicaux libres oxydatifs* lors de la respiration. Les électrons qui circulent dans la chaîne de transport peuvent s'échapper et réagir avec certains composés pour former des radicaux libres oxydatifs fortement mutagènes (Wallace, 2005). Les radicaux libres sont des composés avec un ou plusieurs électrons non appariés sur leur couche externe qui les rendent très instables, c'est à dire susceptibles de réagir avec de nombreux composés. Ces sous-produits de la respiration peuvent directement entraîner des modifications biochimiques sur des protéines et des lipides constituant la mitochondrie, mais également des mutations sur le génome mitochondrial physiquement proche.

Enfin, évolutivement parlant, cette hypermutableté peut être vue comme un paradoxe sachant (i) que de nombreuses maladies pourraient être évitées avec un taux de mutation mitochondrial plus faible (en effet, l'hypermutableté mitochondriale est à l'origine de nombreuses maladies humaines (Wallace, 2005), et particulièrement lorsque des mutations touchent les ARNt (Florentz, 2002)), (ii) que les cellules animales peuvent répliquer un génome fidèlement, le génome nucléaire présentant un taux de mutation 5 à 25 fois plus faible, et que (iii) d'autres cellules eucaryotes ont un taux de mutation mitochondrial beaucoup plus faible, comme les végétaux par exemple.

Chapitre 3

Pourquoi s'intéresser à l'évolution du génome mitochondrial ?

Intérêts et objectifs de cette thèse

J'ai essayé de montrer à travers cette introduction pourquoi le génome mitochondrial constitue un modèle d'étude de choix. Son histoire évolutive est complexe et fascinante, associant une origine endosymbiotique avec des caractéristiques très différentes du génome nucléaire (chapitre 2.3). Les quelques kilobases qui constituent le génome mitochondrial contiennent des gènes essentiels impliqués dans des fonctions métaboliques primordiales. Ces caractéristiques forcent naturellement l'intérêt mais sont renforcées par plusieurs avancées techniques et conceptuelles, aux nombres desquelles on peut citer :

1. De nombreuses avancées dans le domaine de la médecine et de la physiologie permettent, à présent, de mieux comprendre le lien entre l'activité mitochondriale et les mécanismes du vieillissement. Si le lien entre la production de radicaux libres oxydatifs (comme sous produit de la respiration) et le vieillissement est supposé depuis la fin des années 50 (Harman, 1957, voir chapitre 4.1), il a récemment été confirmé par de nombreuses analyses notamment comparatives (Barja & Herrero, 2000). De plus, l'implication du génome mitochondrial, à travers son hypermutabilité, a été récemment proposée par Trifunovic *et al.* (2004). L'hypermutabilité mitochondriale est également à l'origine de nombreuses maladies humaines (chapitre 2.3.3). Ces progrès récents, principalement réalisés dans le domaine biomédical, incitent à intégrer la longévité, avec une approche phylogénétique utilisant les concepts de la biologie évolutive, comme un déterminant potentiel du taux de mutation mitochondrial. Ce travail aura également l'avantage de créer une liaison entre la littérature évolutionniste et la littérature biomédicale.
2. L'ADN mitochondrial est l'outil moléculaire le plus largement utilisé, son statut de favori provenant de ses particularités, notamment la neutralité, l'hypermutabilité, la clonalité et sa facilité d'obtention (chapitre 2.3.2). Cependant, certaines de ces caractéristiques ont été récemment remises en question. L'un des plus sérieux questionnements concerne la neutralité évolutive de l'ADN

mitochondrial (Ballard & Whitlock, 2004; Hurst & Jiggins, 2005; Bazin *et al.*, 2006). Ces résultats importants appellent des analyses complémentaires. Il apparaît, par exemple, important d'intégrer simultanément la mutation, la dérive et la sélection pour bien comprendre la dynamique évolutive de l'ADN mitochondrial.

Le premier objectif de cette thèse (partie 2) consiste à *quantifier le taux de mutation* de l'ADN mitochondrial par une approche phylogénétique. Cette quantification précise du taux de mutation nous permettra de nous intéresser à la *compréhension du processus ultime* du contrôle de l'hypermutableté mitochondriale. Pour ce faire, nous avons choisi de nous intéresser au déterminisme des variations du taux de mutation mitochondrial entre lignées.

Le deuxième objectif de cette thèse consistera à tester l'hypothèse de neutralité (partie 3). Cette partie se concentrera sur le déterminisme de la diversité génétique intra-spécifique mitochondriale en utilisant comme point de départ le travail à large échelle taxonomique de Bazin *et al.* (2006). Nous avons choisi de nous focaliser sur des groupes taxonomiques plus restreints (oiseaux et mammifères) afin de prendre le contre pied de l'étude de Bazin *et al.* (2006), dans laquelle la biologie des groupes étudiés pouvait être très variable, nuisant à la comparabilité. La neutralité sera testée à l'aide de différentes approches et notamment à travers la relation entre diversité génétique et taille des populations. Enfin, nous intégrerons des résultats récents concernant la relation entre efficacité de la sélection et taille des populations (Popadin *et al.*, 2007; Berlin *et al.*, 2007). Nous essaierons de proposer un scénario général permettant de rendre cohérent l'ensemble des résultats sur le taux de mutation, le polymorphisme et la sélection.

Cette thèse a donc deux objectifs distincts qui seront traités dans deux parties (2 et 3) qui peuvent être lues indépendamment l'une de l'autre. Dans la partie finale, nous nous attacherons à faire ressortir les liaisons entre les différents résultats et à dégager le fil conducteur de cette thèse.

Deuxième partie

Quantification et Déterminisme des variations du taux de mutation mitochondrial

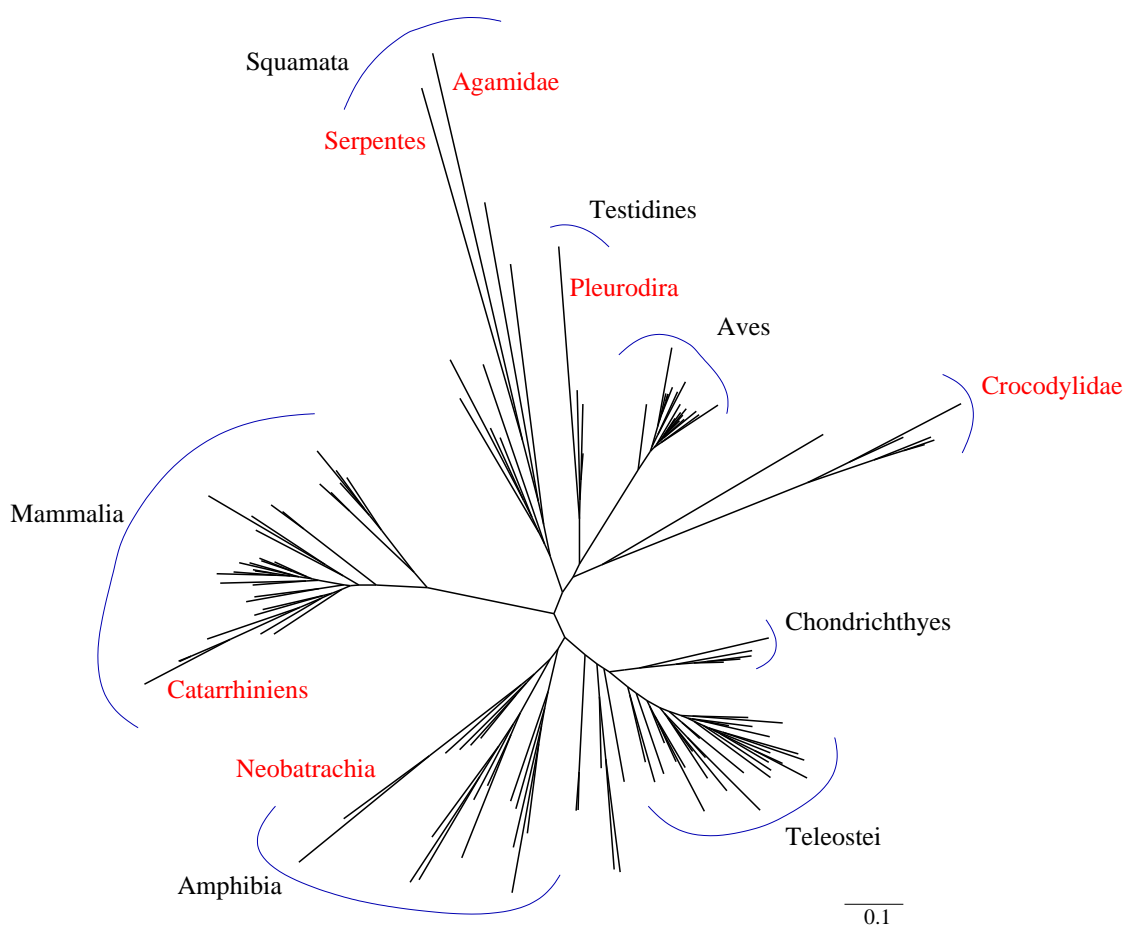


FIG. 3.1 – Estimation des longueurs de branches de la phylogénie de 124 espèces de vertébrés obtenue à partir de la concaténation des séquences protéiques des 13 gènes mitochondriaux. Estimation effectuée par maximum de vraisemblance (mtREV + Γ). Les taxons représentés en rouge sont ceux qui présentent une apparente accélération de leur taux d'évolution.

Comme on l'a vu, l'hypermutable est l'une des caractéristiques les plus visible de l'ADN mitochondrial des animaux. Il existe une série de raisons qui en font un sujet d'étude privilégié : (i) l'évolution de l'hypermutable est difficile à comprendre (chapitre 2.3.3), (ii) elle conditionne probablement la structure du génome mitochondrial (Lynch, 2006), (iii) enfin c'est l'un des caractéristiques qui explique le succès du marqueur mitochondrial dans l'étude de la biodiversité (chapitre 2.3.2).

L'ADN mitochondrial présente également un taux de mutation étonnamment variable entre lignées. Cette variation est bien visible simplement en estimant les longueurs de branches d'une phylogénie réalisée sur la base de marqueurs mitochondriaux. Un exemple particulièrement frappant concerne l'ADN mitochondrial des vertébrés (Figure 3.1). Certaines lignées, comme les crocodiles, certains lézards, batraciens, serpents ainsi que les primates anthropoïdes, semblent avoir des branches qui se dilatent, alors que d'autres groupes, comme les oiseaux, semblent au contraire subir une contraction de leurs branches. Pour certains de ces exemples, des causes sélectives peuvent être invoquées pour expliquer les augmentations du taux d'évo-

lution. C'est le cas, par exemple, des primates anthropoïdes, groupe auquel appartiennent les Catarrhiniens (Grossman *et al.*, 2004) ou, plus récemment, des serpents (Castoe *et al.*, 2008).

Cependant, ces variations ont, sans aucun doute, également une base mutationnelle. Il suffit pour s'en convaincre de constater que des variations de taux d'évolution entre lignées apparaissent aussi au niveau des divergences synonymes, impliquant donc des variations du taux de mutation. Quantifier et comprendre les déterminants de ces variations du taux de mutation mitochondrial est un bon moyen de s'attaquer à la *compréhension de l'hypermutableté mitochondriale*.

Chapitre 4

Etat des lieux des connaissances

Les hypothèses expliquant les variations du taux de mutation mitochondrial entre lignées animales peuvent se répartir en deux catégories : celles que l'on peut qualifier de répllication-indépendantes, c'est à dire qui ne nécessitent pas un changement de qualité de la fidélité de la répllication mitochondriale pour engendrer des variations de taux d'évolution, et au contraire, celles que l'on peut qualifier de répllication-dépendantes, où les changements de taux de mutation entre espèces sont sous le contrôle du mécanisme de répllication mitochondriale.

4.1 Les hypothèses répllication-indépendantes

La littérature évolutionniste est dominée par deux hypothèses principales expliquant les variations de taux de mutation entre lignées. La première de ces hypothèses est celle du *temps de génération*. Pour être transmise à la génération suivante et, *a fortiori*, être observable lors de la comparaison entre séquences homologues d'ADN, la mutation doit se produire dans les cellules germinales des organismes, c'est-à-dire dans les cellules destinées à la production de gamètes. Ces mutations sont différentes de celles qui se produisent dans les cellules de la lignée somatique qui disparaîtront à la mort de l'organisme qui les porte. Cette dichotomie est importante pour la compréhension générale des mécanismes contrôlant le taux de mutation et particulièrement dans le cas de l'hypothèse du temps de génération. Cette hypothèse est basée sur la supposition que, pour une période de temps donnée, le nombre de répllications des cellules germinales dépend de la fréquence des générations. Cette hypothèse prédit l'existence d'une corrélation négative entre le temps de génération et le taux d'évolution. Deux espèces peuvent avoir le même taux d'évolution *par génération* (si elles présentent le même nombre de répllications cellulaires de la lignée germinale par génération) mais des taux d'évolution *par an* qui sont différents (si elles présentent des temps de générations différents). La relation entre taux d'évolution et temps de génération peut être modifiée si le nombre de répllications des cellules germinales par génération est différent entre les espèces. Par exemple, elle pourrait être diminuée si le nombre de répllications des cellules germinales augmente avec le temps de génération (Chang *et al.*, 1994; Goetting-Minesky & Makova, 2006).

L'hypothèse du temps de génération, proposée à la fin des années 60 (Laird *et al.*, 1969), a été démocratisée au milieu des années 80 principalement sous l'impulsion des travaux de Wen-Hsiung Li sur l'horloge moléculaire chez les mammifères (Wu & Li, 1985; Li *et al.*, 1987; Li & Tanimura, 1987; Li *et al.*, 1990). Les premiers résultats proviennent uniquement de la comparaison entre gènes nucléaires de primates et de rongeurs (Wu & Li, 1985). Ils ont été rapidement complétés par des séquences d'artiodactyles et de carnivores. Ces comparaisons ont donné lieu à l'hypothèse de l'“hominoid slowdown” proposant un ralentissement dans la lignée des hominidés par rapport aux autres mammifères et qui s'accroît dans la lignée humaine par rapport aux autres grands singes (Li & Tanimura, 1987). Même si certains résultats restent inexplicables, comme le faible taux d'évolution des lémuriers et des lagomorphes (Li *et al.*, 1987), la vision globale des taux d'évolution chez les mammifères, avec rongeurs > artiodactyles et carnivores > primates, reste largement en accord avec l'hypothèse du temps de génération (Li *et al.*, 1990).

Tous ces travaux ont été réalisés sur la base de résultats obtenus avec des gènes nucléaires. Cependant, l'hypothèse du temps de génération implique un effet équivalent à la fois pour le génome mitochondrial et le génome nucléaire. Il s'agit, en effet, simplement d'un effet d'échelle entre le taux par an et par génération indépendant du génome concerné (voir aussi chapitre 6.4.2). Cet impact équivalent entre les deux génomes est une spécificité de l'hypothèse du temps de génération, les deux autres hypothèses évoquées plus bas étant plus fortement mitochondrie-spécifiques.

Cette hypothèse simple et bien intégrée dans le cadre théorique du modèle presque-neutre (Ohta, 1987; Kimura, 1983; Chao & Carr, 1993), a souvent été reprise dans la littérature évolutionniste, principalement dans des études utilisant les mammifères comme modèles (Catzeflis *et al.*, 1987; Ohta, 1993; Li *et al.*, 1996; Bromham *et al.*, 1996; Pesole *et al.*, 1999). Elle a aussi été invoquée pour expliquer les variations de taux d'évolution au sein des oiseaux (Mooers & Harvey, 1994), de manière moins évidente, pour expliquer le ralentissement du taux d'évolution des acipentiformes (esturgeons) comparés à d'autres téléostéens (Krieger & Fuerst, 2002) ou, entre les plantes annuelles et pérennes (réfutée par Whittle & Johnston, 2003; Smith & Donoghue, 2008).

Facile à mesurer pour des espèces à générations non-chevauchantes, le temps de génération est plus difficile à mesurer chez les espèces à générations chevauchantes comme les mammifères. Dans la plupart des études, le temps de génération est approximé par l'âge de la maturité sexuelle (femelle dans le cas de l'ADN mitochondrial) (Mooers & Harvey, 1994), ou l'âge moyen de naissance de la première portée (Bromham *et al.*, 1996; Welch *et al.*, 2008).

La deuxième hypothèse invoquée pour expliquer les variations des taux d'évolution est l'hypothèse du *taux métabolique*. Cette hypothèse propose une relation entre le taux de mutation et la production de *radicaux libres oxydatifs* (ou ROS pour *Reactive Oxygen Species*). Les ROS sont des sous-produits de la respiration très mutagènes (voir chapitre 2.3.3). Si la quantité de ROS augmente avec le taux métabolique masse-spécifique alors il devrait exister une relation entre le taux métabolique masse-spécifique et taux de mutation à travers la production de ROS. Cette relation devrait tout particulièrement être présente pour le génome mitochondrial,

très proche physiquement de la production de ROS.

Cette hypothèse a été proposée dans les années 90 par Andrew Martin et Stephen Palumbi (Martin *et al.*, 1992). Les travaux de Martin *et al.* s'appuient effectivement sur l'utilisation d'alignements mitochondriaux. Ces auteurs ont ajouté des séquences mitochondriales de Chondrichthyens (requins et raies) au jeu de données de mammifères déjà existant (Martin & Palumbi, 1993; Martin, 1995, 1999). Les requins présentent un taux d'évolution plus faible que les primates pour un temps de génération comparable mais un métabolisme masse-spécifique plus faible (Martin *et al.*, 1992). Ce taux métabolique plus faible s'explique par le fait que les requins sont des organismes hétérothermes (ou poïkilothermes), qui ne régulent pas leur température corporelle alors que les oiseaux et les mammifères sont homéothermes (ou endothermes). Chez les homéothermes, la capacité à réguler sa température interne provient de la capacité de nos mitochondries à transformer les calories en chaleur lors du processus de *découplage* de la phosphorylation oxydative. La chaîne respiratoire est à la fois le lieu de production ATP et de la production de chaleur. Un changement dans l'efficacité du *couplage* mitochondrial détermine l'allocation relative des calories entre l'ATP (pour l'activité) et la chaleur (pour le maintien de la température corporelle) (Wallace, 2005). Les homéothermes possèdent, en outre, des mitochondries plus actives et plus nombreuses par unité de volume de tissu que les hétérothermes (Lane, 2005). Issue de la comparaison hétérotherme-homéotherme, cette hypothèse a été confirmée par d'autres études intra-homéothermes, notamment chez les procellariiformes (oiseaux marins du groupe des albatros et petrels, Nunn & Stanley, 1998) et des colibris (Bleiweiss, 1998a,b), ou intra-hétérothermes comme chez des poissons (cyprinidés) Nord-Américains (Estabrook *et al.*, 2007). Cette hypothèse a récemment été remise sur le devant de la scène par l'étude de Gillooly *et al.* (2005). Ces auteurs ont utilisé une formule décrivant la variation du métabolisme masse-spécifique en fonction de la masse et de la température (Gillooly *et al.*, 2001) pour prédire les variations du taux d'évolution également en fonction de la masse et de la température (Gillooly *et al.*, 2005). La force de ce travail réside dans cette formulation *a priori* d'un modèle statistique qui s'ajuste parfaitement aux données expérimentales obtenues dans divers groupes taxonomiques (hétérothermes et homéothermes).

Bien soutenue par des données récentes et issue d'un mécanisme directement en lien avec la biologie mitochondriale, l'hypothèse du taux métabolique est pourtant l'une de celles qui a le plus de points faibles. D'une part, certaines données expérimentales ne sont pas en accord avec cette théorie, comme le taux d'évolution apparemment plus faible des oiseaux par rapport aux mammifères (Mindell *et al.*, 1996, voir chapitre 6.1), ou l'absence d'un effet du taux métabolique masse-spécifique sur le taux d'évolution à l'échelle des métazoaires (Lanfear *et al.*, 2007). D'autre part, l'hypothèse du taux métabolique peut être critiquée conceptuellement. Elle se base, en effet, sur plusieurs suppositions peu vraisemblables. L'hypothèse du taux métabolique suppose, par exemple, une relation directe entre taux métabolique masse-spécifique et production de ROS. Cependant les études provenant de la physiologie comparative nous apprennent que les espèces peuvent avoir, à taux métabolique constant, une production de ROS différente (*e.g.* Herrero & Barja, 1998 et Ku & Sohal, 1993). Cet ajustement dans la production de ROS provient de l'efficacité des

protéines du couplage de la chaîne respiratoire (Rousset *et al.*, 2004; Speakman, 2005). De la même manière, la relation entre quantité de ROS produite et dégâts engendrés par ces ROS peut être soumis à des ajustements entre les organismes. Ces ajustements peuvent, par exemple, être effectués via la production d’anti-oxydants qui détoxifient les ROS (Sun *et al.*, 2002). Enfin, l’hypothèse du taux métabolique invoque la proportionnalité entre la production de ROS dans les tissus somatiques et dans la lignée germinale. Cependant, bien qu’il y ait peu d’évidences expérimentales, les mitochondries de la lignée germinale sont probablement inactives (Allen, 1996) et donc peu ou pas productrices de ROS. Une autre possibilité serait que les ROS produits dans le soma aient un impact sur la lignée germinale. Cette hypothèse est très improbable du fait de la forte réactivité des ROS, associée à la présence de nombreux anti-oxydants qui leur imposent une durée de vie courte (Turrens, 2003).

L’ensemble de ces remarques jette un doute sur la possibilité de l’existence d’une relation forte et générale entre taux métabolique masse-spécifique et taux d’évolution, comme invoquée par Gillooly *et al.* (2005), sans pour autant réfuter que l’augmentation du taux métabolique puisse expliquer une part des grandes variations de taux d’évolution entre les hétérothermes et les endothermes.

Enfin, une autre hypothèse, rarement invoquée, propose que les réarrangements chromosomiques mitochondriaux pourraient promouvoir de brusques augmentations du taux de mutation par un changement de position des régions génomiques par rapport à l’origine de réplication. La position par rapport à l’origine est connue pour influencer fortement la composition en base de l’ADN mitochondrial, *e.g.* (Faith & Pollock, 2003). Cette hypothèse a été proposée suite à l’étude d’un groupe où les réarrangements chromosomiques mitochondriaux sont fréquents : les arthropodes. Chez les arthropodes, le nombre de réarrangements est apparemment corrélé au taux d’évolution (Xu *et al.*, 2006).

4.2 Taux d’évolution et théorie du vieillissement mitochondrial : une hypothèse réplication-dépendante

Enfin, on peut formuler une dernière hypothèse explicative à partir de la théorie du *vieillessement mitochondrial*. Au milieu des années 50, Denham Harman propose la “théorie des radicaux-libres” du vieillissement humain. Cette théorie postule que la production de ROS générée dans les cellules produit des dommages cumulatifs qui provoquent le vieillissement (Harman, 1957). Cette théorie est notamment soutenue par la corrélation négative entre la fréquence des dommages oxydatifs et la durée de vie maximale chez les mammifères (Barja & Herrero, 2000). L’interprétation actuelle de la théorie de Harman fait le lien entre les dommages oxydatifs et la respiration mitochondriale, qui est le principal producteur de ROS. La théorie du vieillissement mitochondrial postule que les mutations, qui surviennent sur l’ADN mitochondrial, en affectant l’efficacité des protéines de la chaîne respiratoire, entraînent l’augmentation du taux de production de ROS, qui à leur tour participent à la mutation de l’ADN mitochondrial. Cette chaîne de retro-action positive peut, à terme, provoquer le vieillissement de l’organisme (Figure 4.1).

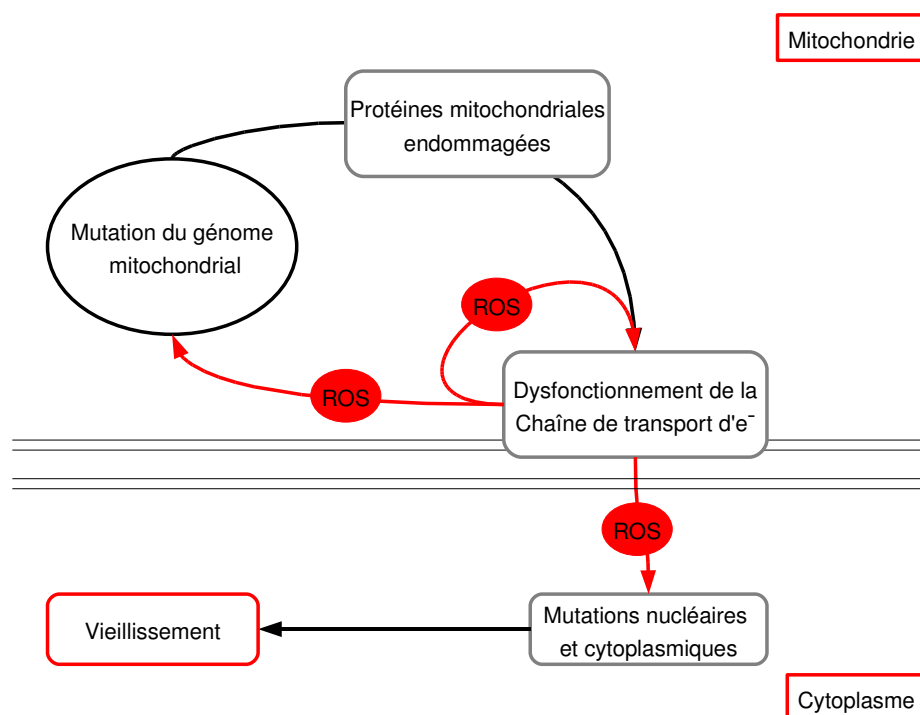


FIG. 4.1 – Schéma représentant le mécanisme de la *théorie du vieillissement mitochondrial*. La production de radicaux libres oxydatifs (ROS) par la chaîne respiratoire augmente si le génome mitochondrial muté produit des protéines défectueuses. Les ROS provoquent à leur tour des mutations sur le génome mitochondrial qui produit des protéines encore plus défectueuses. Il peut en résulter une boucle de rétro-action positive qui provoquent le vieillissement.

La variation dans les taux de mutation pourrait ainsi provenir de différences dans la fidélité de la réplication de l'ADN mitochondrial. En effet, même sans évoquer la chaîne de rétro-action positive, une réplication du génome mitochondrial peu fidèle suffit à expliquer le vieillissement mitochondrial. Il existe un compromis fondamental entre le coût investi dans les mécanismes contrôlant la fidélité de la réplication et le bénéfice à éviter des mutations délétères (*e.g.* Dawson, 1998, 1999; Furió *et al.*, 2005). La sélection naturelle doit maintenir un niveau de mutation "acceptable" pour les populations mais on peut très bien imaginer que ce niveau soit différent entre les organismes. Par exemple, dans le notre cas, le taux de mutation acceptable devrait être négativement corrélé à la longévité maximale.

Sachant l'implication potentielle de l'ADN mitochondrial dans le vieillissement, il est possible qu'un ajustement du taux de mutation mitochondrial puisse exister en fonction de la longévité des organismes.

Cette liaison entre taux de mutation mitochondrial et longévité est soutenue par plusieurs constatations

1. Le génome mitochondrial des tissus somatiques semble accumuler des mutations au cours de la vie de certains mammifères (Corral-Debrinski *et al.*,

- 1992; Schwarze *et al.*, 1995; Khaidakov *et al.*, 2003).
2. L'augmentation artificielle du taux de mutation mitochondrial chez une souris transgénique, avec une polymérase- γ moins fidèle, provoque un vieillissement prématuré (Trifunovic *et al.*, 2004).
 3. Le nombre de répétitions de plusieurs nucléotides identiques, correspondant à des cibles privilégiées pour la mutation, est négativement corrélé à la longévité (Samuels, 2004; Khaidakov *et al.*, 2006). Ce résultat suggère l'existence d'une sélection pour réduire le taux de mutation mitochondrial chez les espèces longévives.

4.3 Les enjeux

Cet état des lieux a permis de montrer que trois hypothèses principales sont en compétition, sans être forcément exclusives, pour expliquer les variations du taux de mutation mitochondrial des animaux. Le point de départ incontournable de toute analyse destinée à évaluer ces hypothèses consiste à estimer précisément et de manière non-biaisée le taux de mutation mitochondrial pour un maximum d'espèces. Dans le chapitre suivant nous exposerons les raisons qui nous ont fait choisir les oiseaux et les mammifères comme modèles, et quelle méthode nous avons utilisée pour estimer précisément les taux de mutation. Ce type d'analyse a également d'autres intérêts que l'étude du déterminisme des variations du taux de mutation :

1. Le caractère hypermutable de l'ADN mitochondrial rend toute estimation des quantités de divergence entre espèces problématique. On peut donc se poser la question de savoir si les études précédentes, utilisant un nombre limité d'espèces, ont réussi à révéler toute la variation du taux d'évolution mitochondrial. Ceci pourra s'effectuer en utilisant la quantité importante de données génétiques mitochondriales produites ces dernières années (voir chapitre 1.2).
2. De nombreuses études d'évolution moléculaire, de génétique des populations et de phylogéographie se basent sur des approximations concernant la valeur absolue du taux d'évolution mitochondrial. Par exemple, les études de phylogéographie utilisent souvent des taux d'évolution obtenus dans des groupes taxonomiques plus ou moins éloignés, (voir Lovette, 2004 pour une revue chez les oiseaux). Dans quelle mesure cette approximation est-elle réaliste? De même, si le débat sur la réalité de l'horloge moléculaire semble réglé chez les mammifères, il est toujours d'actualité en ce qui concerne les oiseaux (Garcia-Moreno, 2004; Ho, 2007). Si certaines études ont déjà montré qu'il existe une variation dans le taux d'évolution de l'ADN mitochondrial des oiseaux (Pereira & Baker, 2006; Lovette, 2004), la célèbre calibration des "2% de divergence par million d'années", publiée en 1979 (Brown *et al.*, 1979; *e.g.* Bermingham *et al.*, 1992; Weir & Schluter, 2007), est défendue par deux analyses récentes (Päckert *et al.*, 2007; Weir & Schluter, 2008). Cette persistance à utiliser l'horloge moléculaire chez les oiseaux a peut-être une réalité biologique qu'il est important de clarifier.

Chapitre 5

Modèles biologiques et approche phylogénétique

5.1 Pourquoi se focaliser sur les oiseaux et les mammifères ?

Les trois hypothèses principales proposées précédemment sont testables à travers l'utilisation d'un trait d'histoire de vie comme marqueur de chaque hypothèse : maturité sexuelle femelle, utilisée comme approximation du temps de génération, masse corporelle, reliée au métabolisme par une relation d'allométrie (Gillooly *et al.*, 2001) et longévité maximale. Cependant, il est statistiquement délicat de discriminer entre les différentes hypothèses étant donnée la corrélation importante entre ces variables (Tableau 5.1). Une solution consiste à maximiser le nombre d'espèces utilisées afin d'avoir le plus de puissance statistique et d'utiliser des espèces aux traits d'histoire de vie les plus contrastés possibles et bien documentés, pour lesquelles des données génétiques sont également disponibles. Dans ces conditions, il s'est rapidement imposé que les mammifères et les oiseaux constituaient les meilleurs modèles biologiques.

Ainsi, de par leur statut de grands vertébrés charismatiques et phylogénéti-

Traits d'histoire de vie	LM	MC	MB	MS
Longévité maximale (LM)	–	0,58	0,42	0,70
Masse corporelle (MC)	0,71	–	0,89	0,62
Métabolisme basal (MB)	0,66	0,97	–	0,28
Maturité sexuelle (MS)	0,77	0,67	0,64	–

TAB. 5.1 – Coefficient de corrélation entre les principaux traits d'histoire de vie chez les vertébrés (au-dessus de la diagonale) et les mammifères (en-dessous). Les variables sont transformées logarithmiquement. La longévité maximale est exprimée en années ; la masse corporelle en grammes ; le métabolisme basal en Watts et la maturité sexuelle (femelle) en jours. Données issues de la base *AnAge*, (de Magalhaes *et al.*, 2005, <http://genomics.senescence.info/>).

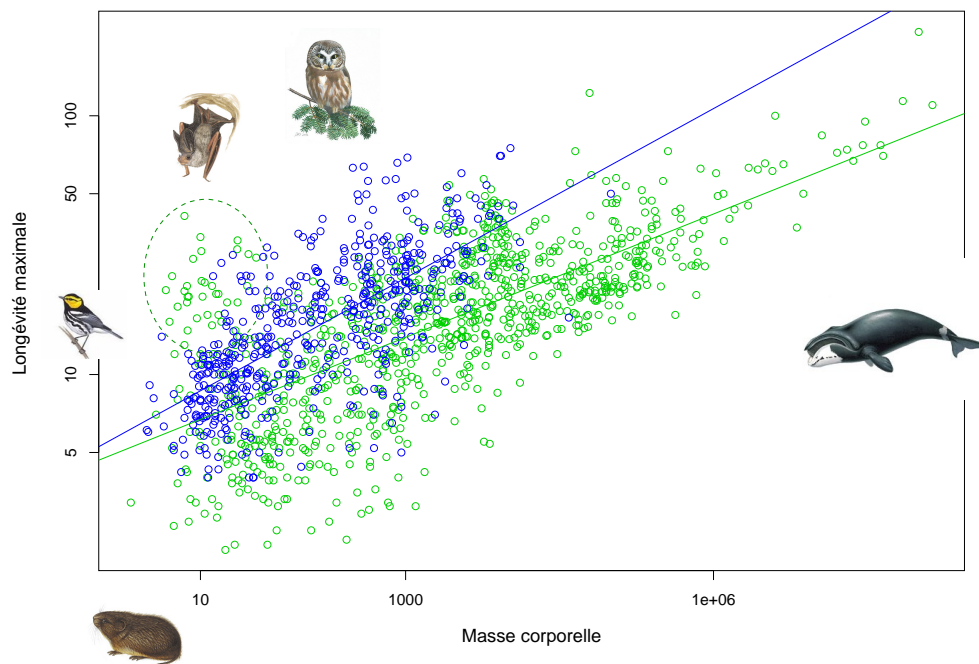


FIG. 5.1 – Relation entre la masse (en grammes) et la longévité maximale (en années) chez les oiseaux, points bleus, et les mammifères, points verts. Le cercle pointillé vert représente approximativement la position des chiroptères, qui présentent en moyenne des longévités élevées par rapport à leur masse.

quement proches de nous, ils sont très largement étudiés, ce qui permet de disposer à la fois de nombreux traits d'histoire de vie et de beaucoup de données génétiques. Ils présentent aussi une grande gamme de variations de traits d'histoire de vie entre espèces. Par exemple, le plus petit mammifère, le pachyure étrusque (*Sunctus etruscus*), pèse 2 à 3 grammes à l'âge adulte, alors que le plus grand, la baleine bleue (*Balaenoptera musculus*), a été enregistré à plus de 100 tonnes. De manière intéressante, ces groupes présentent des espèces avec des traits d'histoire de vie qui s'éloignent des effets d'allométrie classiques. Autrement dit, certaines espèces vivent plus longtemps que la valeur prédite par leur masse corporelle, comme certaines chauve-souris ou certains primates, et au contraire d'autres espèces vivent moins longtemps, comme certains cetartiodactyles ou eulipotyphles (figure 5.1). Enfin, la comparaison oiseaux-mammifères permet également de révéler un contraste intéressant. Les oiseaux présentent un taux métabolique masse-spécifique particulièrement élevé - 1,5 à 2,5 fois plus élevé qu'un mammifère de taille comparable - probablement à cause de la forte demande énergétique induite par le vol (Holmes *et al.*, 2001). Paradoxalement, les oiseaux sont des espèces particulièrement longévives (comparés à des mammifères de masse similaire). Les oiseaux ont des durées de vie trois fois plus longues que des mammifères de taille comparable, exceptées les chauve-souris qui présentent également des longévités élevées (Holmes *et al.*, 2001; Holmes & Ottinger, 2003; Brunet-Rossinni & Austad, 2004 et figure 5.1).

5.2 Inférence du taux de mutation par une approche phylogénétique.

Likewise, rates of molecular evolution are certainly not constant over time. The pertinent question is not whether rates are constant over time but how much do rates change over time.

Jeffrey L. Thorne and Hirohisa Kishino, *Systematic Biology*
51, 2002.

Avant de présenter la méthode d'estimation des taux d'évolution, il est important de clarifier la relation entre substitution, qui correspond au remplacement d'une base par une autre dans une espèce, et mutation, qui correspond à l'apparition d'un variant génétique dans une population. Dans ce travail, nous désirons nous focaliser sur la mutation, mais c'est la substitution qui est mesurée par les méthodes phylogénétiques. Quelle est la relation entre ces deux quantités ?

5.2.1 Relation entre le taux de mutation et le taux de substitution : une mise au point conceptuelle

La méthode la plus simple pour estimer le taux de mutation consiste à séquencer des fractions du génome issues de différentes générations de lignées pures, dites d'"accumulation de mutations", auxquelles on impose des goulots d'étranglement afin de limiter l'effet de la sélection. On peut également procéder ainsi avec des individus dont la généalogie est connue. Cette méthode peut être qualifiée de *méthode directe*. L'avantage de ce procédé réside dans sa simplicité méthodologique : il suffit de diviser le nombre de mutations observées par la taille de la séquence analysée et par le nombre de générations de l'expérience pour obtenir le taux de mutation. En contrepartie, cette méthode est particulièrement lourde à mettre en oeuvre. Les mutations étant des phénomènes relativement rares, cette méthode demande de séquencer de grandes quantités d'ADN pour espérer enregistrer quelques mutations, tout en utilisant des organismes modèles avec des temps de génération courts (Denver *et al.*, 2000; Haag-Liautard *et al.*, 2007; Lynch *et al.*, 2008). Un moyen de contourner ces difficultés méthodologiques consiste à se focaliser sur des régions génomiques hypervariables comme la région de contrôle du génome mitochondrial, ce qui permet d'alléger le protocole expérimental (Howell *et al.*, 2003; Santos *et al.*, 2005).

Alternativement à cette méthode directe, il existe une méthode dite *indirecte*, ou *méthode phylogénétique*. Cette méthode consiste à utiliser l'estimation phylogénétique des divergences moléculaires entre séquences homologues pour extrapoler le taux de mutation via le taux d'évolution. La divergence moléculaire étant le produit du taux d'évolution par le temps de divergence, le taux d'évolution peut être estimé si la date de divergence est connue. Cette date de divergence est typiquement obtenue à partir d'informations paléontologiques ou biogéographiques permettant d'estimer l'âge de la séparation entre deux lignées. La méthode indirecte, si elle présente l'avantage d'utiliser l'information moléculaire accumulée durant des millions d'années et d'être plus facile à mettre en oeuvre sur de nombreuses espèces,

comporte également quelques problèmes méthodologiques qu'il est important de clarifier. Le nombre de substitutions estimé entre deux séquences homologues est la somme des mutations neutres, avantageuses et délétères fixées au cours du temps. Les probabilités de fixation de ces différentes classes de mutations ne dépendent pas des mêmes paramètres. Le nombre de mutations sélectionnées qui se fixent dépend du taux de mutation et du nombre de générations, mais également de la taille des populations et de la valeur sélective de ces mutations. Au contraire, le nombre de mutations neutres qui se fixent ne dépend que du produit entre le taux de mutation et le temps de divergence, indépendamment de la taille des populations (voir chapitre 1.1), et ce quel que soit l'environnement génétique du locus considéré. Par exemple, la liaison génétique avec un locus sous sélection n'a aucune influence sur le taux de substitution neutre (Birky & Walsh, 1988). Le taux de substitution neutre d'un locus mesuré par une méthode phylogénétique entre deux espèces est donc égal au taux de mutation moyen du locus durant les t générations de divergence.

Cette différence de comportement entre mutations neutres et sélectionnées implique que l'estimation d'une divergence moléculaire ne permet pas nécessairement d'inférer le taux de mutation. Il faut utiliser uniquement des mutations neutres qui nous renseignent directement sur le taux de mutation. Un bon moyen de réaliser cette opération consiste à se concentrer sur les parties du génome codant pour des protéines. A l'intérieur des séquences codantes, on peut supposer que les mutations qui entraînent un changement dans la séquence protéique, appelées *mutations non-synonymes*, sont susceptibles d'avoir un effet sur la valeur sélective du mutant, alors que les mutations n'entraînant pas de changement dans la séquence protéique, appelées *mutations synonymes*, sont neutres vis à vis de la sélection. Certaines méthodes utilisant des modèles de substitutions au niveau des codons permettent d'estimer directement les divergences synonymes et non-synonymes (Goldman & Yang, 1994). On peut également considérer comme des substitutions synonymes les substitutions qui ont lieu en troisième position des codons. Cette approximation est particulièrement vraie pour le génome mitochondrial où les transitions ($A \leftrightarrow G$, $C \leftrightarrow T$) en troisième position sont exclusivement synonymes.

5.2.2 Inférence phylogénétique du taux d'évolution : le tour de force des méthodes Bayésiennes

Dans une phylogénie, les paramètres des longueurs de branches correspondent au produit des taux de substitution et des dates de divergence. L'estimation précise des longueurs de branches et des temps de divergence est primordiale pour obtenir une estimation correcte des taux de substitution. L'estimation précise de la divergence moléculaire est particulièrement problématique pour des marqueurs à évolution rapide. Le phénomène dit de *saturation*, correspondant à des substitutions multiples se produisant sur une séquence au même site et dans une seule branche, peut amener à la sous-estimation de la divergence (Figure 5.2). Ce problème de saturation est connu des phylogénéticiens moléculaires à travers la grande quantité d'homoplasies détectées dans les phylogénies réalisées en utilisant l'ADN mitochondrial (Blouin *et al.*, 1998; Springer *et al.*, 2001; Galewski *et al.*, 2006). Limiter la satura-

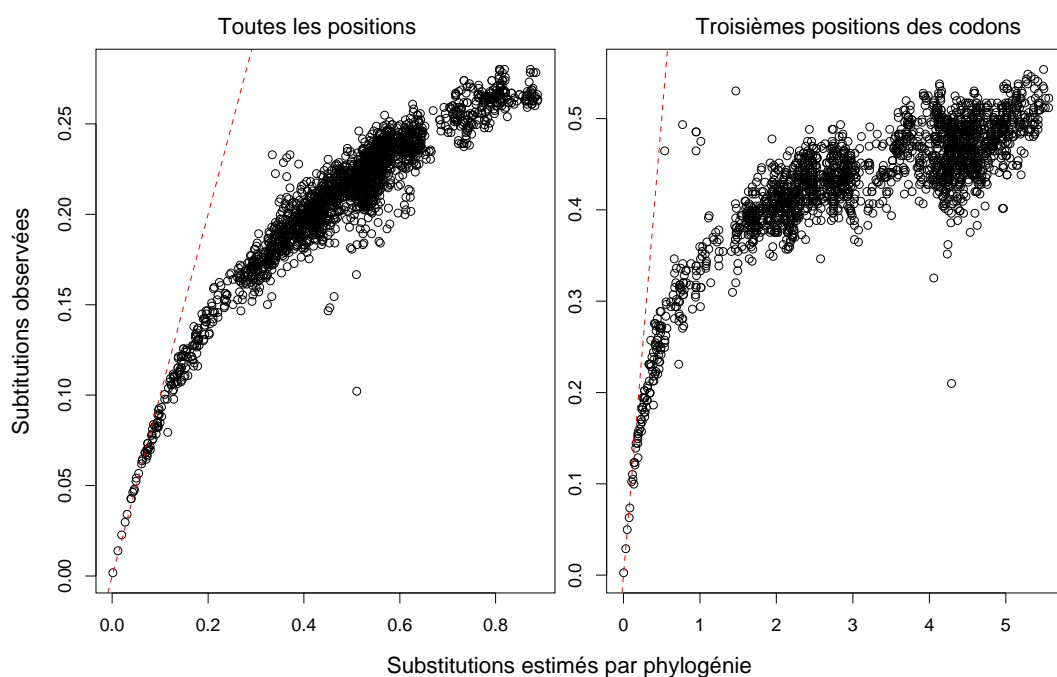


FIG. 5.2 – Nombre de substitutions observées (par site) entre paires de séquences et nombre de substitutions estimées par maximum de vraisemblance (TN93 + Γ). L'alignement est constitué de toutes les positions du cytochrome *b* (à gauche) ou uniquement des 379 troisièmes positions des codons de 65 espèces de Primates. La ligne rouge pointillée correspond à la droite d'équation $y = x$.

tion nécessite l'utilisation de séquences peu divergentes afin de réduire la présence de longues branches.

Les dates de divergence sont, la plupart du temps, issues de points de calibration fossiles, et plus rarement, d'informations biogéographiques telles que la formation de barrières physiques comme l'isthme de Panama, ou la formation de nouveaux habitats comme des îles océaniques. Les points de calibration fossiles sont rares, d'une part du fait de lacunes du registre fossile, mais également à cause de la difficulté méthodologique inhérente à la "conversion" de l'information paléontologique en calibrations phylogénétiques. Par exemple, si l'assignation d'un fossile comme appartenant à un groupe actuel permet de fixer une borne inférieure à une divergence, l'obtention d'une borne supérieure est beaucoup plus problématique et repose sur l'absence de fossiles (voir par exemple Müller & Reisz (2005) et Benton & Donoghue (2007) pour des discussions de ces problèmes).

La rareté des points de calibrations fossiles entraîne un problème récurrent de la datation moléculaire qui consiste à extrapoler les dates de divergence de l'ensemble des noeuds d'une phylogénie à partir de quelques points de calibrations (figure 5.3). L'utilisation de l'hypothèse de l'horloge moléculaire reste la méthode la plus simple. Elle consiste à faire l'hypothèse que les taux d'évolution sont constants dans l'ensemble de la phylogénie. Ainsi, il suffit d'avoir une seule date de divergence connue pour obtenir un taux d'évolution applicable à l'ensemble des divergences de la phylogénie. Cette méthode n'a pas d'intérêt pour notre étude car elle implique que les taux d'évolution ne varient pas entre espèces. On peut toutefois appliquer cette hypothèse en utilisant uniquement des couples d'espèces pour lesquelles la date de divergence est connue (figure 5.3.A). Il suffit de diviser la divergence moléculaire par la date de divergence et de diviser à nouveau par 2 pour obtenir un taux d'évolution espèce spécifique, identique pour les deux espèces selon l'hypothèse d'horloge moléculaire. Dans ce cas, seule la comparaison entre couples d'espèces nous apporte des informations. Cependant, cette méthode comporte trois problèmes : (i) on a vu que les dates de divergence étaient des données rares, en conséquence ; on ne pourra analyser qu'une petite fraction du jeu de données ; (ii) les dates de divergence fossiles comportent souvent des incertitudes qu'il est difficile d'intégrer avec cette méthode ; (iii) enfin, dans le cas d'un changement de taux de d'évolution entre les deux espèces du couple, le taux d'évolution estimé en appliquant l'hypothèse d'horloge moléculaire aboutira à la sous-estimation ou à la surestimation du taux d'évolution (figure 5.3.B).

L'approche phylogénétique permet de traiter simultanément l'ensemble des espèces d'un groupe et d'inférer toutes les divergences dans un cadre phylogénétique. Reste le problème de l'extrapolation des rares dates de divergence à l'ensemble des noeuds d'une phylogénie (figure 5.3.C). Ce tour de force a été rendu possible par l'utilisation de méthodes Bayésiennes. Ces méthodes proposent de *décomposer* les longueurs de branches en taux d'évolution et dates de divergence. Ainsi pour n'importe quel jeu de données X , on peut calculer sa vraisemblance $f(X|T, R, \theta)$, connaissant les dates de divergence T , les taux d'évolution R et les paramètres du modèle de substitution θ . L'avantage des modèles Bayésiens est de pouvoir proposer un "modèle d'évolution des taux d'évolution" *a priori* qui permettra de modéliser les changements de taux entre lignées (Figure 5.3.D). Dans leurs travaux remarquablement précurseurs, Thorne et Kishino ont proposé un modèle d'évolution des taux

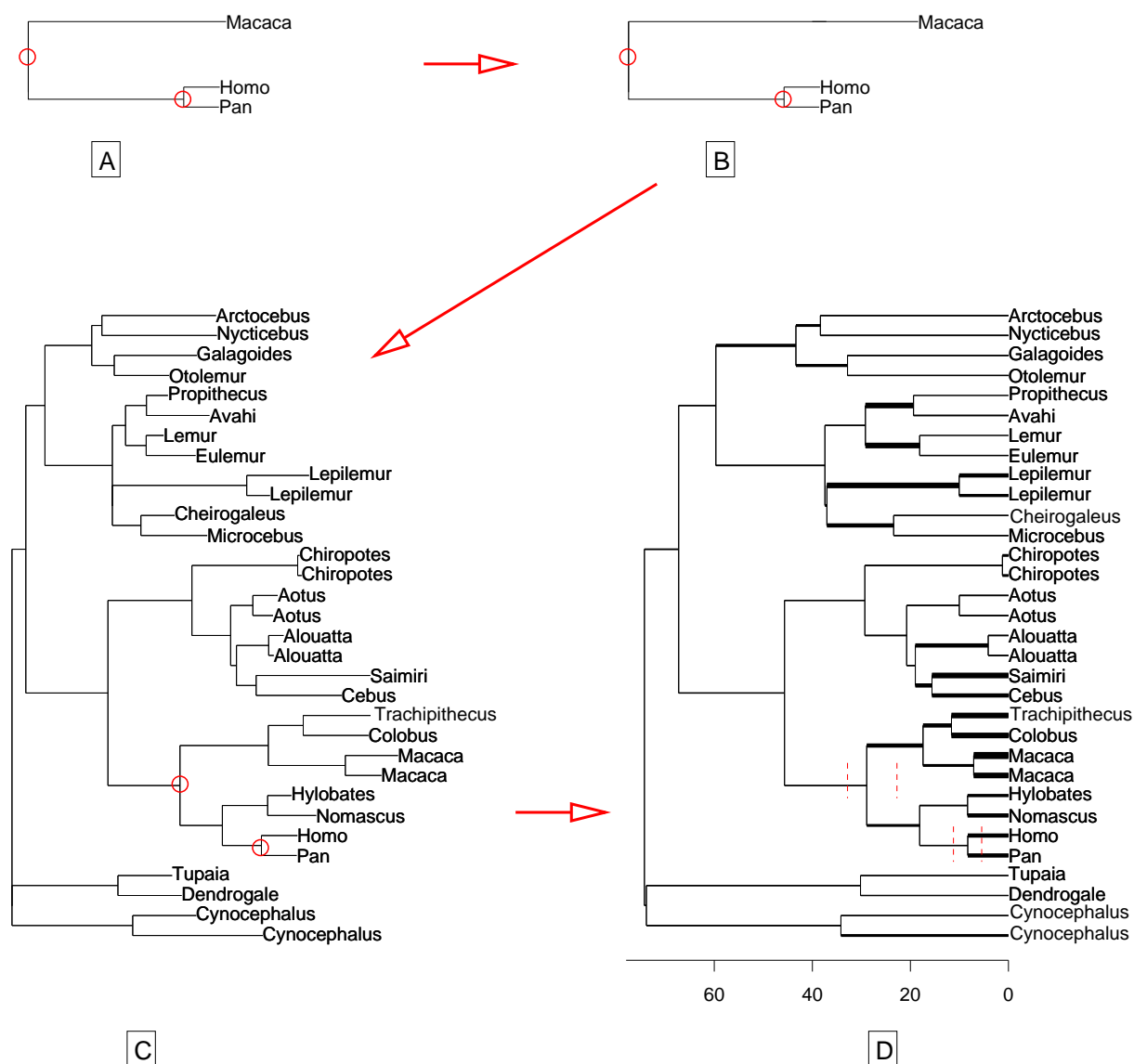


FIG. 5.3 – Les avantages de l’approche phylogénétique. Figure A : Approche classique consistant à utiliser des paires d’espèces avec une date de divergence connue, indiquée par les cercles rouges. Figure B : Changement de taux d’évolution dans la paire. Figure C : Phylogénie des primates, les longueurs de branches illustrent les variations de taux d’évolution, les deux cercles rouges indiquent les dates de divergence connues. Figure D : Extrapolation des dates de divergence, visibles grâce à l’échelle en-dessous de l’arbre, et des taux d’évolution, variables entre lignées, la vitesse du taux d’évolution par branche étant proportionnelle à l’épaisseur de chaque branche. Noter la bonne correspondance entre les longues branches des macaques ou du clade *Saimiri-Cebus* (figure C) et les branches épaisses représentant des taux d’évolution élevés (figure D). Les incertitudes sur les points de calibrations sont représentées par les traits pointillés rouges. Réalisé avec le programme MULTIDIVTIME avec un jeu de données protéiques (mtREV + Γ).

qui fait l’hypothèse que les taux d’évolution sont phylogénétiquement autocorrélés (Thorne *et al.*, 1998; Kishino *et al.*, 2001; Thorne & Kishino, 2002). Si on considère un couple de branches respectivement ancestrale et descendante, le taux d’évolution de la branche descendante correspond à un taux d’évolution échantillonné dans une distribution log-normale de moyenne égale au taux d’évolution de la branche ancestrale, et de variance proportionnelle au temps écoulé entre les deux branches, multiplié par un paramètre d’autocorrélation (ν) (Thorne *et al.*, 1998). Dans ce modèle, la branche descendante possède un taux d’évolution particulier qui est d’autant plus proche de celui de la branche ancêtre que cette branche est courte, et que ν est faible. Une valeur de ν importante correspond à peu d’autocorrélation ; au contraire, une valeur de ν faible correspond à beaucoup d’autocorrélation. A la limite, $\nu = 0$ correspond à l’horloge moléculaire, où toutes les branches ont le même taux. Ce modèle a été implémenté dans le logiciel MULTIDIVTIME (Figure 5.3). Cette méthode fournit un taux évolution par branche. Nous avons choisi de prendre le taux d’évolution de la branche menant au taxon terminal comme approximation du taux d’évolution de l’espèce. L’autocorrélation des taux d’évolution est biologiquement vraisemblable dans la mesure où les espèces issues d’une spéciation héritent bien du taux de mutation de l’espèce ancestrale. En revanche, on peut s’attendre à ce que cette méthode modélise mal les changements brusques de taux d’évolution au sein d’une phylogénie ayant des taux globalement peu variables. Dans le cadre de l’amélioration de cette méthode, plusieurs méthodes Bayésiennes ont récemment été développées pour intégrer des modèles d’évolution des taux différents (Drummond *et al.*, 2006; Rannala & Yang, 2007; Lepage *et al.*, 2007). Principalement développées pour obtenir les dates de divergence des événements de spéciation, ces méthodes permettent également d’intégrer des incertitudes sur les points de calibrations.

En résumé, l’estimation des taux de mutation par une méthode phylogénétique soulève trois problèmes :

1. Nous devons utiliser des mutations neutres afin de bénéficier de l’équivalence entre taux de substitution et taux de mutation.
2. Le taux de mutation élevé de l’ADN mitochondrial des animaux nous oblige à utiliser des divergences modérées entre espèces afin de limiter la saturation.
3. La rareté des points de calibration oblige à utiliser des divergences importantes pour incorporer certains points de calibrations anciens.

Conscients de ces trois problèmes, nous avons proposé une approche en trois étapes qui utilise la double nature (ADN et Acide aminé) des séquences codant pour des protéines.

5.2.3 Diviser pour mieux régner : une approche découplée

Reprenant les problèmes méthodologiques évoqués lors des deux chapitres précédents (5.2.2, 4.1), nous savons que le jeu de données idéal doit remplir deux critères : (i) rassembler le maximum d’espèces possibles afin de maximiser la variance des traits d’histoires de vie, mais également “couper” les longues branches et éviter la saturation et (ii) contenir le maximum de divergences pour lesquelles des

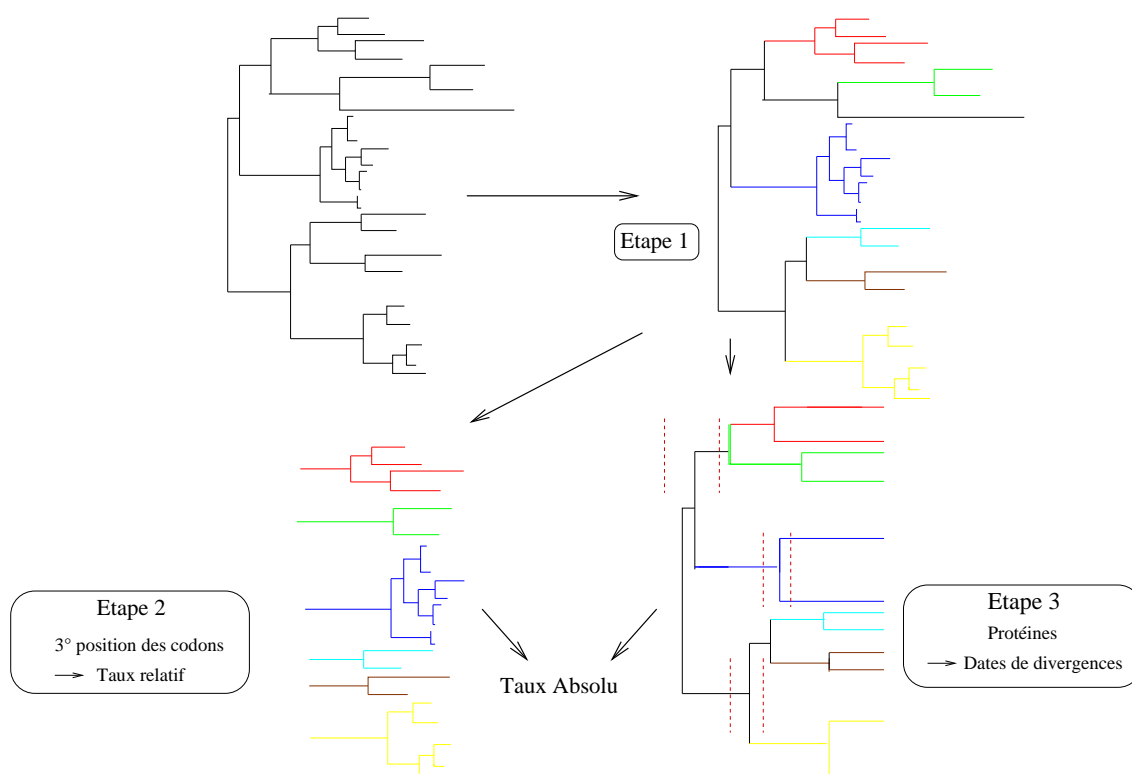


FIG. 5.4 – Schématisation de la procédure d'estimation des taux d'évolution du cytochrome *b*. L'étape 1 consiste à sélectionner des groupes taxonomiques peu divergents. Puis à estimer le taux de substitution relatif des troisièmes positions des codons à l'intérieur de chacun de ces groupes (étape 2) et estimer les dates de divergence à partir des séquences protéiques et de points de calibrations (schématisés par les bornes rouges en pointillées, étape 3). Et enfin, à utiliser ces dates de divergence pour convertir les taux relatifs en taux absolus.

points de calibration fossiles sont disponibles. Nous avons choisi d'utiliser un jeu de données construit exclusivement à partir du gène mitochondrial le plus utilisé chez les oiseaux et les mammifères : le cytochrome *b* (*cytb*) (voir Nabholz *et al.*, 2008c,b). Nous avons obtenu 1938 espèces de mammifères (procédure effectuée en fin 2006) et 1571 espèces d'oiseaux (fin 2007) pour lesquelles le *cytb* complet est disponible. Ce nombre d'espèces est tout à fait remarquable et montre, à nouveau, à quel point le marqueur mitochondrial est largement utilisé. Ainsi, le jeu de données comporte plus d'une espèce de mammifères sur trois, $\sim 36\%$ des 5416 espèces de mammifères connues (Wilson & Reeder, 2005), et près d'une espèce d'oiseaux sur cinq, $\sim 16\%$ des 9934 espèces d'oiseaux recensées (Clements, 2007).

Nous avons choisi de tirer avantage de la dynamique évolutive “découplée” entre la variation non-synonyme (pour dater) et synonyme (pour mesurer les taux). Nous proposons une procédure en trois étapes schématisée dans la figure 5.4 (Nabholz *et al.*, 2008c).

ETAPE 1 – Des groupes monophylétiques de séquences peu divergentes sont formés. Pour réaliser ces groupes, la taxonomie fournie par GenBank est ex-

plorée de manière récursive (super-ordre puis ordre puis sous-ordre puis famille etc...). Dans chacun des groupes taxonomiques explorés, la divergence phylogénétique moyenne entre espèces est calculée en utilisant un modèle TN93 + Γ . Si la divergence moyenne est inférieure à un seuil, fixé empiriquement à 0,4 substitution/site, la procédure récursive est stoppée et le groupe est sélectionné. Afin de maximiser le nombre d'espèces retenues, les genres dont la divergence médiane est supérieure à 0,4 substitution/site sont également sélectionnés.

ETAPE 2 – A l'intérieur de chacun de ces groupes, le taux de substitution neutre lignée-spécifique est estimé de manière relative, c'est à dire sans utiliser de point de calibration fossile, pour chacune des espèces, et en tenant uniquement compte des troisièmes positions des codons. Le taux de substitution est estimé à l'aide du programme MULTIDIVTIME, qui utilise le modèle d'évolution FN84 + Γ . Le taux d'évolution de chaque espèce est défini comme le taux d'évolution assigné par le modèle à la branche terminale menant à l'espèce.

ETAPE 3 – Enfin, une date de divergence est assignée à chaque groupe taxonomique par une datation moléculaire réalisée en utilisant les séquences protéiques, et l'ensemble du jeu de données. Le taux de substitution protéique étant beaucoup plus faible que le taux nucléotidique, la saturation est moins sensible au niveau protéique. Pour des raisons pratiques de temps de calcul, seules deux espèces par groupe taxonomique ont été utilisées pour cette datation. Les deux représentants de chaque groupe ont été sélectionnés aléatoirement en respectant la contrainte que l'ancêtre de ces deux espèces corresponde à la racine du groupe. Les topologies utilisées pour la phase de datation ont été construites manuellement pour être conformes aux phylogénies moléculaires multilocus récentes, soit Springer *et al.* (2003); Stephan *et al.* (2004); Teeling *et al.* (2005) pour les mammifères et Barker *et al.* (2004); Ericson *et al.* (2006); Slack *et al.* (2007) pour les oiseaux. Nous avons sélectionné le maximum de points de calibrations d'origine fossile, tous précédemment utilisés dans des études de datations moléculaires (tableau 5.2, Nabholz *et al.*, 2008c,b) L'obtention d'une date de divergence par groupe taxonomique permet, enfin, de convertir le taux relatif obtenu à l'étape 2 en taux absolu.

Analyse mammifère			
Divergence	Date maximale	Date minimale	Référence
Monotrema/Theria	191,1	162,5	[1]
Didelphimorphia/Australidelphian	71,2	61,5	[1]
Diprotodontian/Dasyuromorphia	74	54	[2]
Phascogale/Dasyurus	10	24	[2]
Bradypus/Dasypus	Libre	58	[3]
Feliformia/Caniformia	63,8	42,8	[1]
Equus/Rhinoceros	58	54	[4]
Hippopotamus/Cetacea	Libre	52	[4]
Pteropodidae/Rhinolophidae	60	43	[4]
Soridae/Erinaceidae	Libre	63	[4]
<i>Bos/Sus</i>	65	48	[4]
<i>Bos/Cervus</i>	23,8	16,4	[5]
<i>Pan/Homo</i>	10	6,5	[1]
<i>Macaca/Homo</i>	33,9	23	[1]
Primates/Rodentia	100,5	61,5	[1]
<i>Apodemus sylvaticus/A. mystacinus</i>	Libre	7	[6]
<i>Gerbillus/Gerbilliscus</i>	9	7	[7]
<i>Xerus/Sciurus</i>	Libre	36	[8]
<i>Cavia/Octodon</i>	31	Libre	[9]
<i>Mus/Rattus</i>	12,3	10	[1]
<i>Bolomys/Akodon</i>	5,3	3,5	[11]
<i>Neofiber/Ondatra</i>	4,3	2,2	[12]
Analyse oiseaux			
Divergence	Date maximale	Date minimale	Référence
Piciformes	Libre	32	[13]
Trogoniformes	Libre	53	[13]
Coliiformes	Libre	55	[13]
Strigiformes	Libre	60	[13]
Spheniscidae	72	62	[14]
Apodiformes	Libre	53	[13]
Trochilidae	Libre	47,5	[13]
<i>Anseranas/Anas</i>	72	66	[14-15]
Galliformes	Libre	53	[13]
Oscine/Suboscine (Passerine)	77*	55	[16-17]

TAB. 5.2 – Points de calibrations utilisés pour contraindre les datations moléculaires. Le point de calibration avec un * correspond à un point de calibration secondaire obtenu sur la base d'une diversification des passereaux correspondant à la vicariance Gondwanienne. La différence entre le nombre de points de calibration fossiles disponibles chez les oiseaux et les mammifères illustre les lacunes du registre fossile des oiseaux. Références : [1] Benton & Donoghue (2007) ; [2] Drummond *et al.* (2006) ; [3] Delsuc *et al.* (2004) ; [4] Springer *et al.* (2003) ; [5] Hassanin & Douzery (2003) ; [6] Michaux *et al.* (2002) ; [7] Neumann *et al.* (2006) ; [8] Mercer & Roth (2003) ; [9] Huchon & Douzery (2001) ; [11] Smith & Patton (1991) ; [12] Repenning (2001) ; [13] Ericson *et al.* (2006) ; [14] Slack *et al.* (2006) ; [15] Clarke *et al.* (2005) ; [16] Boles (1995) ; [17] Barker *et al.* (2004).

Chapitre 6

Variation du taux de mutation et longévité

6.1 Taux de mutation mitochondrial : Différence de comportement entre les oiseaux et les mammifères

La procédure de sélection présentée au chapitre précédent (5.2) a abouti à la création de 175 groupes (dont la divergence médiane intra-groupe est inférieure à 0,4 substitution par site) représentant **1 696** espèces chez les mammifères et 65 groupes représentant **1 571** espèces chez les oiseaux. La variété du niveau taxonomique des groupes sélectionnés suggère déjà une grande hétérogénéité des taux d'évolution. Pour les mammifères, la procédure a abouti à 103 genres, 15 sous-familles, 43 familles, 4 sous-ordres et 10 ordres. En ce qui concerne les oiseaux, nous avons choisi d'éliminer de la procédure de sélection des groupes l'ordre des Ciconiiformes et des Falconiiformes, leurs monophylies étant douteuses (Ericson *et al.*, 2006; Hackett *et al.*, 2008). La procédure a sélectionné beaucoup moins de genres que pour les mammifères : 3 genres, 40 familles, 19 ordres et 1 super-ordre (les ratites).

6.1.1 Variations extrêmes du taux de mutation du cytochrome *b* chez les mammifères

La moyenne du taux de substitution de la troisième position des codons du cytb des mammifères est égale à **0,098** substitution/site/Ma, et la médiane est égale à **0,043**. La distribution des taux de substitution estimés est approximativement de type log-normale mais la caractéristique la plus intéressante de cette distribution est sa très grande variance (figure 6.2) : les valeurs correspondant à 2.5% et 97.5% de la distribution se situent respectivement à 0,007 et 0,643 substitution/site/Ma. Ainsi même en considérant 5% des espèces présentant les taux de mutation les plus extrêmes comme des résultats aberrants, la distribution présente

Espèce 1	Espèce 2	Taux 1	Taux 2	Divergence	Référence
<i>Balaenoptera borealis</i>	<i>B. brydei</i>	0,007 (0,003-0,012)	0,007 (0,003-0,012)	6,90 ± 2,20	[1]
<i>Megaptera novaeangliae</i>	<i>Balaenoptera physalus</i>	0,008 (0,005-0,020)	0,008 (0,005-0,020)	15,3 ± 3,00	[1]
<i>Homo sapiens</i>	<i>Pan Troglodytes</i>	0,011 (0,003-0,009)	0,008 (0,004-0,010)	8,25 ± 1,75	[2]
<i>Mus indutus</i>	<i>Rattus norvegicus</i>	0,223 (0,113-0,413)	0,098 (0,021-0,195)	11,5 ± 0,80	[2]
<i>Apodemus sylvaticus</i>	<i>A. mystacinus</i>	0,220 (0,082-0,500)	0,258 (0,087-0,542)	7,90 ± 0,30	[3]
<i>Gerbillus nigeriae</i>	<i>G. nanus</i>	0,728 (0,087-1,170)	0,741 (0,446-1,074)	6,00 ± 0,92	[4]
				4,12 ± 0,90	[5]

TAB. 6.1 – Taux de substitution d’espèces particulièrement rapides (en bas) et lentes (en haut). Les dates de divergence sont tirées de la littérature citée. Le taux de substitution est exprimé en substitution/site/Ma ; entre parenthèses est exprimé l’intervalle de confiance. Références : [1] Sasaki *et al.* (2005) ; [2] Benton & Donoghue (2007) ; [3] Michaux *et al.* (2002) ; [4] Neumann *et al.* (2006) ; [5] Chevret & Dobigny (2005)

un facteur 100 de variation entre les espèces les plus lentes et les plus rapides. Pour confirmer ce résultat surprenant et s’assurer que cette variation ne provient pas de dates de divergence erronées, nous avons vérifié les estimations sur 6 paires d’espèces proches pour lesquelles la divergence est bien connue par une approche multilocus, indépendamment de cette analyse (tableau 6.1). Les trois premiers couples ont été sélectionnés parmi les espèces à évolution lente et les trois derniers parmi celles à évolution rapide. La gerbille du Balushistan (*Gerbillus nanus*) évolue à 0,74 substitution/site/Ma (C.I 0,45–1,07), soit ~ 100 fois plus rapidement que le rorqual boréal (*Balaenoptera borealis*) qui évolue à 0,007 substitution/site/Ma, (CI : 0,003–0,012). Autrement dit, si certains petits rongeurs, telles que les gerbilles, remplacent toutes les troisièmes positions de leurs cytb en 1 à 2 Ma en moyenne, ce processus prend ~ 100 Ma chez certains cétacés.

La taxonomie explique une grande partie de la variance du taux de mutation entre espèces. Par exemple, 21% de la variance est expliqué par le niveau ordinal. Les rongeurs apparaissent clairement comme les espèces qui, en moyenne, évoluent le plus vite (figure 6.1) avec une moyenne de 0,176 substitution/site/Ma, ce qui est 20 fois plus rapide que les cétacés (0,010 substitution/site/Ma en moyenne). Avec 43% de la variance expliquée, le niveau famille est celui qui explique le plus de variance (ANOVA hiérarchisée sur les niveaux superordre / ordre / famille / genre). Similairement, nous avons détecté une forte inertie phylogénétique, notamment à travers une forte autocorrélation phylogénétique des taux d’évolution estimée à travers l’indice de Moran ($p < 0,001$, Nabholz *et al.*, 2008c).

6.1.2 Taux de mutation du cytochrome *b* chez les oiseaux : moins rapide et moins variable

La moyenne du taux de substitution de la troisième position des codons du cytb des oiseaux est égale à **0,025** substitution/site/Ma, et la médiane est égale à **0,017** substitution/site/Ma, ce qui est effectivement proche de la calibration populaire des 2% de divergence pour l’ensemble des positions de la séquence. Cependant, bien que moins importante que celle rencontrée pour les mammifères, la distribution montre une variance substantielle. Utilisant le même critère que pour les mammifères, la distribution montre une gamme proche d’un facteur 30 entre les lignées d’oiseaux :

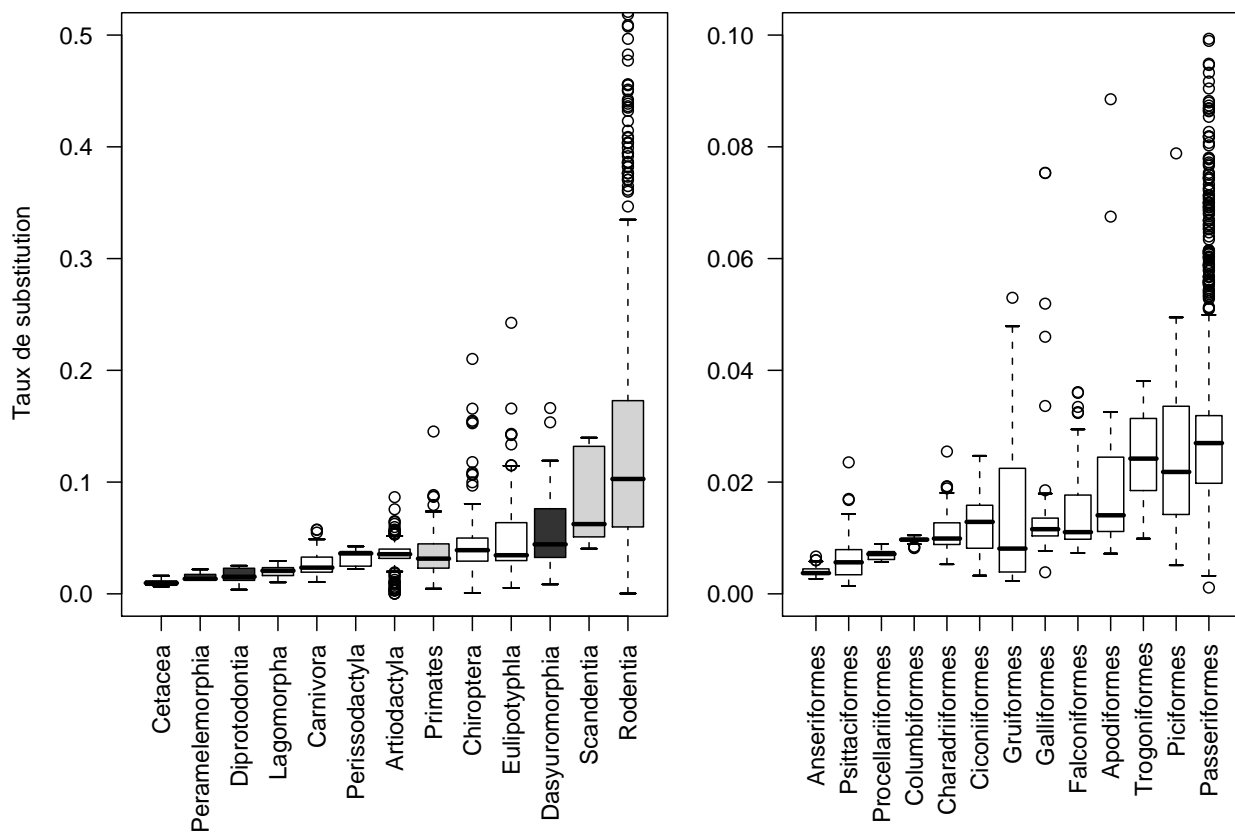


FIG. 6.1 – Répartition des taux de substitution des troisièmes positions des codons du *cytb* en fonction des ordres des mammifères (à gauche) et des oiseaux (à droite). Seuls les ordres avec plus de 5 espèces ont été représentés chez les mammifères et 20 chez les oiseaux. Gris clair : Euarchontoglires ; Blanc : Laurasiatheria ; et Gris foncé : marsupiaux. La barre horizontale correspond à la médiane ; les limites de la boîte donnent les quartiles et les moustaches s'étendent à 1,5 fois la dimension des interquartiles, les autres points étant hors de cet intervalle. Notez la différence d'échelle entre les mammifères et les oiseaux. Le taux de substitution est exprimé en substitution/site/Ma.

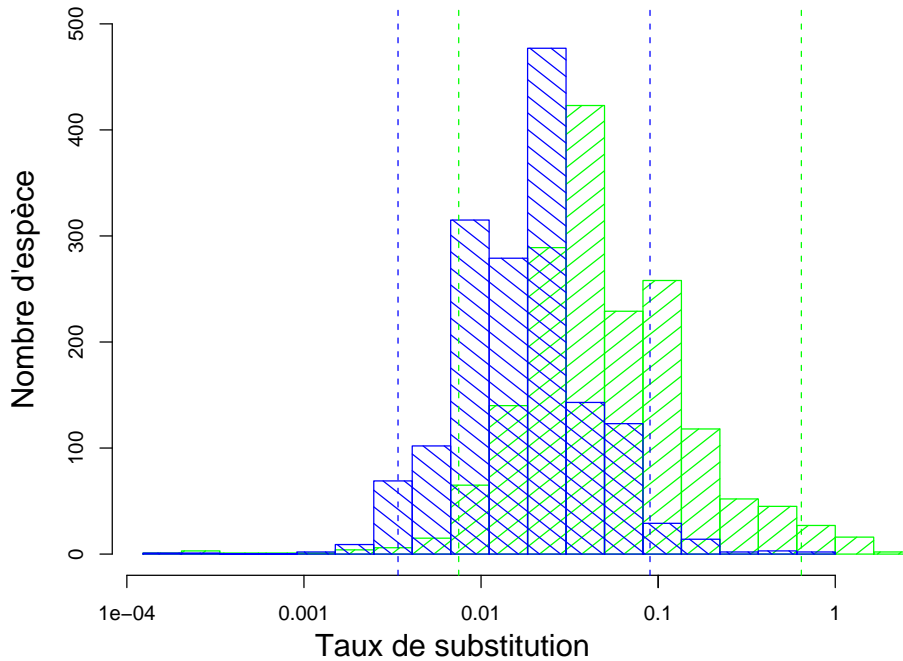


FIG. 6.2 – Distribution des taux de substitution des troisièmes positions des codons du cytochrome *b* de 1696 espèces de mammifères (en vert) et 1571 espèces d'oiseaux en (bleu). Les traits verticaux représentent les limites des 2,5% et 97,5% des distributions. Le taux de substitution est log transformé et exprimé en substitution/site/Ma.

les valeurs correspondant aux 2,5% et 97,5% de la distribution se situent respectivement à 0,003 et 0,090 substitution/site/Ma (figure 6.2). De la même façon que chez les mammifères, la taxonomie explique une grande partie de la variance des taux de mutation entre espèces (55% pour le niveau ordinal et 27% pour le niveau famille). Les Passereaux sont l'ordre (contenant plus de 20 espèces) qui évolue le plus rapidement avec 0,038 substitution/site/Ma et les Anseriformes sont les plus lents (0,004 substitution/site/Ma, voir figure 6.1).

Un résultat intéressant provient de la comparaison entre les taux de mutation des mammifères et des oiseaux. Ces derniers évoluent, globalement, 3,9 fois moins vite, et montrent clairement une gamme de variation plus faible (coefficient de variation : 0,07 pour les oiseaux et 0,32 pour les mammifères, figure 6.2).

		Sans contrôle phylogénétique				Avec contrôle phylogénétique			
		pente	F	R ²	<i>p</i>	pente	F	R ²	<i>p</i>
Mammifères	Masse corporelle	-0,13	134,1	0,25	<0,001	-0,14	81,85	0,17	<0,001
	Maturité	-0,44	144	0,26	<0,001	-0,34	59,82	0,13	<0,001
	Longévité	-0,61	142,7	0,26	<0,001	-0,53	88,42	0,13	<0,001
Oiseaux	Masse corporelle	-0,25	122,5	0,39	<0,001	-0,22	118,2	0,38	<0,001
	Longévité	-0,66	65,3	0,25	<0,001	-0,46	44,2	0,19	<0,001

TAB. 6.2 – Relation entre le taux de substitution de la 3ème position des codons du *cytb* et trois traits d’histoire de vie chez 410 espèces de mammifères et 196 espèces d’oiseaux (la relation avec la maturité sexuelle n’a pas pu être déterminée chez les oiseaux car ce trait d’histoire de vie n’est disponible que pour trop peu d’espèces). Le contrôle phylogénétique est réalisé avec les contrastes phylogénétiques (Felsenstein (1985) implémenté dans le package APE du logiciel R Paradis (2007))

6.2 Déterminisme du taux de mutation mitochondrial - confirmation de l’hypothèse de longévité ?

6.2.1 Résultats mitochondriaux

Comme nous l’avons évoqué au chapitre 4.1, l’obtention des taux de substitution estimés de manière précise et pour le maximum d’espèces est le point de départ nécessaire à l’étude des déterminants du taux d’évolution. Nous nous sommes concentrés sur les trois hypothèses principales - temps de génération, métabolisme et longévité - en utilisant les traits d’histoire de vie correspondant à ces trois hypothèses, soit maturité sexuelle femelle, masse corporelle et longévité maximale. Ces trois traits d’histoire de vie ont été collectés à partir de la base de données **AnAge** (de Magalhaes *et al.*, 2005).

Une analyse intra-classe chez 196 espèces d’oiseaux et 410 espèces de mammifères montre que le taux d’évolution du *cytb* est fortement et négativement corrélé à tous les traits d’histoire de vie testés (tableau 6.2). Ces relations sont en accord avec les trois hypothèses explicatives, et robustes à un contrôle phylogénétique de type contraste indépendant (Felsenstein, 1985). Chez les mammifères, c’est la maturité sexuelle qui explique le plus de variance dans les taux d’évolution ($R^2 = 0,26$). En ce qui concerne les oiseaux c’est la masse qui explique le plus de variance ($R^2 = 0,45$). Compte tenu de la forte co-variance entre les différents traits d’histoire de vie, il est difficile d’interpréter les analyses en utilisant une seule variable explicative. Cependant, il est déjà possible de vérifier la corrélation prédite par Gillooly *et al.* (2005), où la masse corporelle est négativement corrélée au taux de substitution avec une pente de $-1/4$ dans une relation log-log. Dans les deux groupes, cette pente n’est pas retrouvée, bien qu’elle soit assez proche de la valeur prédite chez les oiseaux (tableau 6.2).

Afin de tenter de discriminer entre les hypothèses explicatives, nous avons effectué des régressions multiples combinant plusieurs variables. Une sélection des modèles à partir du critère AIC (“Akaike’s Information Criteria”) correspondant à la différence des vraisemblances entre deux modèles, pondérée par la différence entre le nombre de paramètres, montre que chez les mammifères, le modèle avec le meilleur AIC est un modèle combinant les 3 variables explicatives : (maturité sexuelle + longévité + masse corporelle, $R^2 = 0,30$, $p < 0,001$). Ce résultat montre que, chez les mammifères, malgré la forte corrélation entre les traits d’histoire de vie, les effets de chaque variable apparaissent raisonnablement cumulatifs. Ce n’est pas le cas chez les oiseaux où le modèle avec le meilleur AIC utilise seulement la masse corporelle.

Toujours dans l’optique de discriminer entre les hypothèses explicatives, nous pouvons formuler une prédiction spécifique à l’hypothèse de la longévité. En effet, cette hypothèse est la seule qui nécessite un *processus d’adaptation* qui affecte les espèces *en fonction* de leur longévité. Selon cette hypothèse, la pression de sélection devrait être forte chez les espèces particulièrement longévives et, réciproquement, un relâchement de la pression de sélection devrait affecter les espèces à durée de vie courte. Selon cette hypothèse, nous devrions observer une relation entre la longévité et le taux d’évolution plus forte chez les espèces longévives que chez les espèces à courte durée de vie. Au contraire, aucune différence d’effet n’est attendue dans le cas des deux autres hypothèses, qui peuvent être qualifiées de “mécanistes”. Afin de tester cette prédiction, nous avons découpé le jeu de données en deux sous-groupes sur la base de la longévité médiane (19,8 ans pour les mammifères et 13,4 ans pour les oiseaux). Chez les mammifères, les résultats sont en accord avec la prédiction ; l’effet de la longévité est plus fort chez les espèces longévives ($R^2 = 0,198$) que chez les espèces à durée de vie courte ($R^2 = 0,098$). Cet effet n’est pas retrouvé pour la masse corporelle ou pour la maturité sexuelle femelle, pour lesquelles l’effet est similaire de part et d’autre du seuil de longévité arbitraire (tableau 6.3). Visuellement, ce résultat se caractérise par un changement dans la pente de la relation log-log entre le taux d’évolution et la longévité (figure 6.3). Cette rupture de pente n’est pas détectée pour les deux autres traits d’histoire de vie. De manière congruente, les résultats obtenus avec les régressions multiples montrent que le déterminant principal du taux de substitution mitochondrial des mammifères est la maturité sexuelle femelle pour les espèces à durée de vie courte, et la longévité pour les espèces longévives (tableau 6.3).

L’influence majeure de la longévité n’est pas retrouvée chez les oiseaux. L’utilisation du même protocole que pour les mammifères montre que la masse corporelle reste le déterminant majeur des variations de taux d’évolution, à la fois chez les oiseaux longévifs et chez les oiseaux à durée de vie courte. L’examen des distributions des longévités maximales à l’intérieur des différents ordres d’oiseaux montre qu’il existe une différence claire entre les passereaux et les non-passereaux (voir figure 6.4). Suivant cette apparente bimodalité dans la distribution des longévités, nous avons effectué une analyse basée sur cette dichotomie plutôt que sur la médiane des longévités. Les résultats sont plutôt surprenants et quelque peu inexplicables en ce qui concerne les passereaux (mais voir chapitre 6.3). Les passereaux, ordre monophylétique d’espèces globalement petites, montrent une relation positive entre le taux de substitution et les traits d’histoire de vie. Pour les non-passereaux, groupe

Variable 1	Variable 2	Durée de vie courte (n=205)				Durée de vie longue				
		pente	R ²	p1	p2	pente	R ²	p1	p2	
Mammifères	Maturité	-0,40	0,11	<0,001		-0,43	0,17	<0,001		
	Masse corporelle	-0,15	0,15	<0,001		-0,08	0,10	<0,001		
	Longévité	-0,53	0,10	<0,001		-0,81	0,20	<0,001		
	Maturité		0,13	0,010	0,049		0,21	0,050	<0,001	
	Maturité	Masse corporelle		0,21	0,022	<0,001		<0,001	0,054	
	Masse corporelle	Longévité		0,18	0,250	<0,001		0,135	<0,001	
Oiseaux		Durée de vie courte (n=98)				Durée de vie longue				
	Masse corporelle	-0,20	0,23	<0,001		-0,28	0,23	<0,001		
	Longévité	0,04	0,001	0,82		-0,97	0,13	<0,001		
	Masse corporelle		0,24	<0,001	0,30		0,25	<0,001	0,11	
			Passereaux (n=88)				Non passereaux (n=108)			
	Masse corporelle		0,22	0,11	<0,002		-0,17	0,12	0,002	
Longévité		0,47	0,15	<0,001		-0,41	0,09	0,001		
Masse corporelle	Longévité		0,17	0,14	0,02		0,13	0,013	0,09	

TAB. 6.3 – Relation entre le taux de substitution de la 3ème position des codons du cytb et trois traits d’histoire de vie chez les espèces d’oiseaux et mammifères longévives (respectivement plus de 13,4 ans et 19,8 ans) et à durée de vie courte, ainsi que chez les passereaux et non-passereaux pour les oiseaux (voir texte).

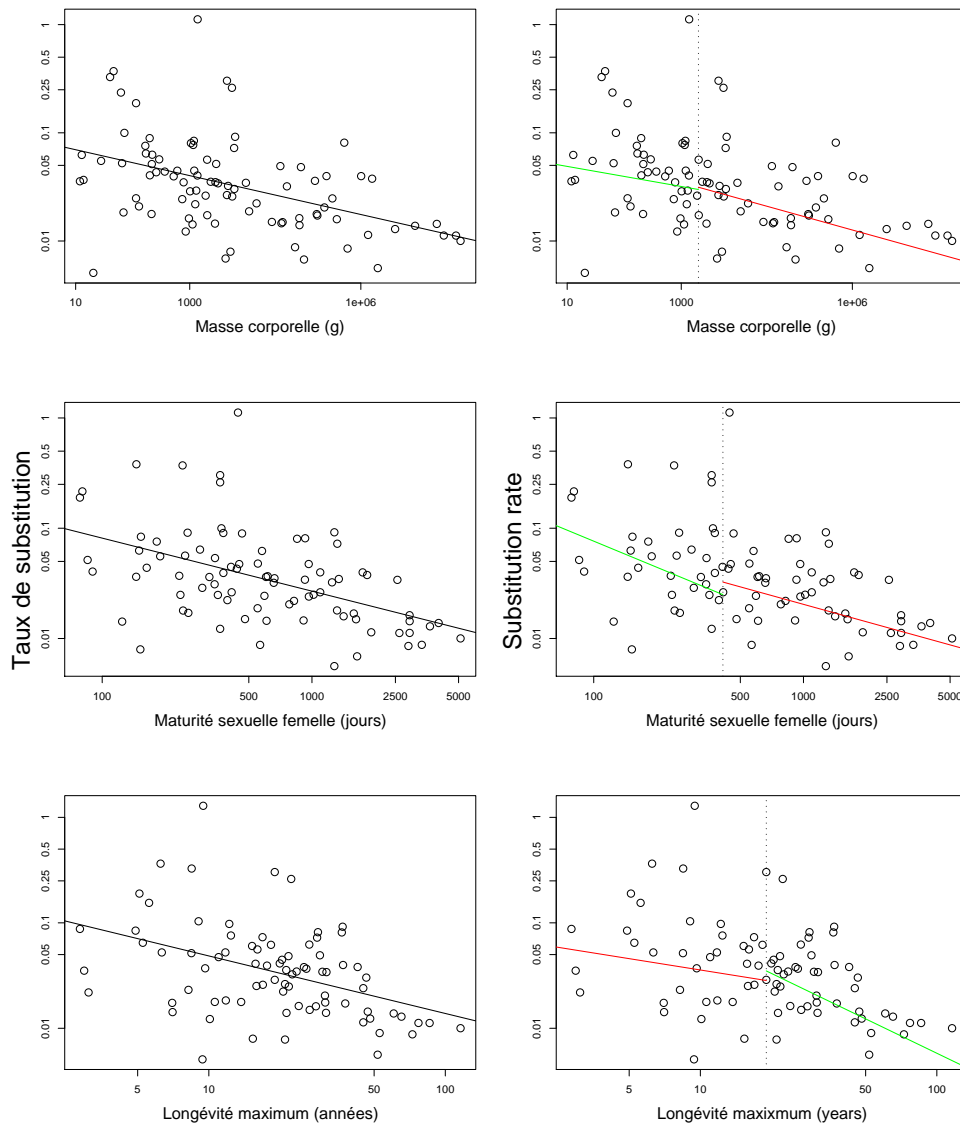


FIG. 6.3 – Corrélation entre les taux de substitution des troisièmes positions des codons du *cytb* et les traits d’histoire de vie des mammifères. Chaque point correspond à une famille. A gauche : les régressions simples. A droite : régressions chez les familles à durée de vie courte ($< 19,8$ ans, en vert) et régressions chez les familles à durée de vie longue ($> 19,8$ ans, en rouge). Les taux de substitutions sont exprimés en substitution par site/Ma.

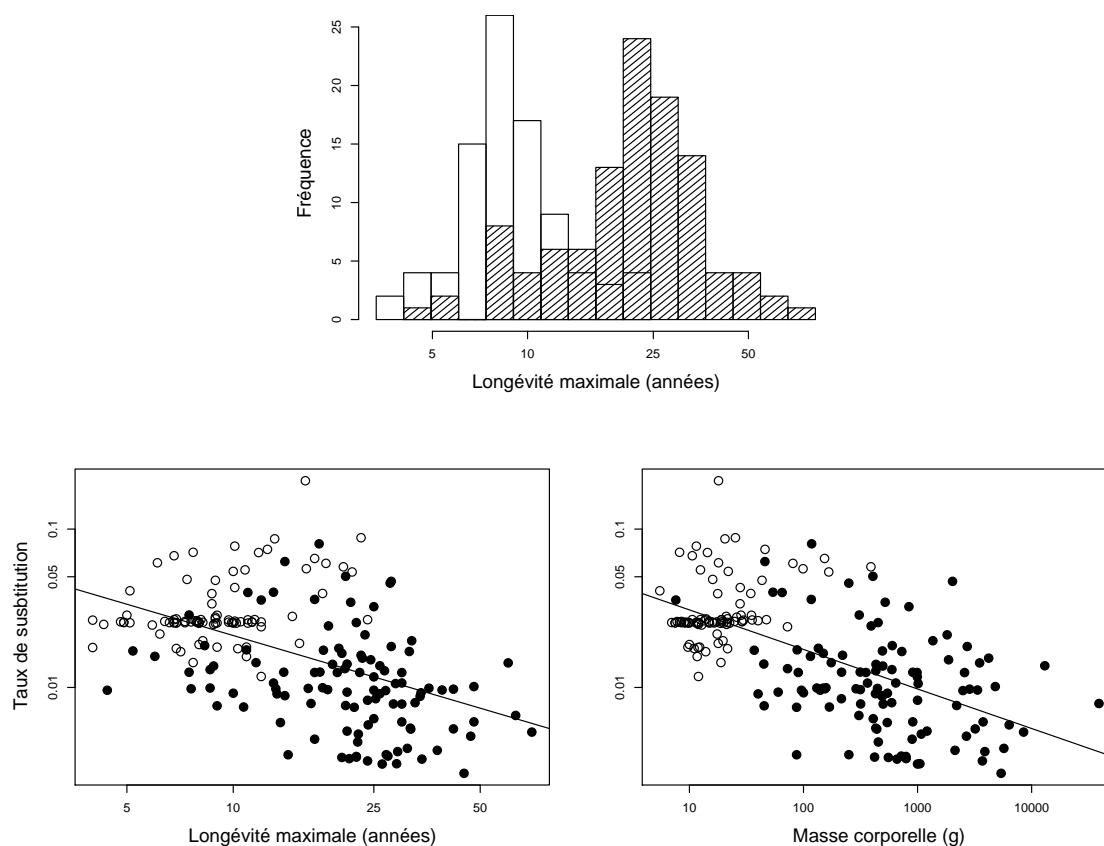


FIG. 6.4 – Figure du haut : Distribution des longévités maximales entre les passereaux (histogramme non hachuré) et les non-passereaux (histogramme hachuré). Figures du bas : Corrélation entre le taux de substitution et la masse corporelle (à gauche) ou la longévité (à droite). Les cercles blancs correspondent aux passereaux ($n = 88$) et les cercles noirs aux non-passereaux ($n = 108$). Les taux de substitution des troisièmes positions des codons du *cytb* sont exprimés en substitution par site/Ma.

constitué essentiellement d'espèces longévives, la masse a toujours un effet plus important que la longévité bien que, quantitativement, la différence entre les deux variables s'estompe quelque peu (tableau 6.3).

Enfin, il est intéressant d'utiliser la comparaison entre les résultats oiseaux et mammifères pour essayer de discriminer entre les hypothèses explicatives. En effet, comme on l'a vu au chapitre 5.1, les oiseaux présentent une longévité exceptionnelle comparés aux mammifères de même masse corporelle. Cette longévité est couplée à un métabolisme masse-spécifique important, caractéristique qu'ils partagent avec les chauve-souris. Afin de quantifier les différences entre les oiseaux et les mammifères, nous avons traité les deux jeux de données simultanément. Pour ce faire, nous avons réalisé une analyse combinant à la fois l'effet de la classe (mammifères *vs* oiseaux) et l'effet des traits d'histoire de vie (figure 6.5). Cette analyse révèle que l'appartenance à l'une ou l'autre des deux classes a une influence forte sur le taux de mutation, même lorsque l'effet des traits d'histoire de vie est pris en compte,

(Masse corporelle + Classe)			
Variabes	effet	valeur t	p
Masse corporelle	-0,25	-8,91	<0,001
Classe (Mammifère)	0,22	3,09	<0,001
Masse corporelle : Classe (Mammifère)	0,13	4,43	<0,001
(Longévité maximale + Classe)			
Variabes	effet	valeur t	p
Longévité maximale	-0,66	-7,16	<0,001
Classe (Mammifère)	0,31	2,50	<0,001
Longévité maximale : Classe (Mammifère)	0,05	0,46	0,65

Tab. 6.4 – Effets des traits d’histoires de vie et du statut oiseaux ($n = 196$) vs mammifères ($n = 500$) (Classe) sur le taux de substitution de l’ADN mitochondrial.

un oiseau évolue en moyenne moins vite qu’un mammifère comparable en termes de traits d’histoire de vie. Il existe une interaction significative entre la classe et la masse corporelle ($p < 0,001$), mais pas entre la longévité et la classe ($p = 0,65$).

Par mesure de comparaison, le taux de substitution au niveau amino-acides des mammifères a, également, été estimé. Les taux de substitution non-synonymes et en troisième position des codons sont bien corrélés ($R^2 = 0,44$, $p < 0,001$). La médiane du taux de substitution protéique est 0,002 substitution/site/Ma. Les rongeurs évoluent en moyenne à 0,0044 substitution/site/Ma ce qui est ~ 10 fois plus rapide que les lagomorphes (0,0004) et ~ 5 fois plus rapide que les cétacés 0,009 substitution/site/Ma. Ce résultat montre que le taux de substitution non-synonyme est moins variable entre espèces que le taux de substitution synonyme.

6.2.2 Résultats nucléaires

Comme nous l’avons évoqué précédemment (chapitre 4.1), l’hypothèse du temps de génération implique un effet équivalent sur le génome mitochondrial et le génome nucléaire.

Afin d’obtenir une image du taux de mutation nucléaire, nous avons effectué une analyse similaire à celle réalisée sur le *cytb* sur quelques gènes nucléaires. Les gènes ont été sélectionnés sur la base de leur échantillonnage ; ils correspondent à des marqueurs utilisés en phylogénie moléculaire : interphotoreceptor retinoid-binding protein exon 1, (IRBP, 664 espèces), α -adrenergic receptor 2B (151 espèces), breast cancer 1 (BRCA 1, 168 espèces), et enfin le von Willebrand factor (VWF, 294 espèces). Tous les résultats obtenus sont similaires entre gènes mais difficilement comparables du fait des différences dans le nombre d’espèces représentées. Nous avons choisi de simplement présenter les résultats obtenus avec l’IRBP. Le *cytb* des mammifères évolue près de 60 fois plus vite que l’IRBP (moyenne du taux de substitution des troisièmes positions des codons : 0,0008 substitution/site/Ma). De plus, l’IRBP ne présente pas les variations extrêmes de taux de substitution entre espèces détectées chez l’ADN mitochondrial. Par exemple, sans exclure aucun point de la distribution des taux de mutation de l’IRBP, l’espèce la plus rapide évolue seulement 9 fois plus vite que la plus lente. Les taux de mutation obtenus pour le *cytb* et l’IRBP sont bien corrélés ($R^2 = 0,27$, $p < 0,001$, $n = 364$). La différence

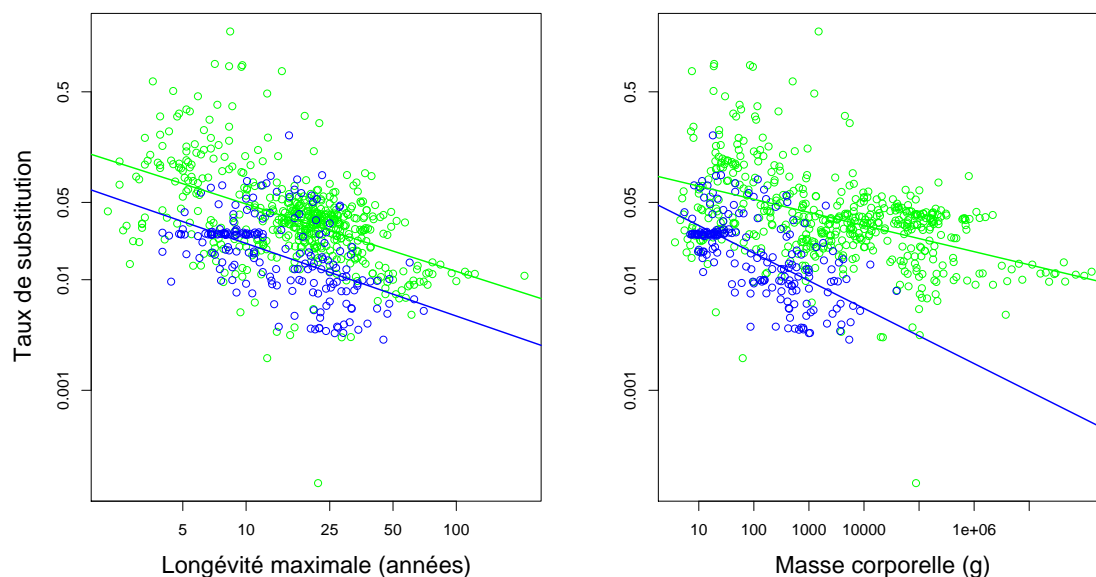


FIG. 6.5 – Relation entre les taux de substitution des troisièmes positions des codons du *cytb* et la longévité (à gauche) ou la masse (à droite) chez 196 espèces d’oiseaux (en bleu) et 500 espèces de mammifères (en vert). Les taux de substitution sont exprimés en substitution par site/Ma.

entre le taux de mutation du *cytb* et de l’IRBP s’accroît chez les espèces rapides, où le taux de mutation mitochondrial peut être 100 fois plus élevé que le nucléaire (*e.g.*, 136 pour Muridae, Rodentia). Ce n’est pas le cas des lentes, où le *cytb* est moins de 20 fois plus rapide que l’IRBP (*e.g.*, 15 pour les Hominidae, Primates). Enfin, nous n’avons pas retrouvé un effet plus fort de la longévité chez les espèces de mammifères longévives lorsque le taux de mutation de l’IRBP est considéré. Ce résultat ne va pas dans le sens de l’hypothèse du temps de génération mais, comme nous le verrons dans la discussion (6.4.2), cela ne disqualifie pas forcément cette hypothèse.

6.3 Robustesse et généralisation des résultats *cytochrome b*

Le matériel et la méthode choisis pour réaliser les analyses sur le taux de mutation mitochondrial nous semblent être les plus appropriés pour révéler la véritable étendue de variation des taux. En effet, les bases de données de séquences nous laissent le choix entre 2 types de jeux de données : un jeu de données de génomes mitochondriaux complets contenant peu d’espèces (~ 150 mammifères et ~ 100 oiseaux), ou un jeu de données à l’échelle d’un gène avec énormément d’espèces (de l’ordre du millier), mais évidemment moins d’information génétique disponible pour chaque espèce. En examinant le jeu de données de génomes mitochondriaux, on peut

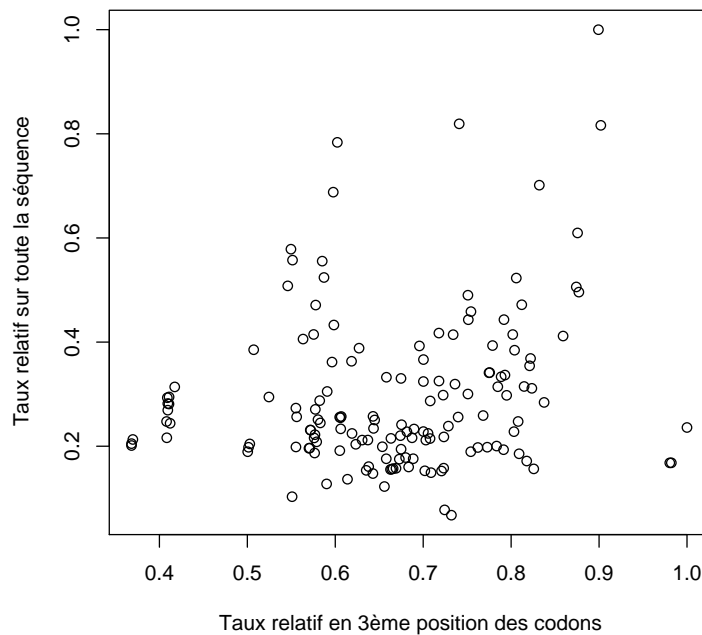


FIG. 6.6 – Corrélation entre les taux relatifs obtenus avec MULTIDIVTIME à partir de toutes les positions du cytochrome *b* ou en utilisant uniquement les troisièmes positions des codons. Les 150 espèces de mammifères utilisées pour réaliser cette analyse correspondent aux espèces pour lesquelles le génome mitochondrial complet est disponible. Les autres gènes présentent le même type de relation.

constater qu'il est déséquilibré du point de vue taxonomique. Certains groupes, tels que les cétacés et les pinnipèdes, sont bien échantillonnés, et pourraient présenter des divergences synonymes peu ou pas saturées. D'autres groupes sont, par contre, clairement sous échantillonnés. C'est particulièrement le cas des rongeurs, un groupe à évolution rapide, pour lequel moins de 20 espèces sont disponibles. Ainsi, dans le jeu de données de génomes mitochondriaux, certains couples d'espèces soeurs présentent des divergences déjà trop importantes pour que l'on puisse espérer estimer correctement la quantité de divergence synonyme. Une illustration de ce phénomène provient de la mauvaise corrélation entre les taux d'évolution estimés à partir des troisièmes positions des codons et ceux estimés avec toute la séquence (figure 6.6).

Enfin, une estimation directe du taux de mutation, en utilisant les troisièmes positions des codons du jeu de données de génome complet, sous-estime la variation des taux de mutation avec d'un facteur 10 entre les espèces les plus rapides et les espèces les plus lentes.

Plusieurs remarques et limites peuvent, néanmoins, être formulées quant à l'utilisation d'un jeu de données uniquement cytb. Dans ce chapitre, nous allons tenter de répondre à trois remarques principales : (i) Les résultats cytb sont-ils généralisables au génome mitochondrial ? (ii) Les datations réalisées uniquement avec le cytb sont-elles fiables ? (iii) Toutes les analyses ont été réalisées avec MULTI-

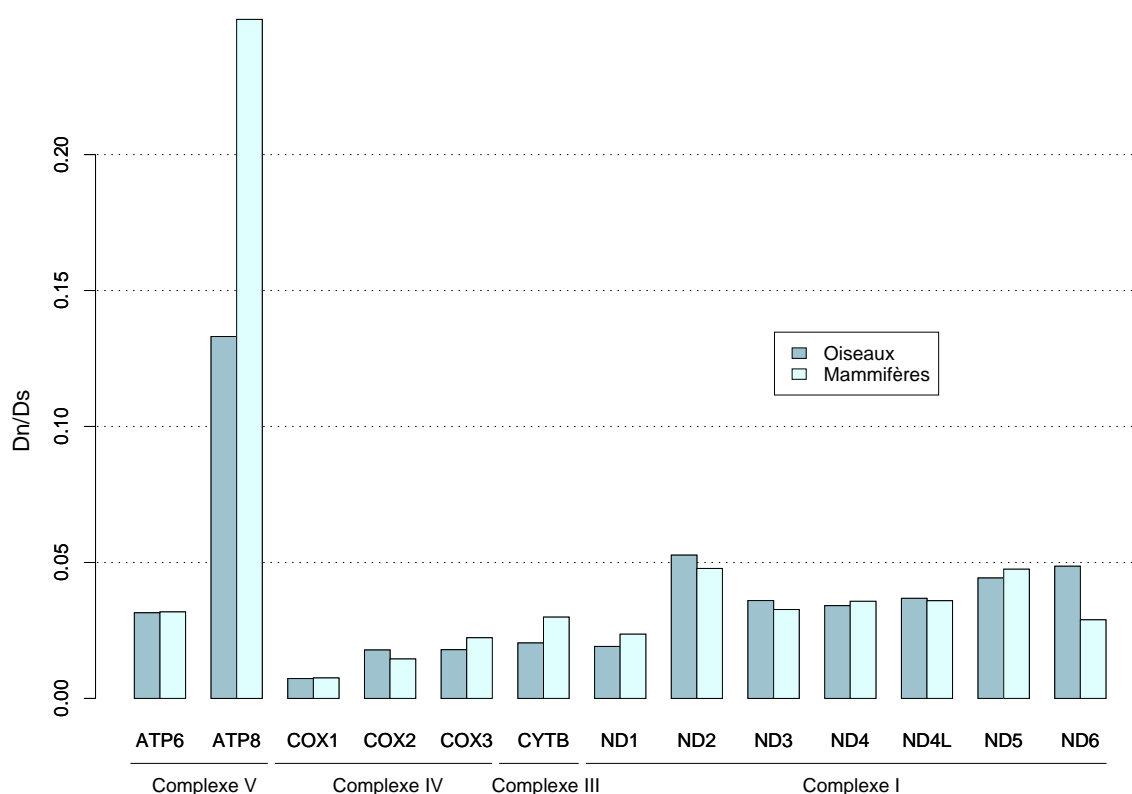


FIG. 6.7 – Ratio de la divergence non-synonyme D_n sur la divergence synonyme D_s estimé à l'échelle des oiseaux et des mammifères pour chacune des treize protéines mitochondriales.

DIVTIME ; dans quelle mesure les résultats sont-ils robustes à l'utilisation d'un autre programme, avec, par exemple, un autre modèle d'évolution des taux d'évolution ?

La première remarque concerne l'utilisation d'un seul gène protéique mitochondrial : le *cytb*. On peut, naturellement, se demander dans quelle mesure les résultats obtenus à partir du *cytb* sont généralisables à l'ensemble du génome mitochondrial. Les gènes protéiques mitochondriaux ont tendance à avoir des taux d'évolution plus corrélés entre gènes appartenant à un même complexe protéique de la chaîne respiratoire (Waddell *et al.*, 2007). Similairement, le niveau de contrainte de chaque gène est également assez bien lié à son appartenance à un complexe protéique particulier (Pesole *et al.*, 1999, voir figure 6.7). Qu'en est-il du taux de mutation entre lignées ? Afin de tester le degré de corrélation entre les taux de mutation du *cytb* avec les autres gènes protéiques, nous avons pu inférer le taux de mutation du génome mitochondrial complet pour deux groupes taxonomiques : les caniformes (un sous-ordre de carnivores) et les cétacés. Dans ces deux groupes, le génome mitochondrial est disponible pour beaucoup d'espèces permettant ainsi de limiter la saturation. Les deux groupes présentent une bonne corrélation entre les taux de mutation *cytb* et le taux obtenu en utilisant la troisième position des codons des 13 gènes protéiques mitochondriaux ($R^2 = 24\%$, $p < 0,01$ pour les 39 espèces de caniformes et $R^2 = 83\%$, $p < 0,001$ pour 26 espèces de cétacés). Très peu de génomes complets sont disponibles pour les espèces à évolution rapide. Par exemple, nous ne disposons à l'heure actuelle que d'une quinzaine de rongeurs appartenant à des familles différentes, donc

Divergence	ADNmt complet	Murphy	Cytb
hominidés/cercopithecidés	34,6 ($\pm 1,6$)	25,5 ($\pm 2,7$)	27,9 ($\pm 3,0$)
homme/gibbon	21,7 ($\pm 1,0$)	15,6 ($\pm 2,1$)	16,1 ($\pm 3,1$)
<i>Ochotona/Lagopus</i>	–	50,5 ($\pm 3,2$)	45,5 ($\pm 5,2$)
Baleine à dents / Baleine à fanons	25,5 ($\pm 2,0$)	27,4 ($\pm 2,8$)	34,4 ($\pm 12,3$)
Rhinocéros blanc/Rhinoceros Indien	26,1 ($\pm 2,3$)	–	24,4 ($\pm 6,9$)
Feliformia/Caniformia	49,0 ($\pm 2,7$)	56,8 ($\pm 3,2$)	53,0 ($\pm 5,6$)

TAB. 6.5 – Comparaison entre les dates de divergence estimées en utilisant uniquement le cytb et les dates estimées par deux analyses multilocus utilisant le génome mitochondrial complet ou l’alignement de Murphy *et al.* (2001) principalement composé de gènes nucléaires. Datations réalisées par Hasegawa *et al.* (2003).

déjà trop divergentes pour espérer estimer correctement le taux de mutation. Un effort de séquençage chez d’autres groupes à évolution rapide serait bénéfique à la compréhension du processus mutationnel mitochondrial, notamment à travers une comparaison rapide *vs* lent.

La deuxième remarque concernant la robustesse des résultats est en lien avec l’étape de datation moléculaire. Cette étape détermine toute la qualité de l’estimation des taux d’évolution neutres (des troisièmes positions des codons) puisque cette datation permet la conversion entre taux d’évolution relatifs, à l’intérieur de chaque groupe taxonomique, et taux d’évolution absolus, comparables entre espèces. Cette étape a été réalisée en utilisant seulement le cytb, une séquence relativement courte pour réaliser une datation à l’échelle des mammifères et des oiseaux, couvrant des dizaines de millions d’années d’évolution. Conscients de la faible quantité des données moléculaires, nous avons maximisé le nombre de points de calibration fossiles (voir tableau 5.2). Cependant, il paraît important d’effectuer des vérifications, d’une part à travers une revue bibliographique, et d’autre part, à l’aide d’analyses complémentaires.

La vérification bibliographique consiste en une recherche et une comparaison de dates de divergence obtenues par d’autres analyses, si possible multi-locus, précédemment publiées. En ce qui concerne les mammifères, nous avons réalisé cette vérification sur les groupes particulièrement rapides et lents présentés dans le tableau 6.1. Les dates de divergence évoquées dans ce tableau sont proches des dates estimées avec le cytb. Par exemple, la divergence entre les deux espèces du genre *Balaenoptera* a été estimée à 6,9 Ma ($\pm 2,2$) par Sasaki *et al.* (2005) dans une analyse utilisant le génome mitochondrial complet, la date cytb étant 9,9 Ma ($\pm 2,5$). De même, pour les deux espèces de gerbilles (*Gerbillus nanus* et *G. nigeriae*), appartenant à un groupe particulièrement rapide, la date de divergence obtenue par Chevret & Dobigny (2005) avec la combinaison de deux gènes mitochondriaux dont le cytb est 4,1 Ma ($\pm 0,9$) ce qui est accord avec notre estimation (6,0 Ma $\pm 0,9$). Utilisant la datation à l’échelle des mammifères réalisée par Hasegawa *et al.* (2003), nous avons généralisé ces comparaisons (tableau 6.5). Les estimations des dates de divergence obtenues avec le cytb sont généralement bien en accord avec celles obtenues par l’approche multi-locus d’Hasegawa *et al.* (2003).

Concernant les oiseaux, la vérification bibliographique est rendue difficile par une littérature clairsemée, probablement à cause du faible nombre de points

de calibrations disponibles. Päckert *et al.* (2007) utilisent des calibrations de type biogéographiques pour estimer les dates de divergence à l'intérieur du groupe des mésanges (famille des paridés). Ces auteurs obtiennent une date de divergence entre *Parus* et *Poecile* d'environ 14 Ma, ce qui est plus récent que notre estimation (25 Ma \pm 8). Pasquet *et al.* (2007) ont réalisé une datation moléculaire du groupe des drongos (famille des corvidés). Leur estimation de la divergence de deux passereaux (*Lanius* et *Corvus*, légèrement supérieure à 30 Ma) est un peu plus récente que notre estimation (44 Ma \pm 8). Il est difficile d'affirmer, uniquement sur la base de ces deux comparaisons, que notre estimation des taux de mutation des oiseaux soit sous estimée, d'autant plus que ces deux études ont également été réalisées en utilisant le cytb comme marqueur moléculaire principal. Il existe deux datations moléculaires à large échelle taxonomique des oiseaux. Pereira & Baker (2006) ont effectué une datation moléculaire des oiseaux et des mammifères en utilisant des génomes mitochondriaux complets. Malheureusement, les dates de divergence obtenues dans cette analyse sont visiblement trop anciennes, par exemple, la divergence *Mus/Rattus* est estimée à plus de 50 Ma, soit 4 fois plus que la calibration fossile (Benton & Donoghue, 2007). Cette sur-estimation des dates provient de l'absence de bornes maximales contraignant l'âge maximal de certaines divergences, ce qui entraîne un vieillissement généralisé des dates de divergence. La deuxième datation à large échelle est celle de Barker *et al.* (2004). Les auteurs de cette étude ont réalisé une phylogénie à l'échelle des passereaux en utilisant deux gènes nucléaires. Cette analyse phylogénétique est assortie d'une datation moléculaire réalisée avec deux méthodes différentes des méthodes Bayésiennes présentées plus haut (Sander-son, 1997, 2002). L'analyse de Barker *et al.* (2004) fournit des dates sensiblement identiques à notre analyse cytb avec, par exemple, 45 Ma (\pm 7) pour l'origine du clade des Passerida contre 50 Ma (\pm 7) pour notre étude ou 65 Ma (\pm 5) pour la divergence basale aux sein des suboscines du nouveau monde (correspondant à la divergence *Tyrannoidea/Furnarioidea*) contre 56 Ma (\pm 9) pour l'estimation cytb.

Contrairement au cas des mammifères, il est difficile de se faire une idée de la qualité des dates de divergence estimées pour les oiseaux à partir de la littérature. Malgré tout, on peut penser que les dates de divergence des oiseaux sont moins fiables à cause du manque de (bons) points de calibrations fossiles. Un exemple particulièrement clair concerne le groupe rapide des passereaux. Ce clade contient la moitié de la diversité spécifique des oiseaux et représente, au minimum, 55 Ma d'évolution (Boles, 1995). Pourtant, un seul point de calibration est disponible pour ce clade, correspondant à la divergence basale du groupe. L'identification d'un point de calibration supplémentaire à l'intérieur des passereaux, qui représenterait l'équivalent de la divergence *Mus/Rattus* par exemple, manque cruellement dans la littérature ornithologique et modifierait probablement notre estimation.

Toujours concernant la robustesse de l'estimation des dates de divergence, nous avons effectué une analyse complémentaire. Notre datation moléculaire cytb est effectuée en sélectionnant deux espèces par groupe taxonomique (voir figure 6.5). Nous avons reproduit l'étape de datation moléculaire en sélectionnant des paires d'espèces différentes. La comparaison de 10 réplifications effectuées chez les mammifères fournit des résultats très similaires (le R^2 des corrélations entre les taux absolus estimés entre les analyses varie entre 84% et 98%), qui prouvent que la qualité de la datation

ne dépend pas des espèces sélectionnées.

Un autre sujet problématique de notre analyse concerne l'utilisation d'un seul programme pour l'ensemble de l'analyse : MULTIDIVTIME. Le découplage entre les taux d'évolution et les dates de divergence est une opération difficile car ces deux variables sont totalement imbriquées dans l'estimation des longueurs de branches. De plus, l'implémentation de programme utilisant le processus MCMC est réputée techniquement difficile. Enfin, MULTIDIVTIME utilise un modèle de changement des taux d'évolution faisant l'hypothèse d'autocorrélation phylogénétique et un modèle d'évolution des séquences de type F84 pour l'ADN et mtREV pour les protéines. Quelle est la performance du programme MULTIDIVTIME par rapport à d'autres programmes implémentés plus récemment ? Nous avons répliqué nos analyses en utilisant le programme MCMCTREE (Yang & Rannala, 2006), à la place de MULTIDIVTIME. Les résultats concernant à la fois les datations moléculaires (utilisant le cytb au niveau amino-acides) et l'estimation des taux en troisième position des codons donnent des résultats très similaires (figure 6.8). Sur les sept groupes de mammifères testés, un seul présente une corrélation non significative (les périssodactyles) entre les 2 méthodes.

Les modèles d'évolution des taux et des séquences sont-ils importants sur les dates de divergence estimées ? Nous avons utilisé un autre modèle implémenté dans MCMCTREE. Ce modèle ne fait pas l'hypothèse de l'autocorrélation des taux d'évolution. Les résultats obtenus avec ce modèle semblent moins stables, fournissant parfois des valeurs de taux extrêmement élevées. Cette observation est confirmée par une publication récente utilisant un autre programme (BEAST Emerson, 2007). Enfin, le modèle d'évolution des séquences : MULTIDIVTIME utilise le modèle d'évolution F84, l'utilisation du modèle HKY dans MCMCTREE ne modifie pas les résultats.

En conclusion, mises à part les datations moléculaires du groupe des oiseaux, où le problème du manque de calibrations fossiles n'est pas spécifique à notre étude, nos résultats semblent robustes en ce qui concerne l'estimation des dates de divergence et les taux de mutation espèces spécifiques.

6.4 Discussion et implication des résultats sur les taux d'évolution

6.4.1 Considérations sur la variation des taux d'évolution

L'un des résultats les plus surprenants obtenus par notre estimation des taux de mutation du cytb concerne les deux ordres de grandeurs obtenus entre les espèces de mammifères. En effet, une variation aussi extrême n'avait jamais été reportée. On peut supposer que ces variations n'avaient précédemment pas été trouvées pour des raisons qui varient d'une étude à l'autre : (i) utilisation de jeu de données plus petits (Martin *et al.*, 1992; Martin & Palumbi, 1993), (ii) absence de calibration fossile

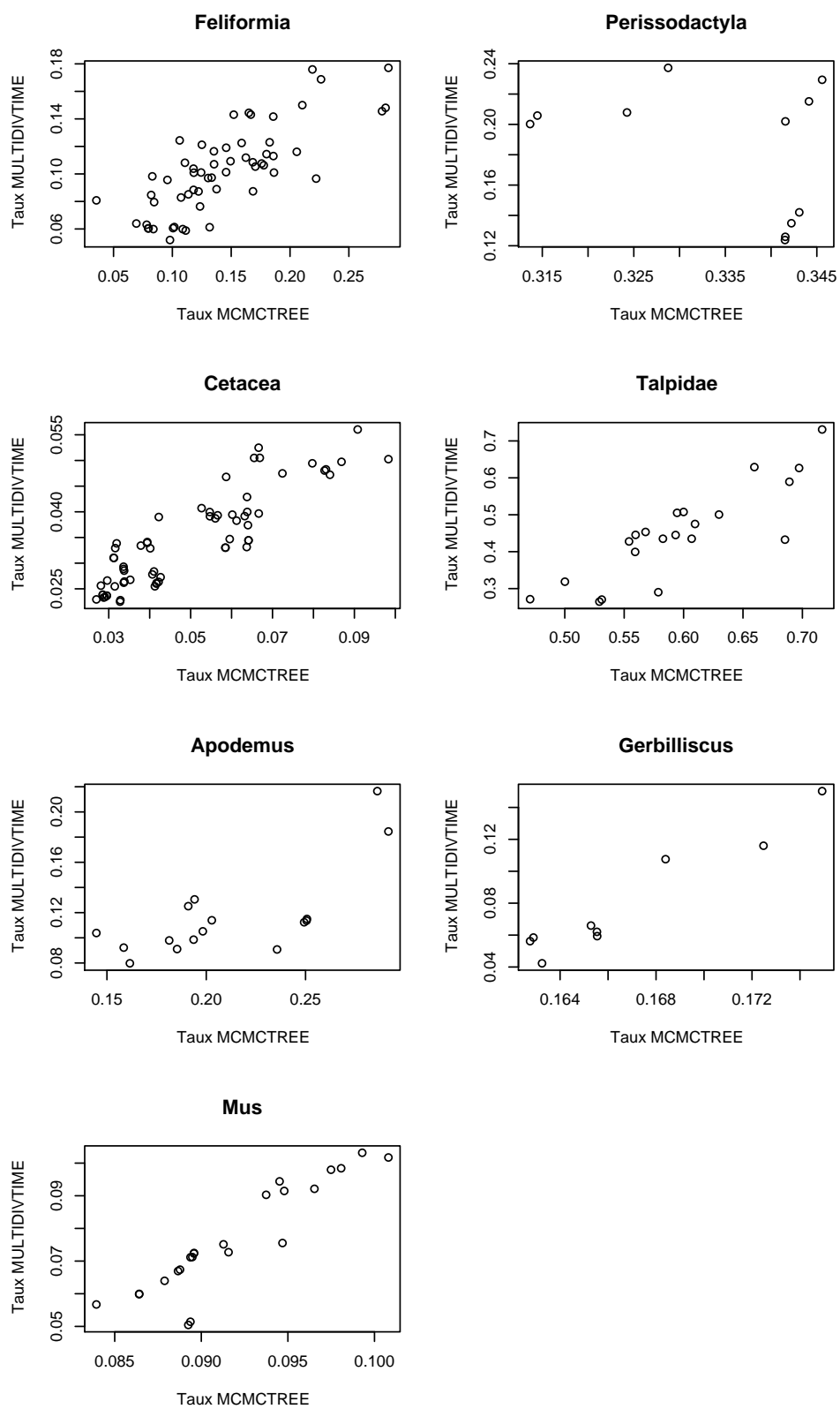


FIG. 6.8 – Résultats dans 7 groupes taxonomiques des taux d'évolution obtenus avec MULTIDIVTIME (F84, taux autocorrélés) vs MCMCTREE (F84, taux autocorrélés).

aboutissant à des taux relatifs non comparables entre espèces (Mindell *et al.*, 1996; Bromham *et al.*, 1996), (iii) enfin, et c'est le point où notre méthodologie amène un apport substantiel, un traitement inapproprié de la saturation (Gissi *et al.*, 2000; Bininda-Emonds, 2007). Chez les oiseaux, la variation est moins étendue que chez les mammifères mais reste non négligeable. En conséquence, l'hypothèse d'horloge est à proscrire, y compris chez les oiseaux, groupe dans lequel cette hypothèse est régulièrement utilisée à travers la règle des 2% de divergence par millions d'années (voir chapitre 5.2).

Ce résultat a des implications pratiques importantes en génétique des populations, évolution moléculaire, et en datation moléculaire. Par exemple, le *cytb* ne permettra pas de réaliser une phylogénie des rongeurs si des relations supérieures à la famille sont considérées, alors qu'il peut être efficace au niveau de l'ordre chez les cétacés. En datation moléculaire, ce résultat souligne tout l'intérêt de l'utilisation de méthodes autorisant une variation des taux entre lignées, méthodes dites d'horloge relâchée. Des articles récents avec des résultats importants ont été obtenus en faisant l'hypothèse de l'horloge moléculaire (Castresana, 2001; Weir & Schluter, 2007). Ces résultats méritent une ré-évaluation en prenant en compte les variations extrêmes reportées dans cette thèse.

Enfin, concernant le mécanisme mutationnel, il faut rappeler que notre estimation des taux d'évolution est effectuée à partir de toutes les troisièmes positions des codons du *cytb*. Cette moyenne des taux de mutation effectuée sur tous les sites ne reflète pas la variation entre sites. On sait par exemple que des points chauds mutationnels existent dans les séquences codantes de l'ADN mitochondrial des mammifères (Galtier *et al.*, 2006). Cette remarque est encore plus valable si on considère la dynamique mutationnelle du génome mitochondrial complet. La position physique des sites sur le génome est, par exemple, connue pour influencer le taux de mutation (Faith & Pollock, 2003). Une approche génomique permettant de décomposer la variance entre sites et la variance entre lignées nécessite un effort de séquençage de génomes mitochondriaux à l'intérieur de groupes à évolution rapide pour compléter les génomes "lents" déjà disponibles (*e.g.* cétacés et carnivores).

6.4.2 Considérations sur le déterminisme des taux d'évolution

Dans cette étude, nous avons essayé de discriminer entre les hypothèses explicatives à travers l'utilisation de traits d'histoire de vie. Mais avant de passer à l'étape de discrimination, nos résultats rappellent, en premier lieu, la forte influence des traits d'histoire de vie, pris dans leur globalité, sur les variations des taux d'évolution. En effet, il ne faut pas oublier que cette influence n'est pas retrouvée dans toutes les analyses (Gissi *et al.*, 2000; Thomas *et al.*, 2006). Il apparaît donc important de souligner la forte influence des traits d'histoire de vie sur le taux de mutation à la fois chez les mammifères et les oiseaux. A ce titre, la comparaison oiseaux *vs* mammifères apporte également une preuve de l'influence des traits d'histoire de vie. En effet, la distribution plus étroite du taux de mutation chez les oiseaux peut être rapprochée de la distribution plus étroite de leurs traits d'histoire de vie. Cela constitue une preuve indirecte de l'influence des traits d'histoire de vie sur les variations du taux de mutation.

Pour renforcer l'effet des traits d'histoire de vie, il faut également noter les sources d'erreurs potentielles sur l'estimation des taux de mutation et des traits d'histoire de vie, qui nous font probablement sous-estimer les niveaux réels de corrélation. Premièrement, les traits d'histoire de vie sont mesurés sur les espèces actuelles, alors que l'estimation des taux de mutation correspond à une moyenne du taux depuis la séparation avec l'espèce soeur. L'espèce peut, par exemple, avoir grossi très récemment, créant un décalage entre le taux estimé et la valeur de la masse corporelle actuelle. Cela constitue une autre raison, avec la saturation, justifiant l'utilisation de nombreuses espèces permettant de réduire la distance entre les espèces. Cette particularité de la variable du taux de mutation entraîne une perte de puissance statistique lors de la réalisation des corrélations avec les traits d'histoire de vie (Welch & Waxman, 2008). Il faut ajouter à cela le fait qu'un changement de taux d'évolution n'est pas répercuté instantanément sur la divergence moléculaire. Les changements récents ne sont probablement pas détectables avec l'approche phylogénétique.

On peut également parler de problèmes affectant la mesure des traits d'histoire de vie. Sur les trois variables correspondant aux traits d'histoire de vie, deux sont difficiles à mesurer : la longévité et la maturité sexuelle. La variable correspondant à la longévité est, en réalité, la longévité maximale représentant, majoritairement, des mesures faites sur des animaux captifs. En pratique, les études utilisant régulièrement la longévité des animaux ont démontré que ces records de longévité correspondent à la meilleure approximation de la longévité "réelle" ou "naturelle" des espèces (de Magalhaes *et al.*, 2005). Cependant, il n'est pas exclu qu'il existe des variations entre la longévité maximale mesurée en captivité et la longévité "réelle" en conditions naturelles sur laquelle la sélection naturelle agit. De même, il peut exister un décalage entre la maturité sexuelle femelle et le temps de génération. En effet, le temps de génération correspond à une variable plus complexe dans les espèces à générations chevauchantes (Charlesworth, 1994). Au contraire, la masse corporelle est facile à mesurer et ne repose sur aucune approximation. Ainsi, dans la confrontation entre les différentes hypothèses explicatives, la masse corporelle, et donc l'hypothèse du taux métabolique, part avec un avantage.

Après ces considérations générales, nous pouvons nous focaliser sur la discrimination entre les différentes hypothèses expliquant les variations de taux de mutation entre lignées. Nous avons vu que ces hypothèses, qui ne sont pas nécessairement exclusives, sont difficiles à séparer à cause de la forte liaison entre les différents traits d'histoire de vie utilisés comme marqueurs de ces hypothèses. Il n'est donc pas étonnant de constater que nos analyses statistiques ne nous permettent pas de réfuter fortement une hypothèse particulière. Cependant, certains motifs se détachent de nos résultats.

Ainsi, l'hypothèse du taux métabolique semble la moins soutenue par l'ensemble de nos résultats. Nous n'avons pas retrouvé la pente de 0,25 prédite par l'étude de Gillooly *et al.* (2005) à la fois chez les mammifères et lorsque les mammifères et les oiseaux sont regroupés dans la même analyse. Par contre, cette pente est retrouvée dans l'étude sur les oiseaux, où, comparée uniquement à la longévité, la masse corporelle est un meilleur prédicteur des variations du taux de mutation. Cependant, la comparaison oiseaux *vs* mammifères fournit une preuve contre l'hy-

pothèse du taux métabolique. Les oiseaux présentent un taux métabolique masse-spécifique plus élevé que les mammifères, combiné à une masse en moyenne plus faible ce qui devrait induire un taux de mutation, en moyenne, plus élevé que les mammifères (Gillooly *et al.*, 2005). Or nos résultats montrent la relation inverse (chapitre 4.1 et plus bas). Enfin, l'ensemble des critiques formulées au chapitre 4.1, et bien synthétisées par Lanfear *et al.* (2007), rend cette hypothèse conceptuellement peu vraisemblable.

Considérons à présent le temps de génération. Malheureusement, il n'a pas pu être testé chez les oiseaux. Chez les mammifères, il constitue le meilleur prédicteur du taux d'évolution lorsque les hypothèses sont considérées séparément, mais son effet diminue chez les espèces longévives au détriment de la longévité. De plus, une des prédictions de l'hypothèse du temps de génération est que son effet doit agir avec la même intensité à la fois sur le génome mitochondrial et sur le génome nucléaire, ce qui n'est pas le cas (voir chapitre 6.2.2 et Welch *et al.*, 2008). Cependant, cela ne disqualifie pas forcément l'hypothèse du temps de génération à cause du mode de transmission différent des deux génomes. Si les mutations de la lignée germinale nucléaire surviennent principalement chez les mâles (l'hypothèse d'évolution "conduite par les mâles", Li *et al.*, 2002) et si le nombre de réplication, par génération de la lignée germinale mâle est moins variable entre espèces que le nombre de réplifications de la lignée germinale femelle, alors la gamme de variation des taux d'évolution du génome nucléaire peut être plus réduite. Des données embryologiques concernant les processus de réplifications des cellules germinales mâles et femelles dans plusieurs espèces de mammifères seraient utiles pour clarifier cette question. Un autre argument en défaveur de l'hypothèse du temps de génération est l'existence de variations fortes dans la composition en bases entre lignées de mammifères (Gibson *et al.*, 2005; Urbina *et al.*, 2006). Ces variations ne sont pas attendues si le processus de substitution synonyme mitochondrial est simplement le reflet d'un effet d'échelle entre les générations et les années.

Enfin, les résultats intra-mammifères et provenant de la comparaison oiseaux *vs* mammifères supportent l'hypothèse d'une réduction sélectionnée du taux de mutation mitochondrial chez les espèces longévives. Ce résultat est en accord avec le rôle des mutations mitochondriales dans le vieillissement (voir Kujoth *et al.*, 2007 pour une revue de la question). Chez les mammifères, le taux de mutation mitochondrial est moins variable et mieux corrélé à la longévité chez les espèces longévives que chez celles à durée de vie courte. Ce résultat ne peut être expliqué par aucune des deux autres hypothèses. Il est intéressant de constater que Welch *et al.* (2008) sont arrivés à des conclusions similaires en utilisant une méthodologie plus directe avec un jeu de données de génomes mitochondriaux complets. La comparaison oiseaux *vs* mammifères fournit également un argument en faveur de l'hypothèse de la longévité. Le taux de substitution mitochondrial des oiseaux est, en moyenne, plus faible que celui des mammifères, et cette différence augmente si la masse corporelle est prise en compte (tableau 6.4). Par exemple, si l'on considère uniquement les espèces de petite taille (masse < 500g), les oiseaux évoluent 4 fois plus lentement que les mammifères. Ainsi, le taux de mutation des oiseaux s'accorde mieux à leur longévité qu'à leur masse corporelle. Similairement, les chauves-souris, qui sont les seules espèces de mammifères pratiquant le vol battu, ont des traits d'histoire de vie

proche des oiseaux (Brunet-Rossini & Austad, 2004). Elles présentent une masse corporelle en moyenne 7 fois plus faible que les rongeurs présents dans notre jeu de données, mais elles vivent en moyenne 2,4 plus longtemps. Il est intéressant de constater que les chauve-souris présentent un taux de mutation 2 fois plus faible que les rongeurs.

L'hypothèse d'une contrainte sélective pour réduire le taux de mutation mitochondrial des espèces longévives est séduisante et supportée par les connaissances sur la physiologie et le vieillissement mitochondrial. Cette hypothèse semble en accord avec nos résultats, à l'exception notable de l'analyse intra-oiseaux. D'autres preuves en faveur de cette hypothèse seraient donc bienvenues. Répliquer l'analyse sur d'autres groupes taxonomiques ne paraît pas évident étant donné que peu de groupes taxonomiques sont aussi bien connus que les oiseaux et les mammifères. Les difficultés engendrées par le peu de séquences disponibles et le manque d'information sur les traits d'histoire de vie rendent cette perspective inenvisageable à court terme. Une perspective plus intéressante consiste à étudier la génétique du contrôle des mutations mitochondriales. Un certain nombre de gènes nucléaires sont impliqués dans la réplication mitochondriale et la gestion du stress oxydatif produit par la respiration (tableau 6.6). Deux mécanismes sont possibles pour expliquer une réduction du taux de mutation chez les espèces longévives :

1. Une pression de sélection directe sur la fidélité de la réplication mitochondriale, par exemple à travers une polymérase γ plus fidèle, permettrait de limiter le nombre de mutations somatiques au cours de la vie d'un individu. Cet ajustement direct du taux de mutation mitochondrial entraînerait un retardement dans la boucle de rétro-action négative présentée précédemment (figure 4.1), impliquant la production de ROS et, *in fine*, le vieillissement.
2. La diminution du taux de mutation mitochondrial peut également passer par une meilleure gestion des ROS, par exemple à travers une détoxification par les anti-oxydants ou par le mécanisme de découplage (tableau 6.6). Ce mécanisme impliquerait également une pression de sélection mais pas directement sur la fidélité de la réplication mitochondriale (voir Kujoth *et al.*, 2007). Ce mécanisme n'explique pas la relation entre le taux de mutation et la longévité maximale des espèces, sauf si les ROS agissent aussi dans la lignée germinale (voir chapitre 4.1)

L'hypothèse de la longévité est la seule à impliquer une pression de sélection destinée à ajuster le taux de mutation mitochondrial (directement ou indirectement) en fonction de la longévité des organismes. La preuve d'une contrainte sélective plus forte ou d'événements d'adaptation agissant sur des gènes contrôlant, directement ou indirectement, le taux de mutation mitochondrial chez des espèces longévives constituerait un argument fort en faveur de l'hypothèse de la longévité, et renforcerait la théorie du vieillissement mitochondrial.

Gène	Fonction générale	Liens avec le taux de mutation mitochondrial	Références
Polymérase γ et TWINKLE	Réplication de l'ADN mitochondrial	Contrôle de la fidélité de la réplication	[1]
ANT1	Echange entre ATP et ADP dans la membrane interne de la mitochondrie	Son inactivation entraîne une production de ROS	[1]
Protéines du découplage (UCP1-3, UCP4, MCP1)	Contrôle du découplage respiration/production de chaleur de la chaîne respiratoire	Réduction de la production de ROS	[2-3]
Superoxyde dismutase (SOD)	Élimination et détoxification des ROS	Détoxification des ROS	[4]
Facteur de transcription mitochondrial A (TFAM)	Régulation de la transcription mitochondrial	TFAM enrôle l'ADN mitochondrial et pourrait être une protection contre les ROS	[1-5]

TAB. 6.6 – Quelques gènes candidats aux contrôles du taux de mutation mitochondrial. Références : [1] Kujoth *et al.* (2007), [2] Speakman *et al.* (2004), [3] Ledesma *et al.* (2002), [4] Sun *et al.* (2002), [5] Krishnan *et al.* (2007)

Troisième partie

Taille des populations et évolution moléculaire mitochondriale

Chapitre 7

Soupçons sur la neutralité de l'ADN mitochondrial

7.1 Historique

L'hypothèse de neutralité est cruciale pour reconstruire l'histoire évolutive d'une espèce. En effet, la sélection naturelle peut effacer les traces de la démographie en éliminant la variation populationnelle pour un locus (voir encadré 1). L'ADN mitochondrial, abondamment utilisé comme outil moléculaire, (voir chapitre 2.3.2) ne devrait "normalement" pas présenter de trace de sélection positive. En effet, un argument suggère que l'évolution du génome mitochondrial est conforme au modèle neutre : les gènes protéiques mitochondriaux ont des fonctions physiologiques fondamentales partagées par l'ensemble des eucaryotes. Ces fonctions ont été optimisées par la sélection naturelle depuis longtemps, et sont aujourd'hui constantes d'une espèce à l'autre. De ce fait, les variations observées aujourd'hui ont toutes les chances d'être neutres ou faiblement délétères.

Ainsi, à travers l'hypothèse de neutralité, les variations de diversité génétique mitochondriale sont classiquement interprétées comme directement sous l'influence de variations dans la taille des populations (Avice, 1994; Frankham, 1997; Spielman *et al.*, 2004). En effet, à l'équilibre mutation/dérive, la diversité génétique d'un locus neutre dans une population diploïde est égale à $4N_e\mu$ (résumé par le paramètre θ). La diversité génétique neutre est donc directement proportionnelle à l'effectif efficace N_e . Elle est également proportionnelle au taux de mutation (μ). Ce paramètre est souvent négligé car ces variations sont classiquement considérées comme beaucoup plus faibles que les variations entre les tailles de population. Cependant, les résultats de la partie 2 de cette thèse nous ont montré que les variations de taux de mutation pouvaient être très importantes entre espèces. Une bonne interprétation des patrons de la diversité génétique mitochondriale passe donc par une prise en compte des variations du taux de mutation. Enfin, avant d'analyser les patrons de diversité génétique en termes populationnels, il convient de tester si l'ADN mitochondrial évolue bien de manière conforme au modèle neutraliste. En effet, durant les dix dernières années, l'hypothèse de neutralité évolutive des variations mitochondriales a subi plusieurs attaques.

Les premières analyses se rattachant à cette problématique montrent que l'évolution des gènes protéiques mitochondriaux n'est pas complètement neutre mais plutôt dominée par la sélection purificatrice. Ces analyses conduites chez la drosophile et les mammifères proviennent de la comparaison entre le ratio des mutations synonymes et non-synonymes dans le polymorphisme (P_n/P_s) et la divergence (D_n/D_s) (Nachman, 1998; Rand & Kann, 1998; Hasegawa *et al.*, 1998; Nielsen & Weinreich, 1999). Mc Donald & Kreitman (1991) ont montré que, sous l'hypothèse de neutralité, ces deux ratios doivent être équivalents. Rand & Kann (1996) ont proposé de récapituler cette information dans un indice de neutralité (NI) :

$$NI = \frac{P_n/P_s}{D_n/D_s},$$

Sous l'hypothèse de stricte neutralité, on attend $NI = 1$, sous sélection positive $NI < 1$, et sous sélection purificatrice $NI > 1$. En effet, les allèles faiblement délétères participent en partie au polymorphisme mais sont rarement fixés ($P_n/P_s > D_n/D_s$). L'inverse est attendu pour les mutations avantageuses qui participent beaucoup à la divergence non-synonyme, et pas au polymorphisme. Dans ce dernier cas, la majorité du polymorphisme non-synonyme correspond à des variants délétères, ce qui est observé chez la drosophile et quelques mammifères (Nachman, 1998; Rand & Kann, 1998; Hasegawa *et al.*, 1998; Nielsen & Weinreich, 1999).

Récemment, la possibilité que l'ADN mitochondrial soit soumis à l'action de la sélection positive a commencé à apparaître dans la littérature. Ballard & Whitlock (2004) font justement remarquer que l'argument verbal consistant à invoquer l'importance des fonctions mitochondriales comme preuve de la neutralité pouvait également être utilisé en sens inverse pour souligner que les mutations mitochondriales peuvent avoir des effets forts sur la valeur sélective. Ces auteurs ont fait la revue des preuves expérimentales de l'influence des variations génotypiques mitochondriales sur la valeur sélective. Chez l'homme, de nombreuses maladies sont associées à des haplotypes mitochondriaux particuliers (Wallace, 2005). De même, certains changements de mode de vie, impliquant des changements physiologiques, peuvent s'accompagner d'une adaptation mitochondriale, par exemple des changements de température chez les organismes hétérothermes (Pörtner *et al.*, 2007). On a, également, évoqué précédemment l'adaptation détectée chez les primates anthropoïdes qui pourrait être due à l'augmentation de la demande énergétique associée au développement du néocortex (Grossman *et al.*, 2004), ou chez les serpents, peut-être associée à une spécialisation vers des proies plus volumineuses (Castoe *et al.*, 2008).

Dans le cas du génome mitochondrial, l'effet d'une mutation favorable est accentué par l'absence de recombinaison créant une liaison complète entre tous les sites du génomes. On ne trouve nulle part, dans le génome nucléaire des vertébrés, une région non-recombinante aussi dense en gènes, tous de potentielles cibles pour la sélection naturelle. De plus, le génome mitochondrial peut être en liaison avec d'autres gènes/génomes transmis maternellement. Cette remarque est particulièrement importante pour les invertébrés, où de nombreux micro-organismes à transmission verticale peuvent influencer le succès reproducteur de leurs hôtes (Hurst & Jiggins, 2005). Ainsi, les parasites à transmission verticale maternelle des genres *Cardinium* et *Wolbachia* sont susceptibles de modifier le sex-ratio de leurs hôtes

(des arthropodes), induisant un déséquilibre de liaison entre la lignée du parasite et l'haplotype mitochondrial. Cette sélection indirecte sur l'ADN mitochondrial a les mêmes effets qu'un événement de sélection positive directe. Toute mutation avantageuse dans le génome du symbionte qui se fixe via un balayage sélectif entraîne avec elle par auto-stop l'haplotype mitochondrial associé (par hasard) à ce cytoplasme femelle, qui se fixera aussi dans la population. De nombreux insectes ont développé des interactions positives avec des micro-organismes à transmission verticale. Ces interactions positives peuvent également déboucher sur des événements de sélection indirecte (Hurst & Jiggins, 2005). Enfin, Berlin *et al.* (2007) ont proposé que des interférences de type Hill-Robertson avec le chromosome W, chez les oiseaux, pourraient influencer la diversité et la dynamique évolutive de l'ADN mitochondrial dans ce groupe. Chez les oiseaux, la femelle est le sexe hétérogamétique. La mitochondrie des oiseaux est donc en liaison génétique complète avec le chromosome W femelle (Berlin & Ellegren, 2001). Les effets de la sélection sur des locus sur le chromosome W peuvent affecter l'ADN mitochondrial à travers un effet d'auto-stop (voir encadré 1). Le chromosome W, de la même manière que le chromosome Y, ne recombine pas et présente un taux d'évolution protéique élevé associé à une diversité intra-spécifique faible (Berlin & Ellegren, 2001, 2004).

7.2 Sélection naturelle et taille des populations

L'ensemble des considérations évoquées précédemment suggère que l'ADN mitochondrial peut être soumis à l'effet de la sélection positive. Une question intéressante consiste à savoir avec quelle fréquence ces événements de sélection positive se produisent dans les populations naturelles. Cette question est importante pour deux raisons : (i) d'une part pour comprendre l'importance des gènes codés par le génome mitochondrial dans la biologie et l'adaptation des espèces et (ii) d'autre part pour savoir si les événements de sélection positive sont assez réguliers pour causer des problèmes lors de l'utilisation du marqueur mitochondrial.

Pour répondre à cette question, Bazin *et al.* (2006) ont étudié les variations du polymorphisme intra-spécifique à l'échelle des animaux. En calculant la moyenne du polymorphisme intra-spécifique mesuré dans plusieurs grands groupes d'animaux, ces auteurs ont constaté que le niveau de polymorphisme est remarquablement constant, en ce qui concerne l'ADN mitochondrial (figure 7.1). Au contraire, la diversité estimée à partir de marqueurs nucléaires et allozymiques varie de manière importante entre les groupes, principalement en fonction du statut vertébrés *vs* invertébrés mais aussi en fonction de grands traits écologiques (par exemple marins *vs* terrestre). Cette variation peut être expliquée par l'influence de la taille des populations, les vertébrés présentant, en moyenne, des effectifs plus faibles que les invertébrés. La théorie neutraliste prédit, en effet, que le niveau de polymorphisme est corrélé avec la taille des populations. En revanche, les résultats obtenus avec l'ADN mitochondrial ne peuvent pas être expliqués par la démographie. En effet, celle-ci a un effet global sur l'ensemble des génomes, et devrait influencer de la même manière les marqueurs mitochondriaux et nucléaires. Au contraire, la sélection a un effet

La dérive et la sélection positive (et négative) sont deux forces qui ont tendance à réduire la diversité génétique. Cependant, cette réduction ne s'effectue pas de la même manière dans les deux cas. La dérive génétique présente la particularité d'avoir un effet global sur l'ensemble du génome d'un organisme. Son intensité est inversement proportionnelle à la taille de la population. Au contraire, la sélection naturelle a un effet localisé sur le génome. Elle provoque une diminution de la diversité génétique par le processus de fixation d'un allèle avantageux ; ce phénomène est appelé balayage sélectif. Cette diminution de diversité se propage mécaniquement aux locus avoisinants, en liaison génétique avec le locus sous sélection : on parle d'auto-stop génétique (Maynard-Smith & Haigh (1974), figure A). C'est donc la recombinaison qui contrôle la dimension du fragment génétique affecté par un balayage sélectif. Ainsi, juste après un balayage sélectif, la diversité autour de la mutation sous sélection ne reflète plus du tout la taille de la population. La sélection purificatrice a également un effet localisé sur la diversité génétique d'un marqueur donné. Par exemple, lors de l'élimination d'un allèle portant une mutation délétère, les locus en liaison génétique sont également éliminés on parle de sélection d'arrière-plan (Charlesworth *et al.*, 1995). Cet effet est moins important, du point de vue de la perte de diversité génétique, que la sélection positive car il ne concerne que les allèles portant la mutation contre-sélectionnée. Une relation positive entre taille des populations et diversité génétique est toujours attendue en présence de sélection d'arrière-plan (Bazin *et al.*, 2006).

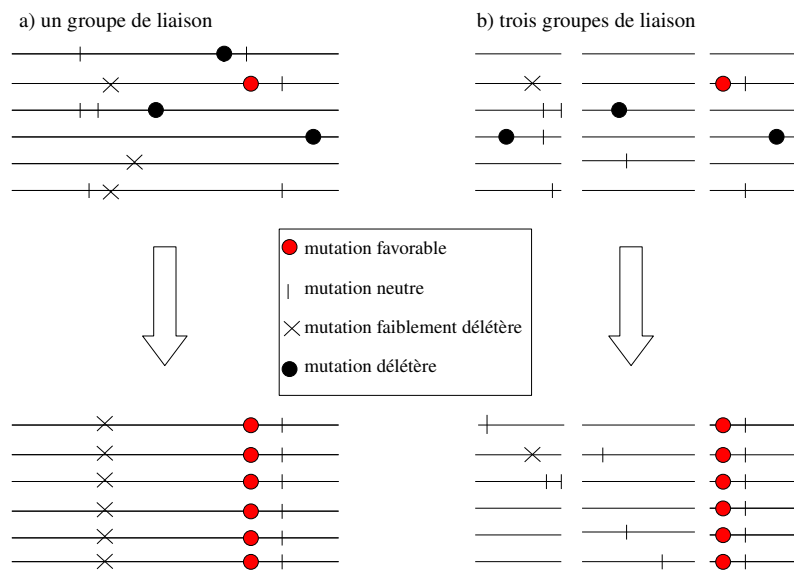


FIG. A - Influence de la liaison génétique sur les effets de la sélection naturelle. a) lorsque que tous les sites sont en liaison génétique la fixation d'une mutation avantageuse entraîne la disparition de la diversité génétique. b) Lorsque les locus sont physiquement indépendants, la fixation d'une mutation avantageuse a un effet local sur la diversité génétique.

Encadré 1: Effet de la recombinaison et de la sélection sur la diversité génétique

localisé sur le génome qui dépend de la recombinaison (voir encadré 1). Ainsi, l'homogénéité de la diversité génétique mitochondriale peut s'expliquer par la fixation récurrente de mutations favorables qui entraînent des pertes de diversité touchant spécifiquement le génome mitochondrial. L'apparition de mutations favorables est d'autant plus fréquente que la taille de la population augmente, limitant l'effet de la dérive génétique. Selon ce modèle, une relation plate, et même légèrement négative, est attendue entre la diversité génétique et la taille des populations (figure 7.2). Ce phénomène a été proposé et modélisé par Gillespie, qui lui a donné le nom de *genetic draft* (Gillespie, 2000, 2001). Pour finir, des valeurs de NI fréquemment inférieures à 1 chez les invertébrés confirment l'hypothèse sélectionniste (Bazin *et al.*, 2006).

La méta-analyse de Bazin *et al.* (2006) confirme, par une approche à large échelle taxonomique, les doutes émis précédemment sur la neutralité évolutive de l'ADN mitochondrial. Le niveau de diversité de l'ADN mitochondrial apparaît imprédictible, et semble refléter le temps écoulé depuis le dernier événement de sélection positive, plutôt que l'histoire démographique des populations.

Cependant, si la force de l'étude de Bazin *et al.* (2006) provient de son approche globale, qui a permis de dégager un motif général à travers la comparaison vertébrés *vs* invertébrés, plusieurs limitations sont également induites par ce type d'approche :

1. L'échelle taxonomique employée implique de comparer des groupes avec des biologies/écologies très différentes. Une étude réalisée à une échelle plus fine permettrait de pouvoir interpréter d'éventuels liens entre biologie et diversité génétique.
2. La partie 2 de cette thèse a révélé des variations importantes du taux de mutation entre espèces. La difficulté méthodologique inhérente à l'inférence du taux de mutation n'a pas permis d'intégrer cette variable dans Bazin *et al.* (2006). On peut se demander dans quelle mesure ces variations influencent le niveau de diversité mitochondriale.
3. L'analyse à l'échelle des animaux a surtout révélé l'effet de la sélection positive chez les invertébrés. Dans le modèle de Gillespie, l'effet de la sélection est surtout visible pour des grandes tailles de populations. Pour les effectifs plus faibles, la dérive génétique a un effet plus fort, et la diversité génétique se comporte comme dans le modèle neutre (partie gauche de la figure 7.2). On peut donc se demander si l'effet de la sélection positive est visible dans l'ensemble des groupes, et particulièrement chez les vertébrés.

L'ensemble de ces considérations nous a poussé à nous focaliser sur le déterminisme des variations de la diversité génétique, et sur l'hypothèse de neutralité, à l'intérieur de groupes taxonomiques plus restreints. Profitant des données déjà obtenues pour le taux d'évolution mitochondrial, nous avons choisi de nous focaliser sur les oiseaux et les mammifères.

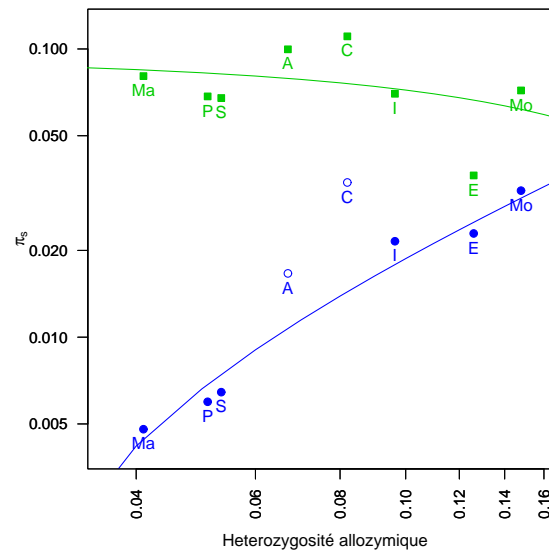


FIG. 7.1 – Diversité intraspécifique allozymique, nucléaire et mitochondriale dans 8 groupes d'animaux (moyenne). Abscisse : moyenne de l'hétérozygotie allozymique. Ordonnée : les ronds correspondent à la diversité synonyme de l'ADN nucléaire; les carrés correspondent à la diversité synonyme de l'ADN mitochondrial. Ma : *Mammalia*; S : *Sauropsida*; A : *Amphibia*; P : poissons; I : *Insecta*; C : *Crustacea*; E : *Echinodermata*; et Mo : *Mollusca*. D'après Bazin *et al.* (2006)

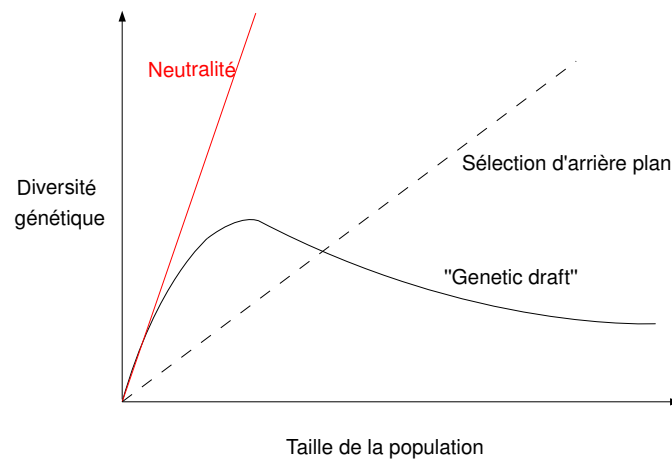


FIG. 7.2 – Relation attendue entre la taille de la population et la diversité génétique mesurée sur un locus neutre respectivement : isolé génétiquement (en rouge); en liaison avec un locus fortement sélectionné positivement (en noir, Gillespie, 2001); ou en liaison avec un locus sélectionné négativement (en pointillé).

Chapitre 8

Taille des populations et évolution moléculaire mitochondriale chez les mammifères et les oiseaux

8.1 Pas de traces de la sélection positive chez les vertébrés

Deux analyses ont été réalisées pour tester la neutralité évolutive de l'ADN mitochondrial chez les oiseaux et des mammifères. La première consiste à rechercher s'il existe une corrélation entre la diversité nucléaire et la diversité mitochondriale. Les variations démographiques ont la particularité d'affecter à la fois la diversité nucléaire et mitochondriale alors que les épisodes sélectifs n'affectent qu'un locus à la fois. Reproduisant la méthode de Bazin *et al.* (2006), nous avons utilisé les données d'hétérozygoties allozymiques, correspondant à des marqueurs nucléaires, ces données étant disponibles pour un grand nombre d'espèces suite à la popularité de ce type de marqueur dans les années 70-80 (Avise, 1994; Nevo *et al.*, 1984). Concernant la diversité mitochondriale, nous avons choisi d'utiliser un seul marqueur mitochondrial : le cytb. La diversité nucléotidique neutre (π_s , Tajima, 1989a) a été estimée à partir d'alignements obtenus via la base de données Polymorphix (Bazin *et al.*, 2005; Nabholz *et al.*, 2008a). Une corrélation positive entre la diversité neutre mitochondriale et l'hétérozygotie allozymique est observée chez les mammifères ($\tau = 2,7$, $n = 29$, $p = 0,006$, test de Kendall). Chez les oiseaux, la relation n'est pas significative mais nous ne disposons que de 11 points ($\tau = 0,07$, $n = 11$, $p = 0,75$, test de Kendall). La mise en commun des deux jeux de données aboutit à une relation positive et significative entre le polymorphisme mitochondrial et l'hétérozygotie allozymique (figure 8.1, $\tau = 3,0$, $n = 40$, $p = 0,003$). Ce résultat contraste avec le résultat obtenu à l'échelle des animaux (figure 7.1, Bazin *et al.*, 2006) et indique que la sélection positive, si elle existe dans l'évolution mitochondriale de ces groupes, n'est toutefois pas assez fréquente pour éliminer la relation entre la diversité nucléaire et mitochondriale.

L'hypothèse la plus naturelle pour expliquer cette corrélation consiste à invo-

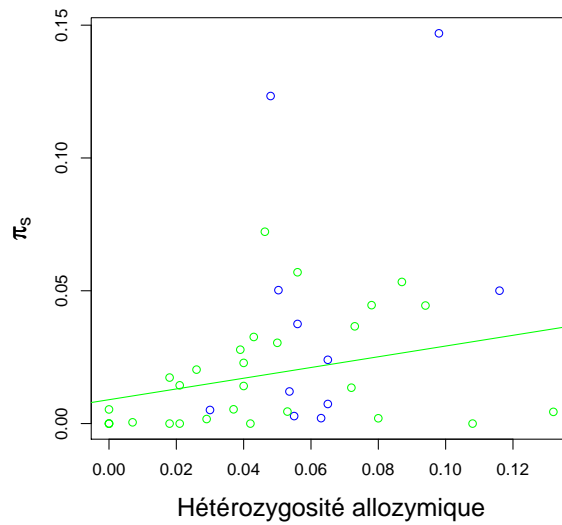


FIG. 8.1 – Relation entre la diversité nucléotidique mitochondriale neutre (π_s) et l'hétérozygotie allozymique chez 29 espèces de mammifères (points verts) et 11 genres d'oiseaux (moyenne des espèces par genre, points bleus).

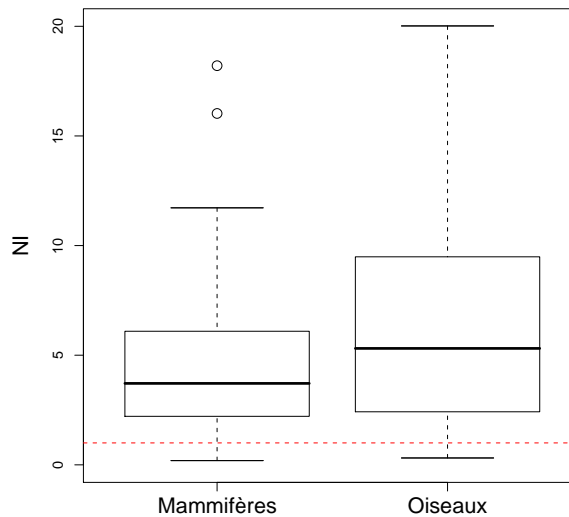


FIG. 8.2 – Distribution de l'index de neutralité (NI) chez les oiseaux ($n = 81$) et les mammifères ($n = 76$). La barre horizontale correspond à la médiane ; les limites de la boîte donnent les quartiles et les moustaches s'étendent à 1,5 fois la dimension des interquartiles, les autres points étant hors de cet intervalle. La ligne horizontale pointillée indique $NI = 1$.

quer une variation de taille des populations entre les différentes espèces. Cependant, cette relation peut également être le résultat de variations corrélées du taux de mutation entre l'ADN mitochondrial et l'ADN nucléaire. Nous avons, précédemment, reporté une corrélation positive entre les taux de mutation de quelques gènes nucléaires (*e.g.* IRPB) et du *cytb* chez les mammifères (chapitre 6.2.2). Cette relation s'explique bien dans le cadre de l'hypothèse du temps de génération, où des changements dans la fréquence des générations influencent à la fois le nombre de réplifications mitochondriales et nucléaires. Nous reviendrons sur la discrimination entre taille des populations et taux de mutation par la suite (chapitre 8.2). Dans tous les cas, cette évolution corrélée des gènes mitochondriaux et nucléaires rejette l'hypothèse d'une sélection positive récurrente, spécifique à l'ADN mitochondrial.

La deuxième analyse pour tester la neutralité évolutive de l'ADN mitochondrial consiste à étudier la distribution des indices de neutralité (NI, voir chapitre 7). Le calcul des NI nécessite la présence d'un groupe externe pour l'estimation de divergences synonymes et non-synonymes (D_n , D_s). Nous avons sélectionné ce groupe externe comme étant le plus proche possible de la séquence consensus du groupe interne. Cette sélection permet de limiter le problème de saturation nécessaire pour une estimation correcte des D_s . Toujours dans cette optique, nous avons également choisi d'estimer les D_n et D_s à l'aide d'une approche phylogénétique par maximum de vraisemblance (avec le logiciel PAML, Yang, 2007). Chez les mammifères comme chez les oiseaux, la majorité des espèces présente des NI > 1 : 91% des espèces de mammifères ($n = 75$) et 93% des espèces d'oiseaux ($n = 81$), ce qui indique la présence dominante de la sélection purificatrice dans l'évolution de l'ADN mitochondrial dans ces deux groupes (figure 8.2). De plus, les NI ne sont pas liés à la masse corporelle, variable potentiellement reliée à la taille des populations (voir plus bas). Ce résultat contraste, à nouveau, avec les résultats obtenus par Bazin *et al.* (2006), où les invertébrés (espèces à grandes tailles de populations) présentent des NI significativement plus faibles que les vertébrés (espèces à petites tailles de populations), en accord avec la théorie du "genetic draft" de Gillespie (Gillespie, 2000, 2001).

Si on compare la distribution des NI entre les oiseaux et les mammifères, on constate que les oiseaux présentent des NI en moyenne plus élevés que les mammifères ($R^2 = 0.058$, $p < 0.01$, Figure 8.2). Cette différence dans les NI provient principalement d'une différence dans le ratio π_n/π_s , significativement plus élevé chez les oiseaux (oiseaux : moyenne du $\pi_n/\pi_s = 0,103 \pm 0,102$; mammifères : moyenne du $\pi_n/\pi_s = 0,057 \pm 0,050$; $R^2 = 0,07$, $p < 0,001$), alors que les mammifères ont un ratio D_n/D_s légèrement supérieur aux oiseaux (oiseaux : $n = 81$, moyenne du $D_n/D_s = 0,019 \pm 0,011$; mammifères : $n = 77$, moyenne du $D_n/D_s = 0,028 \pm 0,049$; $R^2 = 0,01$, $p = 0,11$). Ce résultat est conforme à une hypothèse de réduction de l'efficacité de la sélection chez les oiseaux à travers un phénomène d'interférence de type Hill-Robertson (Berlin *et al.*, 2007). Il indique également un niveau de contrainte fort sur le *cytb* des oiseaux déjà reporté par Stanley & Harrison (1999).

L'ensemble de ces résultats ne supporte pas l'hypothèse d'un rôle prééminent de l'adaptation dans l'évolution du génome mitochondrial des oiseaux et des mammifères, même si ce type d'événement peut se produire ponctuellement (*e.g.* Bensch

et al., 2006 et chapitre 7). L'évolution de l'ADN mitochondrial chez ces grands vertébrés semble compatible avec le modèle d'évolution presque-neutre, où la démographie a un effet majeur sur le niveau de diversité (figure 7.2) et la probabilité de fixation des mutations neutres et faiblement délétères (figure 1.1). Les mammifères et les oiseaux semblent bien appartenir à la zone où la dérive est prédominante dans le modèle de Gillespie (partie gauche de la figure 7.2). Après avoir écarté la possibilité que la sélection positive élimine les traces de la démographie, nous pouvons nous focaliser sur l'influence du taux de mutation et de la taille des populations dans le déterminisme du polymorphisme mitochondrial.

8.2 Diversité génétique mitochondriale : quand l'effectif ne compte plus

8.2.1 Taux de mutation et polymorphisme mitochondrial

La diversité génétique neutre est déterminée par deux forces évolutives : la dérive génétique et le taux de mutation. L'absence de trace de l'adaptation dans la dynamique évolutive du *cytb* permet d'attendre la taille des populations comme le déterminant principal du polymorphisme mitochondrial. Il convient, au préalable, d'évaluer l'influence du taux de mutation, dont nous avons mis en évidence la grande variabilité (chapitre 6.1). Utilisant ces estimations, nous avons constaté l'existence d'une relation forte entre le taux de mutation et le polymorphisme mitochondrial des mammifères et des oiseaux (mammifères : $n = 123$, $R^2 = 0,14$, $p < 0,001$; oiseaux : $n = 46$, $R^2 = 0,10$, $p = 0,03$, figure 8.3). Deux conclusions peuvent être tirées de ce résultat. Premièrement, le taux de mutation apparaît comme un déterminant important de la diversité génétique neutre. Son effet devra donc être impérativement pris en compte lors de l'analyse de l'effet de la taille des populations. Deuxièmement, la relation plus faible chez les oiseaux peut, en partie, être expliquée par une interférence Hill-Robertson avec les chromosomes W qui rajoute du "bruit" dans la relation entre le taux de mutation et la diversité. Il est d'ailleurs intéressant de constater que le π_s des oiseaux est significativement plus faible que le π_s des mammifères ($R^2 = 0,09$, $p < 0,001$). Ce résultat est l'argument principal utilisé par Berlin *et al.* (2007) pour soutenir l'hypothèse d'une interférence de type Hill-Robertson. Ce résultat est conforté par l'absence de différence entre l'hétérozygotie des allozymes chez les oiseaux et les mammifères, qui va à nouveau dans le sens de Berlin *et al.* (2007) (moyenne des oiseaux : $H = 0,053 \pm 0,028$, $n = 42$; moyenne des mammifères $H = 0,046 \pm 0,033$, $n = 164$, mammals : , $R^2 = 0,01$, $p = 0,08$). Hickey (2008) fournit un commentaire particulièrement intéressant : il évoque la possibilité que la diversité faible des oiseaux pourrait être expliquée par leur faible taux de mutation plutôt que par l'interférence avec le chromosome W (le commentaire est d'autant plus remarquable que l'auteur ne disposait pas de données sur le taux d'évolution des oiseaux). Il propose que la bonne gestion des ROS par les oiseaux limite leur taux de mutation mitochondrial (voir chapitre 2.3.3 et Hickey, 2008). En effet, la différence de π_s entre les oiseaux et les mammifères est fortement réduite, mais pas totalement éliminée, lorsque le taux de mutation est pris en compte ($R^2 = 0,03$, $p = 0,03$, figure 8.4). Le

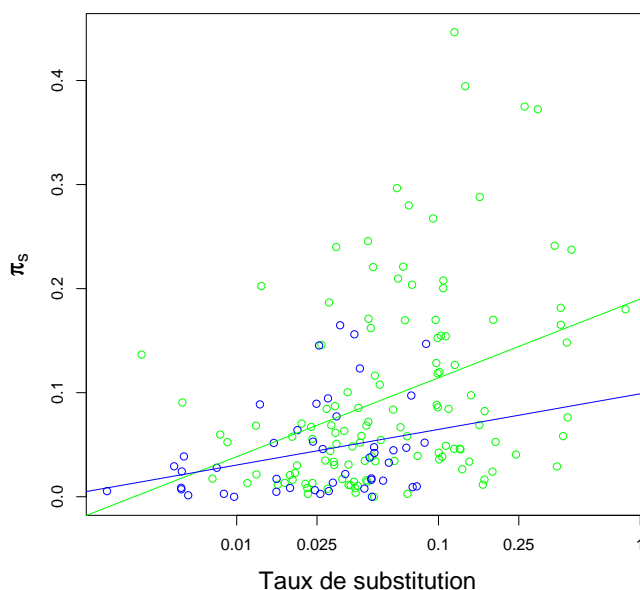


FIG. 8.3 – Relation entre la diversité nucléotidique neutre du *cytb* (π_s) et le taux de substitution neutre chez les mammifères ($n = 123$, points verts), et les oiseaux ($n = 46$, points bleus). Le taux de substitution est représenté en échelle logarithmique et en unité de substitution par site par million d'années.

taux de mutation semble expliquer une grande partie de la différence de diversité intra-spécifique moyenne entre les oiseaux et les mammifères.

8.2.2 Taille des populations et polymorphisme mitochondrial

L'évaluation de l'influence de la taille des population est rendue difficile par le peu d'informations disponibles sur ces tailles dans la majorité des espèces sauvages. Une exception notable concerne les oiseaux. L'engouement des naturalistes amateurs et professionnels pour ce groupe charismatique associé à leur observation facile (comparés aux mammifères, par exemple) a permis la mise en œuvre de programmes à large échelle géographique destinés à estimer l'abondance des espèces nicheuses. Ainsi, les espèces Nord Américaines font l'objet d'un suivi annuel depuis 1966, encadré par le North American Breeding Bird Survey (<http://www.pwrc.usgs.gov/>). A partir de ce suivi, l'extrapolation de l'effectif de nombreuses espèces strictement Nord Américaines a été rendue possible (Rich *et al.*, 2004). Ce type d'estimation n'est pas disponible pour les mammifères, où nous avons choisi d'utiliser la maximum de traits d'histoire de vie et de variables écologiques comme indicateurs potentiels de la taille des populations. 5 traits d'histoire de vie, 8 variables écologiques ainsi que le statut IUCN (Iucn, 2006) ont été choisis (Nabholz *et al.*, 2008a). Parmi les variables supposées avoir une influence substantielle sur la taille des populations, on peut citer

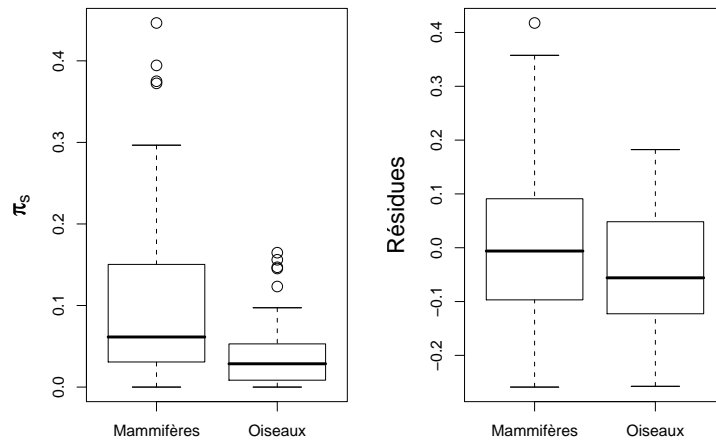


FIG. 8.4 – Diversité nucléotidique neutre de l'ADN mitochondrial (π_s) chez les oiseaux ($n = 46$) et les mammifères ($n = 123$). A gauche, la diversité nucléotidique neutre simple. A droite, les résidues de la relation entre π_s et le taux de mutation.

la masse corporelle, la taille de l'aire de répartition ou le statut IUCN. Une relation entre la masse corporelle et l'abondance semble exister à travers l'augmentation de la surface nécessaire à la demande énergétique d'une espèce avec l'augmentation de sa masse corporelle, limitant mécaniquement l'abondance de celle-ci (Damuth, 1991; Silva *et al.*, 2001; Finlay, 2002). Utilisant les données disponibles chez les oiseaux Nord Américains, nous avons retrouvé l'existence de cette relation entre la taille des populations et la masse ($n = 229$, $R^2 = 0,19$, $p < 0,001$).

De manière surprenante, aucun des traits d'histoire de vie ou des variables écologiques n'a d'effet fort sur la diversité mitochondriale (tableau 8.1). Certaines variables ont un effet significatif, comme la masse, mais qui s'avère peu robuste, par exemple face à un contrôle taxonomique (tableau 8.1) ou à un contrôle par le taux de mutation. L'absence d'effet de la masse corporelle est d'autant plus étonnant que cette variable devrait à la fois refléter les variations de taille des populations et de taux de mutation. En effet, les résultats de la partie 2 nous ont confirmé l'existence d'une corrélation négative entre la masse corporelle et le taux de mutation mitochondrial. Similairement, il est étonnant de constater que des variables pour lesquelles un lien avec la taille des populations était attendu, comme la taille de l'aire de répartition ou la classification sur liste rouge, n'ont pas d'effet sur le niveau de diversité mitochondriale (figure 8.5, Spielman *et al.*, 2004). L'analyse de modèles linéaires plus complexes combinant plusieurs variables afin de révéler d'éventuels effets cumulatifs déterminant la diversité mitochondriale ne nous ont pas permis d'identifier un déterminisme quelconque (Nabholz *et al.*, 2008a).

L'analyse chez les oiseaux confirme les conclusions obtenues chez les mammifères. La masse n'est pas corrélée à la diversité ($R^2 = 0,019$, $p = 0,30$, $n = 56$) et la taille des populations des 28 espèces strictement Nord Américaines de notre jeu de données n'a pas d'influence sur le niveau de diversité mitochondriale (figure 8.6,

	N	Sans contrôle phylogénétique			Avec contrôle Taxonomique	
		pente	R ²	<i>p</i>	R ²	<i>p</i>
Masse corporelle	107	-0,004	0,06	0,01	0,01	0,19
Maturité sexuelle	59	-0,018	0,13	<0,01	0,02	0,30
Reproduction	71	-0,032	0,03	0,13		
Longévité	53	-0,029	0,13	<0,01	0,07	0,06
Fécondité	87	0,018	0,008	0,41		
Biogéographie	138		0,115	0,05		
Régime alimentaire	107		0,195	<0,01	0,05	0,47
Taille de l'aire de répartition	117		0,04	0,16		
Habitat	108		0,0010	0,80		
Mode de vie	112		0,106	0,04	0,06	0,23
Système de reproduction	71		0,045	0,21		
Migrateur	94		0,002	0,66		
Environment	111		0,007	0,40		
Statut IUCN	138		0,03	0,82		

TAB. 8.1 – Effet des traits d’histoire de vie et de l’écologie sur la diversité mitochondriale neutre (π_s) chez les mammifères. Le contrôle taxonomique correspond à l’utilisation des résidus d’une ANOVA hiérarchisée sur la taxonomie (super-ordre, ordre, famille), ce qui correspond à retirer la variance expliquée par la taxonomie.

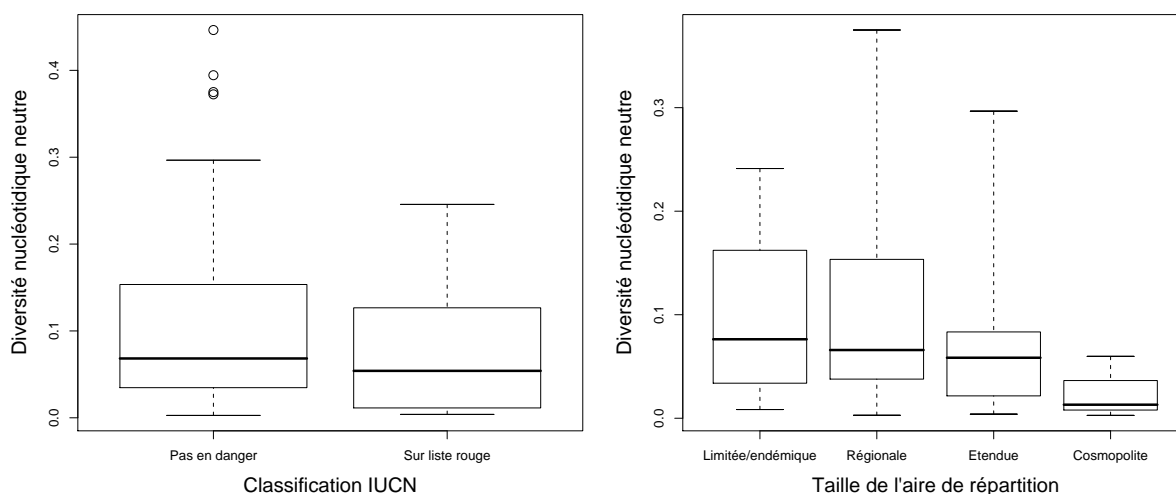


FIG. 8.5 – Relation entre la diversité nucléotidique mitochondriale des mammifères et le statut sur liste rouge (à gauche) ou la taille de l’aire de répartition (à droite).

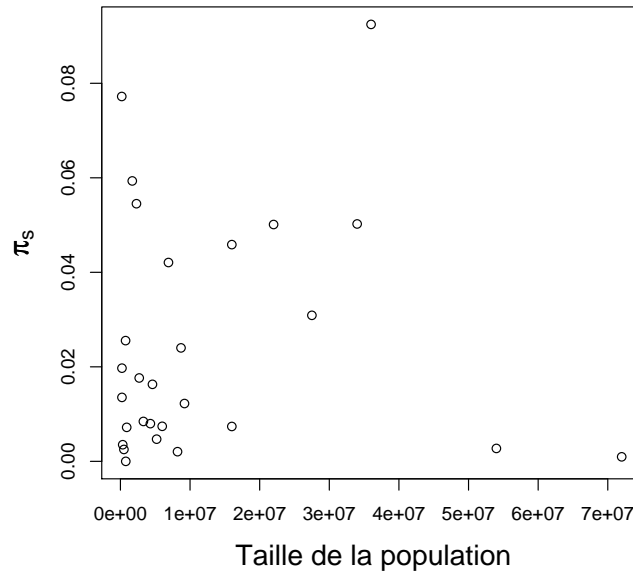


FIG. 8.6 – Relation entre la diversité nucléotidique neutre des 28 espèces d’oiseaux Nord Américaines et leurs effectifs (Rich *et al.*, 2004).

$t = 0,031$, $p = 0,81$, test de Kendall).

Les résultats obtenus apparaissent relativement contre-intuitifs et surprenants. En effet, les variations de taille des populations sont, sans aucun doute, importantes entre espèces et devraient théoriquement entraîner des variations de diversité génétique (Frankham, 1997; Spielman *et al.*, 2004). De plus, il est particulièrement étonnant de constater que le taux de mutation apparaît comme l’un des meilleurs déterminants de la diversité génétique neutre. Son effet est, par exemple, robuste à un contrôle taxonomique chez les mammifères, et il retire l’effet de la masse corporelle dans une régression multiple, suggérant que l’effet faible enregistré entre la masse et le polymorphisme pourrait être un effet du taux de mutation plutôt que de la taille des populations.

8.3 Efficacité de la sélection : quand l’effectif compte

L’absence d’effet de la taille des populations sur le polymorphisme mitochondrial est en contradiction avec les résultats récents obtenus par Popadin *et al.* (2007). Dans le cadre du modèle presque-neutre, une relation négative entre taille des populations et taux d’évolution non-synonyme est attendue (voir chapitre 1.1). Les grandes populations sont, en effet, théoriquement capables de contre-sélectionner un plus grand nombre de mutations faiblement délétères que les petites taille de populations. Partant de cette prédiction théorique, Popadin *et al.* (2007) ont examiné la relation entre le ratio D_n/D_s et la masse chez 110 espèces de mammifères. Par une approche phylogénétique et en utilisant les génomes mitochondriaux complets,

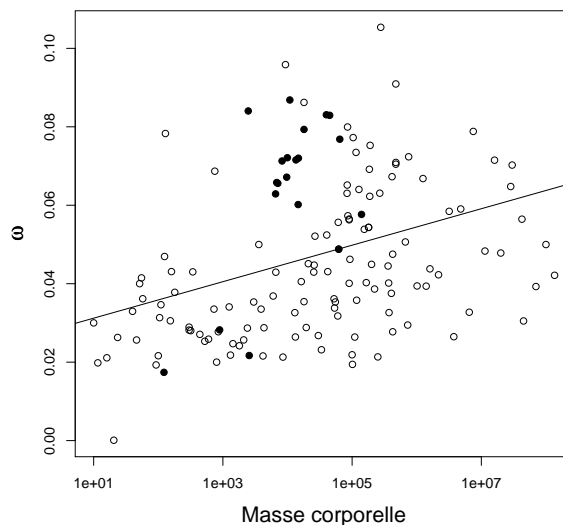


FIG. 8.7 – Relation entre le ratio D_n/D_s (ω) et la masse corporelle (en gramme) chez 147 espèces de mammifères. Le ratio D_n/D_s est estimé phylogénétiquement par la même méthode que Popadin *et al.* (2007). Les primates sont représentés par des points pleins.

ces auteurs ont assigné un D_n/D_s spécifique à chaque espèce comme étant le D_n/D_s estimé dans la branche terminale menant à l'espèce en question. Dans le contexte d'une variation forte du taux de mutation, il est important de rappeler que ce ratio présente l'avantage de s'affranchir du taux de mutation qui affecte à la fois le D_n et le D_s . L'existence d'une forte relation positive entre la masse et le D_n/D_s espèce-spécifique est en accord avec le modèle presque-neutre, les petites espèces (par leur masse corporelle) présentant probablement, en moyenne, des tailles de population plus grandes que les grandes espèces. Reproduisant l'approche de Popadin *et al.* (2007), nous avons effectivement constaté l'existence d'une corrélation positive entre la masse et le D_n/D_s de 147 espèces de mammifères ($R^2 = 0,13$, $p < 0,001$, figure 8.7). L'amélioration de cette relation par le retrait des primates du jeu de données ($n = 125$, $R^2 = 0,18$, $p < 0,001$) confirme ce résultat, ce groupe étant connu pour son évolution mitochondriale particulière (Grossman *et al.*, 2004).

Ce résultat reliant la taille des populations avec l'efficacité de la sélection apparaît d'autant plus fort que : (i) la saturation a tendance à biaiser le D_n/D_s vers des valeurs hautes pour les espèces avec un taux de mutation élevé (par une sous-estimation du D_s), donc, en moyenne, les petites espèces ; (ii) comme indiqué par Popadin *et al.* (2007), ce résultat est robuste à un contrôle taxonomique (c'est à dire en retirant l'effet expliqué par les niveaux taxonomiques super-order/ordre/famille) ; (iii) et enfin, contrairement à l'étude sur la diversité, nous avons utilisé la masse corporelle de manière "brutale" comme marqueur de la taille des populations sans faire d'effort pour raffiner notre analyse.

Afin de pouvoir intégrer les données sur les oiseaux et de limiter l'effet de la saturation, nous avons choisi d'examiner à nouveau cette relation en utilisant les données cytb. Nous avons découpé notre jeu de données en groupes correspondant

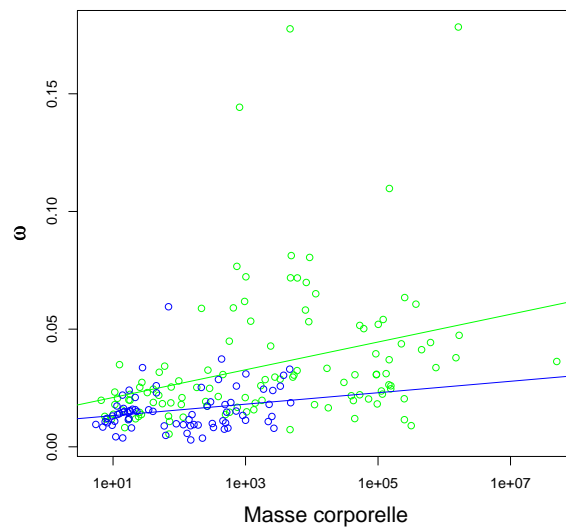


FIG. 8.8 – Relation entre le ratio D_n/D_s (ω) et la masse corporelle moyenne (en gramme) chez 117 genres de mammifères (en vert) et 75 genres d'oiseaux (en bleu). Le ratio D_n/D_s est contraint à être constant dans chacun des genres.

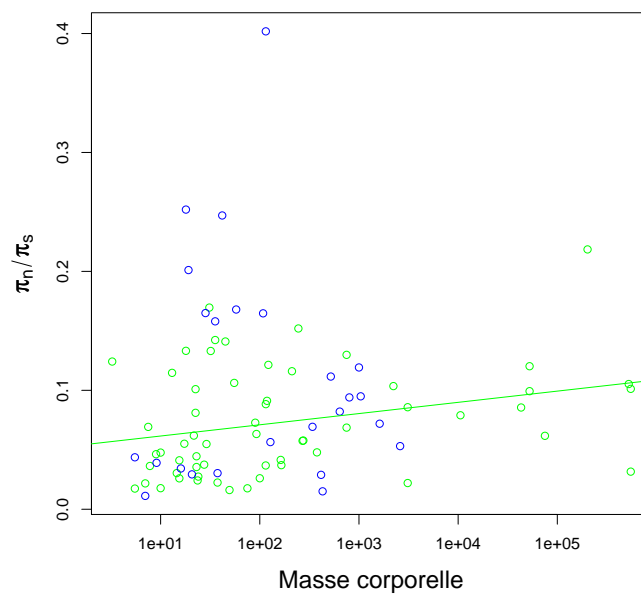


FIG. 8.9 – Relation entre le ratio π_n/π_s et la masse corporelle moyenne (en gramme) chez 60 espèces de mammifères (en vert) et 25 espèces d'oiseaux (en bleu).

aux genres et nous avons contraint le D_n/D_s à être constant à l'intérieur de chacun de ces genres (traités indépendamment les uns des autres, les genres monospécifiques étant exclus de l'analyse). Globalement la relation est très significative ($n = 192$, $R^2 = 0,17$, $p = < 0,001$, figure 8.8) mais apparaît beaucoup moins soutenue chez les oiseaux que chez les mammifères (oiseaux : $n = 75$, $R^2 = 0,05$, $p = 0,047$; mammifères $n = 117$, $R^2 = 0,11$, $p = < 0,001$).

Nous nous sommes également intéressés à l'efficacité de la sélection dans le polymorphisme. Utilisant les alignements intra-spécifiques de cytb, nous avons corrélié le ratio π_n/π_s , calculé d'après la formule de Tajima (1989a), avec la masse corporelle (figure 8.9). Comme pour la divergence, et même de manière plus tranchée, les résultats apparaissent différents entre les oiseaux et les mammifères. Les mammifères présentent une bonne relation entre la masse et π_n/π_s , mais pas les oiseaux (mammifères : $n = 60$, $R^2 = 0,16$, $p = 0,001$, oiseaux : $n = 25$, $R^2 < 0,01$, $p = 0,68$, figure 8.9).

L'ensemble de ces résultats nous montre que la taille des populations, à travers la masse corporelle, a un effet prononcé sur l'efficacité de la sélection. Ce résultat confirme que l'effet de la sélection positive, s'il existe, n'élimine pas l'influence de la taille des populations dans la dynamique évolutive de l'ADN mitochondrial.

Cela étant dit, il nous reste à expliquer pourquoi l'effet de la taille des populations est bien visible sur le ratio π_n/π_s chez les mammifères et pas chez les oiseaux. Il est possible que cette différence provienne de la liaison génétique entre le chromosome W des femelles d'oiseaux et le génome mitochondrial (Berlin *et al.*, 2007, voir chapitre 8.1). La mise en commun de données de diversité, de divergence et de taux de mutation, nous a permis d'apporter un nouvel éclairage sur l'hypothèse d'une interférence de type Hill-Robertson avec le chromosome W des oiseaux (Berlin *et al.*, 2007). Cette hypothèse séduisante avait suscité quelques réactions (Marais, 2007; Hickey, 2008; Lane, 2008). Nous avons ainsi pu confirmer l'intuition de Hickey (2008) en montrant que le taux de mutation explique en grande partie la différence de polymorphisme mitochondrial entre les oiseaux et les mammifères. Néanmoins, l'ensemble des comparaisons concernant l'efficacité de la sélection, qui s'affranchissent du taux de mutation, associé à un effet de la taille des populations sur l'efficacité de la sélection en général plus faible chez les oiseaux vont bien dans le sens de l'hypothèse de Berlin *et al.* (2007). Ce résultat issu d'une seule comparaison oiseaux/mammifères incite à l'étude d'autres couples de clades au déterminisme du sexe hétérogamique mâle et femelle, comme certains téléostéens (Mank *et al.*, 2006), afin de déterminer le degré de généralité de cette hypothèse.

8.4 Interprétation des résultats : l'instabilité populationnelle

Suite à l'ensemble de nos résultats sur l'influence de la taille des populations nous nous retrouvons face à un paradoxe. En effet nos résultats montrent que : (i) l'ADN mitochondrial des oiseaux et des mammifères ne montre pas de trace préminente de sélection positive, (ii) la taille des populations influence l'efficacité

de la sélection, (iii) mais elle n'influence pas le polymorphisme et enfin (iv) le taux de mutation influence le polymorphisme. Deux questions ressortent naturellement : 1°) Pourquoi le polymorphisme est-il influencé par la mutation, et non pas par la taille des populations ? 2°) Pourquoi l'effet de la taille des populations est visible sur l'efficacité de la sélection et pas sur le polymorphisme ?

Concernant cette deuxième question, une explication possible pour l'absence d'influence de la taille des populations sur la diversité génétique mitochondriale pourrait être le manque de puissance statistique. Nous ne disposons, par exemple, que de 28 espèces avec une taille de population connue, et d'environ une centaine d'espèces en ce qui concerne les traits d'histoire de vie. Néanmoins, la relation entre taille des populations et diversité génétique neutre devrait théoriquement être forte, surtout en considérant la grande gamme de variation des tailles de populations (couvrant plus de deux ordres de grandeurs chez les oiseaux de notre jeu de données, voir figure 8.6). Concernant les traits d'histoire de vie, leur lien avec la taille des populations est peut-être ténu. Cependant, Popadin *et al.* (2007) ont utilisé la masse comme marqueur de la taille des populations chez les mammifères et, en utilisant un jeu de données de dimensions comparables au notre, obtiennent une relation forte entre la taille des populations et nombre de mutations délétères fixées. Ces résultats ont été retrouvés et étendus aux oiseaux (voir chapitre 8.3). Enfin, la relation avec le taux de mutation apparaît assez forte et, au-moins chez les mammifères, robuste à un contrôle taxonomique. La résolution du paradoxe évoqué plus haut est, sans aucun doute, à rechercher ailleurs que dans un simple manque de puissance statistique. Cependant, il est probable que les données de polymorphisme contiennent intrinsèquement plus de variance que les données de divergence, du fait à la fois de la variance du processus de coalescence (Hudson, 1991) et d'éventuels problèmes d'échantillonnage pouvant conduire à une sous-estimation de la diversité réelle.

Concernant la première question - pourquoi le taux de mutation, et non pas la taille des populations, détermine la diversité génétique des espèces d'oiseaux et de mammifères - une première partie de la réponse est à rechercher dans les résultats de la partie I de cette thèse. En effet, une influence forte du taux de mutation nécessite que les taux de mutation soient effectivement variables entre espèces. Nous avons montré que c'était effectivement le cas, particulièrement chez les mammifères. L'importance du taux de mutation dans le déterminisme de la diversité mitochondriale a probablement été sous-estimée par le passé, précisément car les variations du taux de mutation étaient sous-estimées.

Néanmoins, les tailles de population sont également, et probablement plus, variables entre espèces. Toutefois, si l'on considère que ces tailles de populations ne sont pas stables dans le temps, mais qu'elles fluctuent assez fortement, par exemple à la faveur d'un changement climatique majeur comme un épisode de glaciation (voir chapitre 1.2), alors c'est le taux de mutation qui devient le déterminant majeur de la diversité, et pas la taille des populations.

Ayant également constaté que la taille des populations n'influence pas la diversité génétique mitochondriale, Piganeau & Eyre-Walker (in prep.) ont proposé un autre mécanisme explicatif. A l'équilibre mutation/dérive, le taux de mutation qui contrôle le polymorphisme est le taux de mutation par génération, et non pas le taux

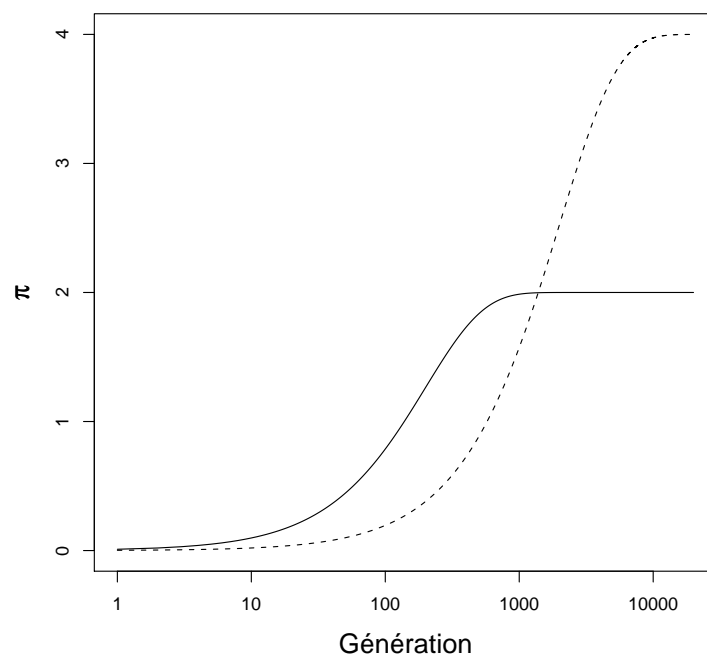


FIG. 8.10 – Evolution de la diversité nucléotidique (π) après goulot d'étranglement dans deux populations avec des paramètres différents : la population représentée par un trait plein a un effectif 10 fois plus faible mais un taux de mutation 5 fois plus élevé que la population représentée par un trait pointillé. Pendant la phase transitoire, c'est le taux de mutation qui détermine la diversité. π est calculée d'après l'équation tirée de Tajima (1989b) : $\pi = \theta + (\pi_o - \theta)^{(-t/(2N_e))}$ où $\theta = 4N_e\mu$, N_e est la taille de la population, μ le taux de mutation et π_o la diversité nucléotidique de départ (nulle dans ce cas).

de mutation par année. Les grandes espèces ont, en moyenne, un temps de génération plus long que les petites espèces. Ainsi, on peut imaginer que les grandes espèces ont un taux de mutation par génération plus important que les petites même si elles ont un taux de mutation par année plus faible. Le taux de mutation par génération peut compenser l'effet de la taille des populations et ainsi faire disparaître l'effet de la taille des populations sur la diversité. En effet, le taux de mutation par génération est corrélé positivement à la masse, et donc, probablement, négativement avec la taille des populations. Cependant, le taux de mutation par génération n'est pas corrélé à la diversité génétique mitochondrial, le meilleur prédicteur du polymorphisme mitochondrial restant le taux de mutation par année et non pas par génération. Ce résultat est en accord avec l'hypothèse d'une stochasticité populationnelle forte.

Ainsi, même en l'absence d'effet de la sélection positive, la diversité de l'ADN mitochondrial des oiseaux et des mammifères est essentiellement imprédictible connaissant leur biologie et la taille actuelle de leurs populations. Selon notre hypothèse, la diversité génétique reflète plutôt le produit entre la date du dernier événement de goulot d'étranglement populationnel et le taux de mutation. Cela explique pourquoi certaines espèces peuvent avoir une aire de répartition étendue ou une forte densité tout en ne montrant qu'une faible diversité génétique (*e.g.* Hendrickson *et al.*, 2003). Notre analyse indique que, contrairement aux invertébrés, l'ADN mitochondrial des oiseaux et des mammifères, peu ou pas affecté par la sélection positive, peut toujours être utilisé comme marqueur de l'histoire populationnelle. Il n'est pas, en revanche, un bon indicateur de la taille des populations actuelles, tout comme chez les invertébrés.

Généralisant le travail de Nei & Graur (1984) et de Tajima (1989b), Iizuka *et al.* (2002) ont montré que lorsque la taille de population fluctuait aléatoirement entre deux valeurs distinctes, la diversité devient déconnectée des valeurs de taille de population, et le taux de mutation peut devenir le déterminant majeur de la diversité génétique. On peut comprendre le phénomène par le fait que la dynamique de restauration de la diversité génétique, après un goulot d'étranglement populationnel, et essentiellement dépendante du taux de mutation, alors que le niveau de diversité à l'équilibre mutation/dérive est à la fois influencé par le taux de mutation et la taille des populations (voir figure 8.10). Si l'hypothèse de stochasticité populationnelle nous permet d'expliquer les résultats sur le déterminisme du polymorphisme, il reste néanmoins à prouver que cette instabilité est compatible avec les résultats sur l'efficacité de la sélection. Comment, et sous quelles conditions, la taille de population peut influencer le D_n/D_s sans influencer le polymorphisme? Nous tenterons d'apporter des éléments de réponse à cette question dans le chapitre suivant 8.5.

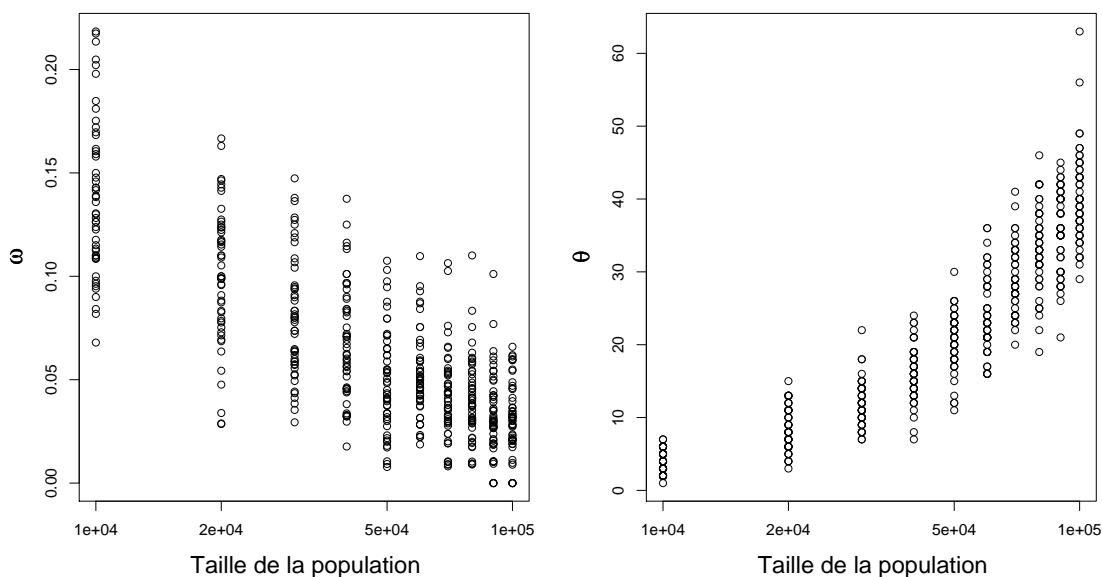


FIG. 8.11 – Effet de la taille efficace sur les patrons génomiques dans une populations à l'équilibre mutation/dérive. A gauche : Relation entre la taille des populations et le D_n/D_s (ω). $D_s = 2\mu t$ et $D_n = 2N_e p\mu t$, où μ est le taux de mutation par génération, N est l'effectif de la population, t le nombre de génération et p la probabilité de fixation d'une nouvelle mutation (Kimura, 1962). A droite : Relation entre la taille des populations et le nombre de mutations polymorphes dans une population ($\theta = 4N_e\mu$). 50 répliques ont été effectuées pour chaque taille de population en introduisant une variance poissonnienne destinée à simuler la variance mutationnelle.

8.5 Modélisation de l'évolution moléculaire en population fluctuante

8.5.1 Le modèle

Le contraste entre les résultats concernant l'influence de la taille des populations sur le diversité génétique mitochondriale neutre et sur l'efficacité de la sélection pose une question claire : Pourquoi l'effet de la taille des populations est visible sur l'efficacité de la sélection et pas sur le polymorphisme ? En effet, dans le cas d'une population idéale de Wrigth/Fisher, la taille des populations a un effet d'intensité similaire à la fois sur l'efficacité de la sélection (représenté par le ratio D_n/D_s) et sur le nombre de sites polymorphes dans la population (figure 8.11).

Dans la partie précédente, nous avons proposé que ce paradoxe pourrait être expliqué par une instabilité populationnelle forte. Réunissant des travaux précédents traitant de la probabilité de fixation d'un mutant d'une part et de l'évolution de la diversité dans des tailles de population fluctuante d'autre part, nous proposons ici de tester par simulation, de manière préliminaire, l'hypothèse verbale d'instabilité populationnelle.

Variations de la taille des populations : Nous avons choisi de reprendre le modèle de Iizuka *et al.* (2002). Considérons que la taille des populations peut fluctuer aléatoirement entre deux états (N_{max} et N_{min}). Cette fluctuation est modélisée comme un processus de Markov à deux états où γ_i correspond à la probabilité de passer de la taille de population N_i à la taille N_j . Ainsi, pour une espèce donnée, le temps passé à l'état N_i suit une distribution exponentielle de paramètre γ_i . Dans un modèle représentant une espèce avec une taille de population (N_{max}) subissant régulièrement des goulots d'étranglements (N_{min}), γ_{max} représente la fréquence des goulots et γ_{min} , l'inverse de la durée des goulots d'étranglements.

Nombre de sites polymorphes : Tajima (1989b) donne la formule permettant de calculer le nombre de sites polymorphes attendu entre deux séquences, S_2 , en fonction du temps, t , après un changement de taille de population et du nombre S_0 de sites polymorphes à l'état initial :

$$S_2 = \theta + (S_0 - \theta)^{-t/(2N_e)}, \quad (8.1)$$

où θ correspond au nombre de sites attendus à l'équilibre mutation/dérive, soit $4N_e\mu$ pour une population diploïde, et $2N_{ef}\mu$ (où N_{ef} est l'effectif efficace des femelles) dans le cas de l'ADN mitochondrial.

Divergence synonyme et non-synonyme : On définit p_t comme la probabilité de fixation d'un allèle apparaissant dans une population à la génération t . Pour une population fluctuante, Otto & Whitlock (1997) donnent l'équation de récurrence qui permet de calculer p_t connaissant p_{t+1} et la différence dans la taille de la population entre les deux générations, $\Delta N_t = N_{t+1} - N_t$:

$$1 - p_t = \exp\left[-(1+s)\left(1 + \frac{\Delta N_t}{N_t}\right)p_{t+1}\right] \quad (8.2)$$

(Ewens, 1967). Otto & Whitlock (1997) proposent un effectif efficace *de fixation*, défini comme :

$$N_{efix} = p_t N_t / 2s. \quad (8.3)$$

Connaissant la probabilité de fixation d'un nouvel allèle au temps t avec l'équation 8.2, nous pouvons en déduire l'effectif efficace de fixation (N_{efix}) pour calculer la probabilité de fixation d'un nouvelle allèle à l'aide de l'équation de Kimura (1962) :

$$P_t = \frac{1 - \exp[-4N_{efix}sf]}{1 - \exp[-4N_{efix}s]}, \quad (8.4)$$

où f est la fréquence de l'allèle mutant, soit $1/N$ pour un allèle qui apparaît dans une population.

L'ensemble de ces équations sont valable pour un allèle avec un effet favorable sur la valeur sélective de l'individu que la porte. Un allèle délétère a une probabilité de fixation plus faible, mais, comme le soulignent Otto & Whitlock (1997), *conditionnellement à la fixation*, le temps moyen de fixation est le même pour un

		$\gamma_{max} = 0,001$	
		$\gamma_{min} = 0,1$	$\gamma_{min} = 0,01$
D_n/D_s	$N_{min} = 1000$	0,80	0,66
S_2	$N_{min} = 1000$	0,76	0,40
D_n/D_s	$N_{min} = 100$	0,85	0,80
S_2	$N_{min} = 100$	0,56	0,35
D_n/D_s	$N_{min} = 10$	0,84	0,78
S_2	$N_{min} = 10$	0,42	0,34
		$\gamma_{max} = 0,01$	
		$\gamma_{min} = 0,1$	$\gamma_{min} = 0,01$
D_n/D_s	$N_{min} = 1000$	0,64	0,16
S_2	$N_{min} = 1000$	0,49	0,02
D_n/D_s	$N_{min} = 100$	0,76	0,36
S_2	$N_{min} = 100$	0,07	0,05
D_n/D_s	$N_{min} = 10$	0,78	0,41
S_2	$N_{min} = 10$	0,06	0,01

TAB. 8.2 – Valeurs des R^2 entre le ratio D_n/D_s , le nombre de site polymorphe S_2 et différentes tailles des populations. γ_{max} correspond à la probabilité de passer de N_{max} à N_{min} et inversement pour γ_{min} . Nous avons choisi de faire varier la taille des populations maximale (N_{max}) entre 5000 et 500000 ($\mu = 10^{-4}$, moyenne de $s = -10^{-4}$).

allèle délétère ou un allèle favorable. N_e mesure l'effectif efficace durant l'événement de fixation, il est donc le même pour un allèle délétère ou un allèle favorable ayant la même magnitude. Ainsi, pour n'importe quel allèle, la valeur absolue de $|s|$ peut être utilisé dans l'équation 8.2 pour déterminer N_{efix} .

8.5.2 Les simulations

Nous avons choisi les paramètres afin que les valeurs de divergence et de nombre de sites polymorphes reflètent celles que nous avons obtenues précédemment avec des données. Nous avons considéré uniquement le cas d'allèles délétères dont la valeur sélective suit une distribution gamma avec un paramètre de forme égal à 0,5 qui se situe dans la gamme de valeurs estimées à partir de données mitochondriales (Piganeau & Eyre-Walker, 2003). Pour chaque gamme de paramètre, nous avons fait varier N_{max} mais nous avons utilisé un N_{min} commun à l'ensemble des espèces.

Le tableau 8.2 récapitule l'intensité de la relation entre le D_n/D_s , le nombre de sites polymorphes S_2 et la taille de population maximale N_{max} dans différents scénarios d'instabilité populationnelle. La relation est toujours meilleure entre le D_n/D_s et le N_{max} , et la différence s'accroît surtout avec la fréquence des goulets d'étranglement. Graphiquement, l'instabilité populationnelle se traduit par une augmentation de la variance de S_2 avec l'augmentation de la taille des populations, créée par la limite supérieure de S_2 à l'équilibre mutation/dérive (figure 8.12). Il

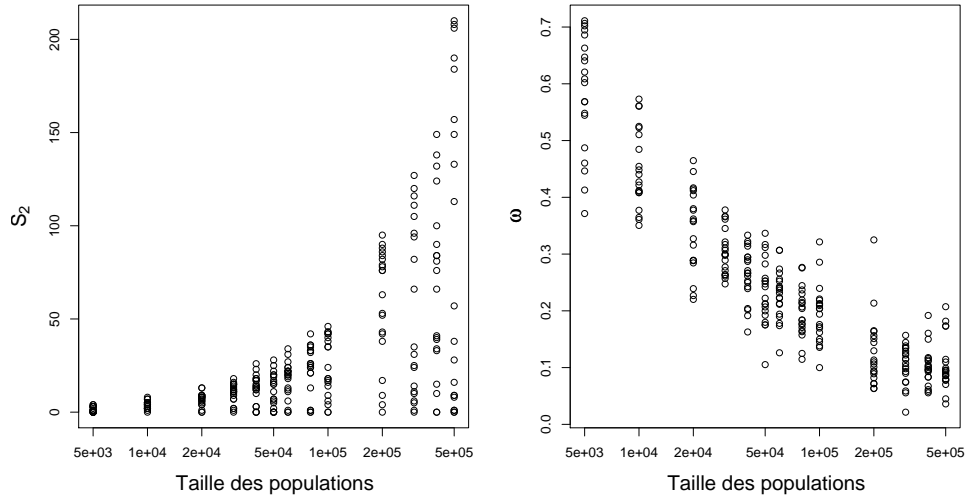


FIG. 8.12 – A gauche : relation entre la taille des populations et le nombre de site polymorphe entre deux séquence (S_2). A droite : relation entre la taille des populations et le ratio D_n/D_s (ω). Paramètre : 20 répliquions par taille de population ; 1 000 000 ; $\mu = 10^{-4}$; $\gamma_{max} = 0,001$; $\gamma_{min} = 0,01$; $N_{min} = 100$

est intéressant de constater que cette apparente augmentation de la variation avec S_2 se retrouve aussi dans les données réelles à travers la relation entre la diversité nucléotidique neutre (π_s) la masse corporelle (figure 8.13). Cette forme en “triangle” de la relation entre le π_s et la taille des populations n’est pas attendue dans le cas de l’hypothèse de Piganeau & Eyre-Walker (in prep.), où le taux de mutation par génération compense la variation dans la taille des populations, prédisant une relation plate entre π_s et la taille des populations.

Ces résultats préliminaires indiquent que la relation entre le D_n/D_s et la taille des population est beaucoup plus robuste à la stochasticité populationnelle que la relation entre le π_s et la taille des populations. Ils s’expliquent par le fait que le ratio D_n/D_s est dépendant de l’ensemble des tailles de population passées ; la taille efficace qui contrôle le nombre de fixation de mutation faiblement délétères est, en effet, proche de la moyenne harmonique des tailles de population passées. C’est également le cas de la diversité moyenne sur l’ensemble de l’histoire de l’espèce mais pas de la diversité finale. En effet, la diversité neutre finale (S_2 dans notre modèle) n’est dépendante que du temps écoulé depuis le dernier événement de goulot d’étranglement populationnel. Si ces événements sont fréquents, la diversité génétique ne reflète plus la taille de la population.

Cependant, ces résultats ne répondent pas à toutes les questions engendrées par les résultats obtenus précédemment. Des simulations plus réalistes seront nécessaires pour intégrer l’effet des variations dans le taux de mutation, l’effet du temps de génération variable entre espèces, et enfin intégrer la variance sur le S_2 liée au processus de coalescence.

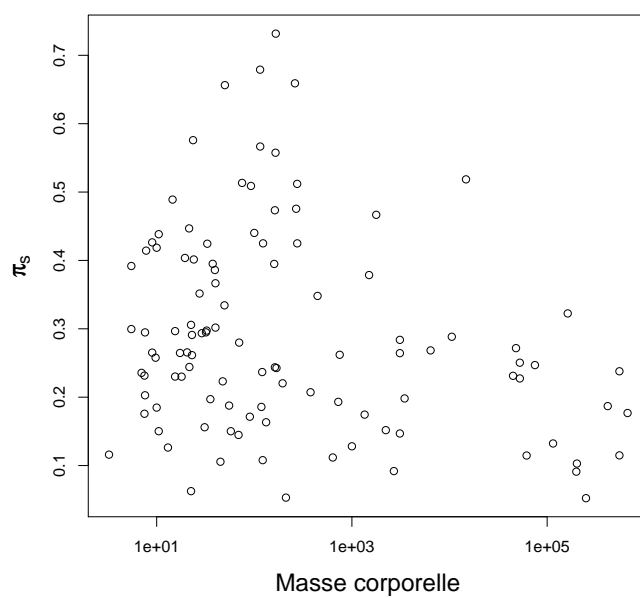


FIG. 8.13 – Relation entre la masse corporelle (en grammes) et la diversité nucléotidique synonyme (π_s) de 107 espèces de mammifères.

Quatrième partie

Conclusion

Chapitre 9

Les apports de cette thèse

We suggest that there are many interesting questions about mitochondrial molecular evolution that have been overlooked by viewing it mainly as a tool.

William J. Ballard et Michael C. Whitlock, *Molecular Ecology*, 2004.

Au cours des dix dernières années, notre vision collective de l'évolution de l'ADN mitochondrial a beaucoup évolué. L'augmentation de la qualité des données moléculaires a permis la mise à l'épreuve des hypothèses classiquement admises, jusqu'au début des années 2000, par les utilisateurs de l'ADN mitochondrial comme marqueur moléculaire (Moritz *et al.*, 1987; Harrison, 1989; Avise, 1994). Beaucoup des présuppositions traditionnelles ont ainsi été remises en question, comme la neutralité (Ballard & Whitlock, 2004; Hurst & Jiggins, 2005; Bazin *et al.*, 2006; Dowling *et al.*, 2008) ou la clonalité (Piganeau *et al.*, 2004) mais fondamentalement, ces recherches ont fait avancer la compréhension des forces évolutives qui gouvernent l'évolution de l'ADN mitochondrial. Le fait que l'ADN mitochondrial ait été largement utilisé avant que l'on s'intéresse à son évolution peut paraître paradoxal mais vient de la nécessité d'avoir à la fois une grande quantité de données génétiques, pour obtenir la puissance statistique nécessaire, et des techniques bioinformatiques permettant d'analyser ces données, pour bien comprendre l'évolution moléculaire de la mitochondrie. Cette thèse s'inscrit dans la tendance d'un approfondissement de l'étude de l'évolution de l'ADN mitochondrial à la lumière des données et techniques actuelles. Les principales avancées permises par cette thèse peuvent être récapitulées en examinant les trois principales forces évolutives qui gouvernent l'évolution de l'ADN mitochondrial.

Concernant *le taux de mutation*, l'hypothèse d'horloge moléculaire, notamment la célèbre calibration des "2% de divergence par million d'années" (Brown *et al.*, 1979), a longtemps été considérée comme valide chez les mammifères et, reste toujours débattue en ce qui concerne les oiseaux (Bermingham *et al.*, 1992; Lovette, 2004; Garcia-Moreno, 2004; Pereira & Baker, 2006; Weir & Schluter, 2007; Päckert *et al.*, 2007; Weir & Schluter, 2008). Nous avons montré que le taux de mutation mitochondrial des animaux pouvait être extrêmement variable entre espèces. Le niveau de variation reporté entre lignées de mammifères est très supérieur à ce qui

avait été estimé précédemment. Ce résultat nous apprend que l'hypothèse de l'horloge moléculaire est véritablement à bannir lorsque l'ADN mitochondrial est utilisé. Ce résultat a été obtenu par une prise en compte rigoureuse du problème de la saturation, et illustre comment les substitutions multiples peuvent rapidement dissoudre le signal évolutif, et, par exemple, créer de l'homoplasie dans un alignement de séquences mitochondriales (voir Galtier *et al.*, 2006).

D'un point de vue fondamental, cette variation pourrait être en lien avec la longévité des organismes à travers une sélection (directe ou indirecte) pour une réduction du taux de mutation mitochondrial chez les espèces longévives. Ce résultat amène beaucoup de perspectives, notamment dans la compréhension proximale du mécanisme de réduction du taux de mutation, à travers l'identification de gènes candidats chez les espèces présentant des durées de vie élevées. Ce résultat conforte également l'implication des fonctions mitochondriales dans le vieillissement, déjà connue à travers la théorie du vieillissement mitochondrial (Barja, 2004; Harman, 1957).

Enfin, le taux de mutation mitochondrial est apparu comme le déterminant principal de la diversité mitochondriale neutre. Là où la taille des populations était attendue, nous avons trouvé la mutation. En effet, et de manière surprenante, la diversité mitochondriale reste, comme chez les invertébrés, indépendante de la taille des populations.

Concernant *la sélection*, l'absence d'effet de la taille des populations sur le polymorphisme mitochondrial constaté dans la comparaison entre les vertébrés et les invertébrés va particulièrement dans le sens du renforcement de l'importance de la sélection positive dans l'évolution de l'ADN mitochondrial (Bazin *et al.*, 2006). Nous avons approfondi les résultats de Bazin *et al.* (2006) en nous focalisant sur les oiseaux et les mammifères. Nous avons montré que, dans le cas de ces grands vertébrés, la sélection positive n'agit pas fréquemment sur l'évolution de l'ADN mitochondrial. Nos résultats montrent plutôt les traces proéminentes de la sélection purificatrice. Ainsi, les résultats que nous avons obtenus pourraient être expliqués par une différence de taille de populations entre vertébrés et invertébrés conformément au modèle du "genetic draft" de Gillespie (Gillespie, 2000, 2001).

Concernant *la dérive*, nous avons abordé, à travers l'exploration de l'hypothèse de Berlin *et al.* (2007), comment des gènes en liaison génétique pouvaient interférer entre eux, notamment par une réduction de l'effectif efficace au travers des effets de type Hill-Robertson. L'un des résultats les plus surprenants obtenu durant ce travail de thèse est le contraste entre l'effet de la taille des populations sur le polymorphisme et sur l'efficacité de la sélection. En effet, nous avons trouvé que la taille des populations a un effet très faible sur le niveau de diversité neutre du cytochrome *b* des oiseaux et des mammifères, alors qu'elle a un effet fort sur l'efficacité de la sélection contrôlant la fréquence de fixation des mutations faiblement délétères. Pour expliquer ce résultat, nous avons proposé un scénario invoquant une instabilité populationnelle forte. Ce scénario a des implications au-delà du domaine pur de l'évolution moléculaire, par exemple en biogéographie, écologie et phylogéographie.

Notre conception de l'évolution de l'ADN mitochondrial a radicalement changé en l'espace d'un dizaine d'années. Cette thèse s'inscrit dans ce mouvement, notamment à travers une réappréciation de l'importance du taux de mutation et une

meilleure compréhension du déterminisme de la diversité génétique et du taux de mutation mitochondrial. Il y a fort à parier que les progrès dans la compréhension de la dynamique évolutive de l'ADN mitochondrial vont continuer avec l'augmentation du volume de données génétiques et le raffinement des théories de génétique des populations.

Chapitre 10

Traits d'histoire de vie et évolution moléculaire

Nothing makes sense in evolution except in the lighth of population genetics.

Michael Lynch, *The Origins of Genome Architecture*, 2007.

Au delà d'un bilan, cette conclusion est également l'occasion de proposer une autre perspective de lecture. Tout au long de cette thèse nous avons montré comment les traits d'histoire de vie pouvaient directement ou indirectement influencer l'évolution moléculaire. Les effets de traits d'histoire de vie sur l'évolution moléculaire peuvent être assez complexes par leurs actions multifactorielles et/ou indirectes. Par exemple, nous avons vu que la masse corporelle influence la taille des populations et agit, ainsi, à la fois sur la dérive et sur la sélection. Mais elle peut, également, influencer le taux de mutation à travers une modification du taux métabolique. On peut également citer le cas de la longévité, qui semble influencer le taux de mutation chez les espèces longévives à travers une sélection directe sur la réplication ou indirecte sur les effets mutagènes des ROS. Enfin, certains traits d'histoire de vie peuvent avoir une action plus directe sur l'évolution moléculaire. Par exemple, on peut imaginer une influence "mécaniste" du temps de génération sur le taux de mutation (figure 10.1).

Les espèces acquièrent au cours de leurs trajectoires évolutives des traits d'histoire de vie spécifiques. Ces changements imprévisibles, contrôlés par l'histoire phylogénétique, l'adaptation et la contingence, modifient les forces évolutives qui, finalement, contrôlent les patrons moléculaires (figure 10.1). Au cours de cette thèse, nous avons illustré le cas de l'évolution de l'ADN mitochondrial chez les oiseaux et les mammifères mais il existe d'autres exemples impliquant d'autres parties du génome, d'autres organismes et d'autres traits d'histoire de vie que ceux que nous avons rencontrés.

Michael Lynch a récemment proposé une série d'hypothèses assez radicales invoquant des processus non-adaptatifs, comme la mutation et la dérive, pour expliquer l'évolution de la complexité des génomes de l'ensemble des organismes (Lynch & Conery, 2003; Lynch, 2006; Lynch *et al.*, 2006; Lynch, 2007b,a). Il existe, en effet, des patrons globaux qui tendent vers une augmentation de la complexité des génomes

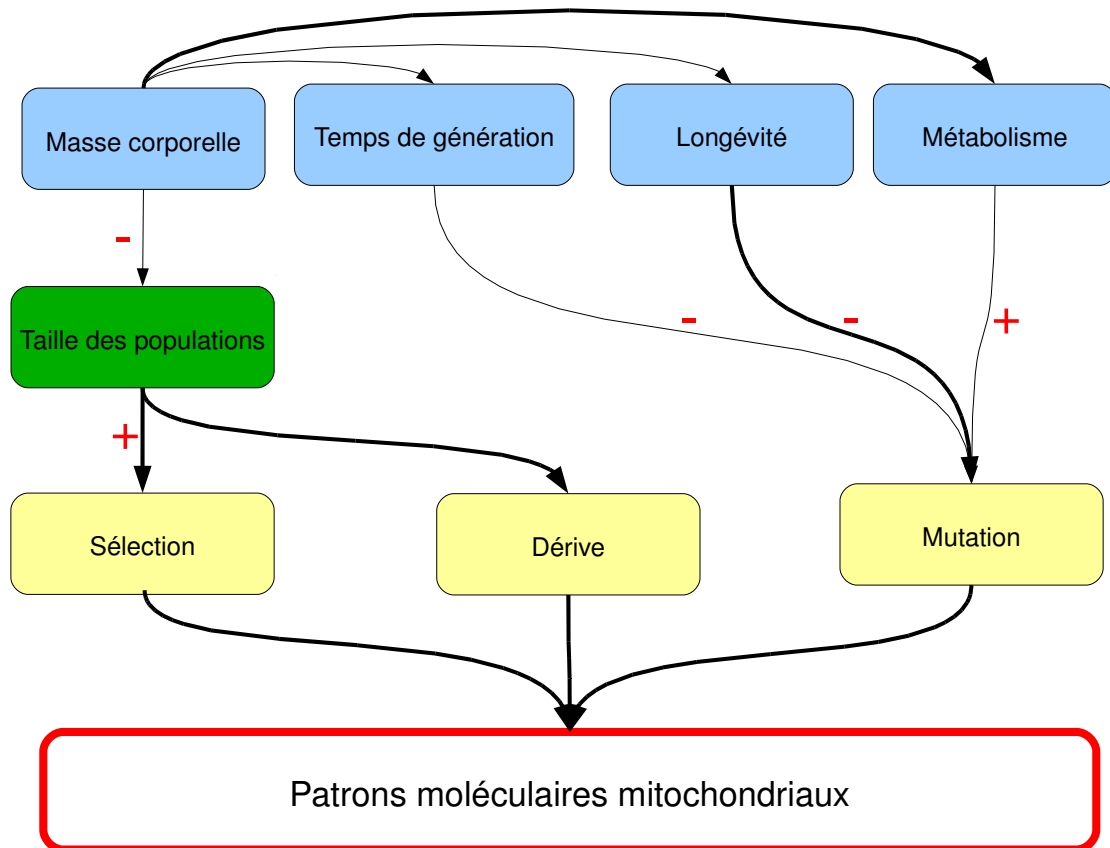


FIG. 10.1 – Schéma représentant l'influence des traits d'histoire de vie (en bleu) sur les patrons moléculaires à travers leurs actions sur les forces évolutives. La masse corporelle a un effet indirect sur la dérive et sur la sélection à travers la taille des populations et sur la mutation à travers les relations d'allométries qu'elle entretient avec les autres traits d'histoire de vie. Les variations dans les traits d'histoire de vie peuvent être positivement (+) ou négativement (-) corrélées à l'intensité des forces évolutives. L'épaisseur des flèches représente l'intensité, supposée, de la relation.

(séquences répétées, introns, duplications de gènes, augmentation du nombre de gènes notamment) avec l'augmentation de la masse corporelle des organismes. Selon Lynch & Conery (2003), cette complexité pourrait être une réponse passive à la diminution de la taille des populations qui accompagne l'augmentation de taille des organismes. Lynch *et al.* (2006) proposent également que l'augmentation de la taille du génome mitochondrial des végétaux pourrait être permise par la réduction de leur taux de mutation, notamment car la prolifération des introns pourrait être contrôlée par le produit de la taille des populations et du taux de mutation. A nouveau, on peut trouver l'influence d'un trait d'histoire de vie, à savoir le métabolisme (voir chapitre 2.3.3) sur une force évolutive, la mutation, qui influence un patron moléculaire.

Une perspective de ce travail de thèse consiste à évaluer la part des patrons moléculaires expliqués par des changements dans les traits d'histoire de vie des espèces. Les hypothèses proposées ci-dessus sont séduisantes mais traitent de différences dans les patrons moléculaires observés à une très grande échelle taxonomique (unicellulaire/métazoaire pour la complexité des génomes, plantes/animaux en ce qui concerne la taille du génome mitochondrial), et regardent essentiellement l'influence de la taille des populations sur les patrons moléculaires. Pour avoir une image complète de la dynamique évolutive des génomes il convient :

1. De se focaliser sur des différences à une échelle taxonomique fine, comme nous l'avons fait dans cette thèse. Nous avons vu comment des variations dans les traits d'histoire de vie entre espèces proches pouvaient influencer la dynamique évolutive de l'ADN mitochondrial. Paland & Lynch (2006) ont également montré comment le passage de la reproduction sexuée à la clonalité, entre populations d'une même espèce de crustacés, influence l'efficacité de la sélection.
2. D'étudier l'influence d'autres traits d'histoire de vie, par exemple, le système de reproduction, le statut libre ou parasite/symbiotique d'un organisme (Mc Cutcheon & Moran, 2007), ou d'autres variables écologiques qui restent à découvrir.
3. D'affiner, enfin, les modèles expliquant l'évolution des patrons moléculaires, notamment en prenant en compte d'autres forces évolutives que celles évoquées dans cette thèse. Par exemple, concernant les systèmes de reproduction, autogamie et allogamie chez les végétaux ou allofécondation et hermaphrodisme chez les animaux, des effets sont attendus à travers la diminution de l'effectif efficace mais également à travers l'effet de la recombinaison efficace plus faible chez les autogames et les hermaphrodites (Charlesworth & Wright, 2001). Malgré une théorie claire prédisant une réduction de l'efficacité de la sélection chez les autogames, cet effet n'est pas retrouvé dans la majorité des études sur des données réelles. Les raisons de ce décalage proviennent de l'omission dans le modèle d'autres forces évolutives comme la purge des mutations délétères (Wright *et al.*, 2008) ou le biais de conversion génique (Haudry *et al.*, 2008).

La réduction des coûts associée à l'augmentation de la capacité du séquençage laissent espérer que dans un futur proche des données de polymorphisme et de divergence soient disponibles pour un grand nombre d'espèces. Cette augmentation de l'échantillonnage permettra d'appréhender si les effets des traits d'histoire de vie sont *généralisables à l'ensemble des organismes*, et d'affiner notre compréhension des

forces évolutives agissant sur les génomes. On sait que l’évolution convergente est très fréquente dans l’histoire des organismes. Les convergences nous apportent autant des répliquions indépendantes pour étudier dans quelle mesure les mêmes changements de traits d’histoire de vie aboutissent aux mêmes changements de patron moléculaire.

Enfin, cette thèse illustre bien comment les problématiques de l’évolution moléculaire sont connectées avec d’autres problématiques de la biologie évolutive. La compréhension de l’évolution moléculaire de l’ADN mitochondrial nécessite, mais également apporte, des éléments de compréhension sur la biologie et la physiologie mitochondriale, la biologie du vieillissement, la mort cellulaire, l’évolution des traits d’histoire de vie, ou encore les caractéristiques démographiques et écologiques des populations. Ce positionnement à l’interface entre de nombreux concepts fondamentaux de la biologie évolutive rend la compréhension de l’évolution moléculaire de l’ADN mitochondrial complexe mais fascinante.

Bibliographie

- Allen, J F. 1993. Control of gene-expression by redox potential and the requirement for chloroplast and mitochondrial genomes. *Journal of Theoretical Biology*, **165**(4), 609–631.
- Allen, J F. 1996. Separate sexes and the mitochondrial theory of ageing. *Journal of Theoretical Biology*, **180**(2), 135–140. 246.
- Allen, J F, & Raven, J A. 1996. Free-radical-induced mutation vs redox regulation : costs and benefits of genes in organelles. *Journal of Molecular Evolution*, **42**(5), 482–492.
- Andersson, S G E, Zomorodipour, A, Andersson, J O, Sicheritz-Ponten, T, Alsmark, U C M, Podowski, R M, Naslund, A K, Eriksson, A S, Winkler, H H, & Kurland, C G. 1998. The genome sequence of rickettsia prowazekii and the origin of mitochondria. *Nature*, **396**(6707), 133–140. 215.
- Avise, J C. 1994. *Molecular markers, natural history and evolution*. London : Chapman and Hall.
- Avise, J C. 1998. The history and purview of phylogeography : a personal reflection. *Molecular Ecology*, **7**(4), 371–379.
- Ballard, J W O, & Whitlock, M C. 2004. The incomplete natural history of mitochondria. *Molecular Ecology*, **13**(4), 729–744. 84.
- Barja, G. 2004. Free radicals and aging. *Trends in Neurosciences*, **27**(10), 595–600.
- Barja, G., & Herrero, A. 2000. Oxidative damage to mitochondrial dna is inversely related to maximum life span in the heart and brain of mammals. *FASEB Journal*, **14**(2), 312–318.
- Barker, F K, Cibois, A, Schikler, P, Feinstein, J, & Cracraft, J. 2004. Phylogeny and diversification of the largest avian radiation. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(30), 11040–11045.
- Bazin, E., Duret, L., Penel, S., & Galtier, N. 2005. Polymorphlix : a sequence polymorphism database. *Nucleic Acids Research*, **33**(Jan 1), D481–D484. 77.
- Bazin, E, Glemin, S, & Galtier, N. 2006. Population size does not influence mitochondrial genetic diversity in animals. *Science*, **312**(5773), 570–572.

- Bell, G. 2001. Ecology - neutral macroecology. *Science*, **293**(5539), 2413–2418.
- Bensch, S, Irwin, D E, Irwin, J H, Kvist, L, & Akesson, S. 2006. Conflicting patterns of mitochondrial and nuclear dna diversity in *phylloscopus* warblers. *Molecular Ecology*, **15**(1), 161–171.
- Benton, M J, & Donoghue, P C J. 2007. Paleontological evidence to date the tree of life. *Molecular Biology and Evolution*, **24**(1), 26–53.
- Berlin, S, & Ellegren, H. 2001. Evolutionary genetics. Clonal inheritance of avian mitochondrial DNA. *Nature*, **413**(6851), 37–38.
- Berlin, S, & Ellegren, H. 2004. Chicken W : a genetically uniform chromosome in a highly variable genome. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(45), 15967–15969.
- Berlin, S, Tomaras, D, & Charlesworth, B. 2007. Low mitochondrial variability in birds may indicate Hill-Robertson effects on the W chromosome. *Heredity*, **99**(4), 389–396. 219.
- Bermingham, E, Rohwer, S, Freeman, S, & Wood, C. 1992. Vicariance biogeography in the Pleistocene and speciation in North American wood warblers : a test of Mengel's model. *Proceedings of the National Academy of Sciences of the United States of America*, **89**(14), 6624–6628.
- Bininda-Emonds, Orp. 2007. Fast Genes and Slow Clades : Comparative Rates of Molecular Evolution in Mammals. *Evolutionary Bioinformatics*, **3**, 59–85.
- Birky, C W, & Walsh, J B. 1988. Effects of linkage on rates of molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **85**(17), 6414–6418. 221.
- Birky, C W, Maruyama, T, & Fuerst, P. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics*, **103**(3), 513–527.
- Bleiweiss, R. 1998a. Relative-rate tests aid biological causes of molecular evolution in hummingbirds. *Molecular Biology and Evolution*, **15**(5), 481–491.
- Bleiweiss, R. 1998b. Slow rate of molecular evolution in high-elevation hummingbirds. *Proceedings of the National Academy of Sciences of the United States of America*, **95**(2), 612–616.
- Blouin, M S, Yowell, C A, Courtney, C H, & Dame, J B. 1998. Substitution bias, rapid saturation, and the use of mtDNA for nematode systematics. *Molecular Biology and Evolution*, **15**(12), 1719–1727.
- Boles, W.e. 1995. The World's oldest songbirds. *Nature*, **374**, 21–22. pdf.
- Bowler, P J. 2003. *Evolution. The history of an idea*. University of California press.

- Bromham, L., Rambaut, A., & Harvey, P.h. 1996. Determinants of rate variation in mammalian dna sequence evolution. *Journal of Molecular Evolution*, **43**(6), 610–621.
- Brown, W M, George, M, & Wilson, A C. 1979. Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **76**(4), 1967–1971.
- Brunet-Rossinni, A K, & Austad, S N. 2004. Ageing studies on bats : a review. *Biogerontology*, **5**(4), 211–222.
- Castoe, T A, Jiang, Z J, Gu, W, Wang, Z O, & Pollock, D D. 2008. Adaptive evolution and functional redesign of core metabolic proteins in snakes. *PLoS ONE*, **3**(5), e2201–e2201.
- Castresana, J. 2001. Cytochrome b phylogeny and the taxonomy of great apes and mammals. *Molecular Biology and Evolution*, **18**(4), 465–471.
- Catzefflis, F M, Sheldon, F H, Ahlquist, J E, & Sibley, C G. 1987. DNA-DNA hybridization evidence of the rapid rate of muroid rodent DNA evolution. *Molecular Biology and Evolution*, **4**(3), 242–253.
- Cavalier-Smith, T. 1987. Eukaryotes with no mitochondria. *Nature*, **326**(6111), 332–333.
- Chang, B H, Shimmin, L C, Shyue, S K, Hewett-Emmett, D, & Li, W H. 1994. Weak male-driven molecular evolution in rodents. *Proceedings of the National Academy of Sciences of the United States of America*, **91**(2), 827–831.
- Chao, L, & Carr, D E. 1993. The molecular clock and the relationship between population-size and generation time. *Evolution*, **47**(2), 688–690.
- Charlesworth, B. 1994. *Evolution in age-structured populations*. Cambridge university press edn.
- Charlesworth, D, & Wright, S I. 2001. Breeding systems and genome evolution. *Current Opinion in Genetics & Development*, **11**(6), 685–690.
- Charlesworth, D, Charlesworth, B, & Morgan, M T. 1995. The pattern of neutral molecular variation under the background selection model. *Genetics*, **141**(4), 1619–1632.
- Chevret, Pascale, & Dobigny, Gauthier. 2005. Systematics and evolution of the subfamily Gerbillinae (Mammalia, Rodentia, Muridae). *Molecular Phylogenetics and Evolution*, **35**(3), 674–688.
- Cho, Yangrae, Mower, Jeffrey P, Qiu, Yin-Long, & Palmer, Jeffrey D. 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(51), 17741–17746.

- Clark, C G, & Roger, A J. 1995. Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*. *Proceedings of the National Academy of Sciences of the United States of America*, **92**(14), 6518–6521.
- Clarke, Julia A, Tambussi, Claudia P, Noriega, Jorge I, Erickson, Gregory M, & Ketcham, Richard A. 2005. Definitive fossil evidence for the extant avian radiation in the Cretaceous. *Nature*, **433**(7023), 305–308.
- Clements, J F. 2007. *The Clements Checklist of Birds of the World, 6th Edition*. Cornell University Press.
- Corral-Debrinski, M, Horton, T, Lott, M T, Shoffner, J M, Beal, M F, & Wallace, D C. 1992. Mitochondrial DNA deletions in human brain : regional variability and increase with advanced age. *Nature Genetics*, **2**(4), 324–329.
- Damuth, J. 1991. Ecology - of size and abundance. *Nature*, **351**(6324), 268–269.
- Dawid, I B, & Blackler, A W. 1972. Maternal and cytoplasmic inheritance of mitochondrial-DNA in xenopus. *Developmental Biology*, **29**(2), 152–158.
- Dawson, K J. 1998. Evolutionarily stable mutation rates. *Journal of Theoretical Biology*, **194**(1), 143–157.
- Dawson, K J. 1999. The dynamics of infinitesimally rare alleles, applied to the evolution of mutation rates and the expression of deleterious mutations. *Theoretical Population Biology*, **55**(1), 1–22.
- de Magalhaes, J P, Costa, J, & Toussaint, O. 2005. HAGR : the human ageing genomic resources. *Nucleic Acids Research*, **33**(Jan 1), D537–D543.
- Delsuc, F, Vizcaíno, S F, & Douzery, E J P. 2004. Influence of Tertiary paleoenvironmental changes on the diversification of South American mammals : a relaxed molecular clock study within xenarthrans. *BMC Evolutionary Biology*, **4**(Apr, 28), 11–11.
- Denver, D R, Morris, K, Lynch, M, Vassilieva, L L, & Thomas, W K. 2000. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science*, **289**(5488), 2342–2344.
- Dowling, D K, Friberg, U, & Lindell, J. 2008. Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends in Ecology & Evolution*, **23**(10), 546–554.
- Drummond, A J, Ho, S Y W, Phillips, M J, & Rambaut, A. 2006. Relaxed phylogenetics and dating with confidence. *Plos Biol*, **4**(5), e88–e88.
- Embley, T M, Horner, D A, & Hirt, R P. 1997. Anaerobic eukaryote evolution : hydrogenosomes as biochemically modified mitochondria?. *Trends in Ecology & Evolution*, **12**(11), 437–441.

- Emerson, B C. 2007. Alarm bells for the molecular clock? No support for Ho et al.'s model of time-dependent molecular rate estimates. *Systematic Biology*, **56**(2), 337–345.
- Ericson, P G P, Anderson, C L, Britton, T, Elzanowski, A, Johansson, U S, Kallersjo, M, Ohlson, J I, Parsons, T J, Zuccon, D, & Mayr, G. 2006. Diversification of neoaves : integration of molecular sequence data and fossils. *Biology Letters*, **2**(4), 543–547.
- Estabrook, G F, Smith, G R, & Dowling, T E. 2007. Body mass and temperature influence rates of mitochondrial dna evolution in north american cyprinid fish. *Evolution*, **61**(5), 1176–1187. 241.
- Ewens, W J. 1967. Probability of survival of a new mutant in a fluctuating environment. *Heredity*, **22**, 438–443.
- Eyre-Walker, Adam. 2006. The genomic rate of adaptive evolution. *Trends in Ecology & Evolution*, **21**(10), 569–575.
- Faith, J J, & Pollock, D D. 2003. Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomes. *Genetics*, **165**(2), 735–745.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *American Naturalist*, **125**(1), 1–15.
- Felsenstien, J. 2004. *Inferring phylogenies*. Sinauer Associates.
- Finlay, B J. 2002. Global dispersal of free-living microbial eukaryote species. *Science*, **296**(5570), 1061–1063.
- Florentz, C. 2002. Molecular investigations on tRNAs involved in human mitochondrial disorders. *Bioscience Reports*, **22**(1), 81–98.
- Frankham, R. 1997. Do island populations have less genetic variation than mainland populations?. *Heredity*, **78**(Mar), 311–327.
- Furió, Victoria, Moya, Andrés, & Sanjuán, Rafael. 2005. The cost of replication fidelity in an RNA virus. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(29), 10233–10237.
- Galewski, T, Tilak, M, Sanchez, S, Chevret, P, Paradis, E, & Douzery, E J P. 2006. The evolutionary radiation of arvicolinae rodents (voles and lemmings) : relative contribution of nuclear and mitochondrial dna phylogenies. *BMC Evolutionary Biology*, **6**(Oct 9), 80–80.
- Galtier, N, Enard, D, Radondy, Y, Bazin, E, & Belkhir, K. 2006. Mutation hot spots in mammalian mitochondrial DNA. *Genome Research*, **16**(2), 215–222. 235.
- Garcia-Moreno, J. 2004. Is there a universal mtdna clock for birds?. *Journal of Avian Biology*, **35**(6), 465–468. 189.

- Gibson, A., Gowri-Shankar, V., Higgs, P.g., & Rattray, M. 2005. A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic methods. *Molecular Biology and Evolution*, **22**(2), 251–264.
- Gillespie, J H. 1991. *The cause of molecular evolution*. Oxford University Press.
- Gillespie, J H. 2000. Genetic drift in an infinite population : the pseudohitchhiking model. *Genetics*, **155**(2), 909–919.
- Gillespie, J H. 2001. Is the population size of a species relevant to its evolution ?. *Evolution*, **55**(11), 2161–2169.
- Gillespie, J H. 2004. *Population genetics : A concise guide (2 eds)*. The Johns Hopkins University Press.
- Gillooly, J F, Brown, J H, West, G B, Savage, V M, & Charnov, E L. 2001. Effects of size and temperature on metabolic rate. *Science*, **293**(5538), 2248–2251.
- Gillooly, J F, Allen, A P, West, G B, & Brown, J H. 2005. The rate of dna evolution : effects of body size and temperature on the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(1), 140–145.
- Gissi, C, Reyes, A, Pesole, G, & Saccone, C. 2000. Lineage-specific evolutionary rate in mammalian mtDNA. *Molecular Biology and Evolution*, **17**(7), 1022–1031.
- Gissi, C, Iannelli, F, & Pesole, G. 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity*, **101**(4), 301–320.
- Goetting-Minesky, M P, & Makova, K D. 2006. Mammalian male mutation bias : impacts of generation time and regional variation in substitution rates. *Journal of Molecular Evolution*, **63**(4), 537–544.
- Goldman, N, & Yang, Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution*, **11**(5), 725–736.
- Gould, S J, & Lewontin, R C. 1979. Spandrels of san-marco and the panglossian paradigm - a critique of the adaptationist program. *Proceedings Of The Royal Society Of London Series B-biological Sciences*, **205**(1161), 581–598.
- Grossman, L I, Wildman, D E, Schmidt, T R, & Goodman, M. 2004. Accelerated evolution of the electron transport chain in anthropoid primates. *Trends in Genetics*, **20**(11), 578–585.
- Haag-Liautard, C, Dorris, M, Maside, X, Macaskill, S, Halligan, D L, Houle, D, Charlesworth, B, & Keightley, P D. 2007. Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature*, **445**(7123), 82–85.
- Hackett, S J, Kimball, R T, Reddy, S, Bowie, R C K, Braun, E L, Braun, M J, Chojnowski, J L, Cox, W A, Han, K-L, Harshman, J, Huddleston, C J, Marks, B D, Miglia, K J, Moore, W S, Sheldon, F H, Steadman, D W, Witt, C C, & Yuri, T. 2008. A phylogenomic study of birds reveals their evolutionary history. *Science*, **320**(5884), 1763–1768. 242.

- Haldane, J B S. 1957. The cost of natural selection. *Journal of Genetics*, **55**, 511–524.
- Harman, D. 1957. Aging : a theory based on free radical and radiation chemistry. *Journal Of Gerontology*, **2**, 298–300.
- Harrison, R G. 1989. Animal mitochondrial-dna as a genetic-marker in population and evolutionary biology. *Trends in Ecology & Evolution*, **4**(1), 6–11.
- Hasegawa, M, Cao, Y, & Yang, Z. 1998. Preponderance of slightly deleterious polymorphism in mitochondrial DNA : nonsynonymous/synonymous rate ratio is much higher within species than between species. *Molecular Biology and Evolution*, **15**(11), 1499–1505.
- Hasegawa, M, Thorne, J L, & Kishino, H. 2003. Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. *Genes & Genetic Systems*, **78**(4), 267–283.
- Hassanin, A, & Douzery, E J P. 2003. Molecular and morphological phylogenies of ruminantia and the alternative position of the moschidae. *Systematic Biology*, **52**(2), 206–228.
- Haudry, A, Cenci, A, Guilhaumon, C, Paux, E, Poirier, S, Santoni, S, David, J, & Glémin, S. 2008. Mating system and recombination affect molecular evolution in four Triticeae species. *Genetical Research*, **90**(1), 97–109.
- Hellberg, M E. 2006. No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. *BMC Evolutionary Biology*, **6**(Mar 16), 24–24.
- Hendrickson, S I, Bleiweiss, R, Matheus, J C, de Matheus, L S, Jacome, N I, & Pavez, E. 2003. Low genetic variability in the geographically widespread andean condor. *Condor*, **105**(1), 1–12.
- Herrero, A, & Barja, G. 1998. H₂O₂ production of heart mitochondria and aging rate are slower in canaries and parakeets than in mice : sites of free radical generation and mechanisms involved. *Mechanisms of Ageing and Development*, **103**(2), 133–146.
- Hewitt, G. 2000. The genetic legacy of the quaternary ice ages. *Nature*, **405**(6789), 907–913.
- Hickey, A J R. 2008. Avian mtDNA diversity? : An alternate explanation for low mtDNA diversity in birds : an age-old solution?. *Heredity*, **100**(5), 443–443.
- Ho, S Y W. 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. *Journal of Avian Biology*, **38**(4), 409–414.
- Holmes, D J, & Ottinger, M A. 2003. Birds as long-lived animal models for the study of aging. *Experimental Gerontology*, **38**(11-12), 1365–1375.

- Holmes, D J, Flückiger, R, & Austad, S N. 2001. Comparative biology of aging in birds : an update. *Experimental Gerontology*, **36**(4-6), 869–883.
- Howell, N, Smejkal, C B, Mackey, D A, Chinnery, P F, Turnbull, D M, & Herrnstadt, C. 2003. The pedigree rate of sequence divergence in the human mitochondrial genome : there is a difference between phylogenetic and pedigree rates. *American Journal of Human Genetics*, **72**(3), 659–670.
- Hubbell, S P. 1997. A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. *Coral Reefs*, **16**, S9–S21.
- Huchon, D, & Douzery, E J P. 2001. From the old world to the new world : a molecular chronicle of the phylogeny and biogeography of hystricognath rodents. *Molecular Phylogenetics and Evolution*, **20**(2), 238–251.
- Hudson, R. 1991. *Gene genealogies and the coalescent process*. Vol. 7. 200. Pages 1–44.
- Hurst, G D D, & Jiggins, F M. 2005. Problems with mitochondrial dna as a marker in population, phylogeographic and phylogenetic studies : the effects of inherited symbionts. *Proceedings Of The Royal Society B-biological Sciences*, **272**(1572), 1525–1534.
- Hutchiso, C A, Newbold, J E, Potter, S S, & Edgell, M H. 1974. Maternal inheritance of mammalian mitochondrial-DNA. *Nature*, **251**(5475), 536–538.
- Iizuka, M, Tachida, H, & Matsuda, H. 2002. A neutral model with fluctuating population size and its effective size. *Genetics*, **161**(1), 381–388.
- Iucn. 2006. IUCN Red List of Threatened Species. Downloaded on 4 May 2006.
- Keeling, P J. 1998. A kingdom's progress : archezoa and the origin of eukaryotes. *BioEssays*, **20**(1), 87–95.
- Khaidakov, M, Heflich, R H, Manjanatha, M G, Myers, M B, & Aidoo, A. 2003. Accumulation of point mutations in mitochondrial DNA of aging mice. *Mutation Research/dna Repair*, **526**(1-2), 1–7.
- Khaidakov, M, Siegel, E R, & Reis, R J S. 2006. Direct repeats in mitochondrial dna and mammalian lifespan. *Mechanisms of Ageing and Development*, **127**(10), 808–812.
- Kim, Kyu-Heon, Eom, Keeseon S, & Park, Joong-Ki. 2006. The complete mitochondrial genome of *Anisakis simplex* (Ascaridida : Nematoda) and phylogenetic implications. *International Journal for Parasitology*, **36**(3), 319–328.
- Kimura, M. 1962. On the probability of fixation of mutant genes in a population. *Genetics*, **47**(Jun), 713–719.
- Kimura, M. 1968a. Evolutionary rate at molecular level. *Nature*, **217**(5129), 624–626.

- Kimura, M. 1968b. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genetical Research*, **11**(3), 247–269.
- Kimura, M. 1983. *The neutral theory of molecular evolution*. Cambridge University Press.
- King, J L, & Jukes, T H. 1969. Non-darwinian evolution. *Science*, **164**(3881), 788–798.
- Kishino, H, Thorne, J L, & Bruno, W J. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Molecular Biology and Evolution*, **18**(3), 352–361.
- Krieger, J, & Fuerst, P A. 2002. Evidence for a slowed rate of molecular evolution in the order acipenseriformes. *Molecular Biology and Evolution*, **19**(6), 891–897.
- Krishnan, K J, Greaves, L C, Reeve, A K, & Turnbull, D. 2007. The ageing mitochondrial genome. *Nucleic Acids Research*, **35**(22), 7399–7405.
- Ku, H H, & Sohal, R S. 1993. Comparison of mitochondrial pro-oxidant generation and anti-oxidant defenses between rat and pigeon : possible basis of variation in longevity and metabolic potential. *Mechanisms of Ageing and Development*, **72**(1), 67–76.
- Kujoth, G C, Bradshaw, P C, Haroon, S, & Prolla, T A. 2007. The role of mitochondrial DNA mutations in mammalian aging. *Plos Genetics*, **3**(2), 161–173.
- Laird, C D, Mcconaughey, B L, & Mccarthy, B J. 1969. Rate of fixation of nucleotide substitutions in evolution. *Nature*, **224**(5215), 149–154.
- Lane, N. 2005. *Power, Sex, Suicide. Mitochondria and the Meaning of Life*. Oxford University Press.
- Lane, N. 2008. Mitochondria and the W chromosome : low variability on the W chromosome in birds is more likely to indicate selection on mitochondrial genes. *Heredity*, **100**(5), 444–445.
- Lanfear, R, Thomas, J A, Welch, J J, Brey, T, & Bromham, L. 2007. Metabolic rate does not calibrate the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America*, **104**(39), 15388–15393.
- Lang, B F, Gray, M W, & Burger, G. 1999. Mitochondrial genome evolution and the origin of eukaryotes. *Annual Review of Genetics*, **33**, 351–397.
- Ledesma, A, de Lacoba, M, & Rial, E. 2002. The mitochondrial uncoupling proteins. *Genome Biol*, **3**(12).
- Lepage, T, Bryant, D, Philippe, H, & Lartillot, N. 2007. A general comparison of relaxed molecular clock models. *Molecular Biology and Evolution*, **24**(12), 2669–2680.

- Li, W H, & Tanimura, M. 1987. The molecular clock runs more slowly in man than in apes and monkeys. *Nature*, **326**(6108), 93–96.
- Li, W H, Tanimura, M, & Sharp, P M. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *Journal of Molecular Evolution*, **25**(4), 330–342.
- Li, W H, Gouy, M, Sharp, P M, O’huigin, C, & Yang, Y W. 1990. Molecular phylogeny of Rodentia, Lagomorpha, Primates, Artiodactyla, and Carnivora and molecular clocks. *Proceedings of the National Academy of Sciences of the United States of America*, **87**(17), 6703–6707.
- Li, W H, Ellsworth, D L, Krushkal, J, Chang, B H, & Hewett-Emmett, D. 1996. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Molecular Phylogenetics and Evolution*, **5**(1), 182–187.
- Li, W H, Yi, S, & Makova, K. 2002. Male-driven evolution. *Current Opinion in Genetics & Development*, **12**(6), 650–656.
- Loeb, L A, & Monnat, R J. 2008. DNA polymerases and human disease. *Nature Reviews Genetics*, **9**(8), 594–604.
- Lovette, I J. 2004. Mitochondrial dating and mixed-support for the "2% rule" in birds. *Auk*, **121**(1), 1–6. 247.
- Lynch, M. 2006. The origins of eukaryotic gene structure. *Molecular Biology and Evolution*, **23**(2), 450–468.
- Lynch, M. 2007a. *The Origins of Genome Architecture*. Sinauer associates edn.
- Lynch, M, & Conery, J S. 2003. The origins of genome complexity. *Science*, **302**(5649), 1401–1404.
- Lynch, M, Koskella, B, & Schaack, S. 2006. Mutation pressure and the evolution of organelle genomic architecture. *Science*, **311**(5768), 1727–1730.
- Lynch, M, Sung, W, Morris, K, Coffey, N, Landry, C R, Dopman, E B, Dickinson, W J, Okamoto, K, Kulkarni, S, Hartl, D L, & Thomas, W K. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, **105**(27), 9272–9277.
- Lynch, Michael. 2007b. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proceedings of the National Academy of Sciences of the United States of America*, **104 Suppl 1**(May, 15), 8597–8604.
- Madsen, O, Scally, M, Douady, C J, Kao, D J, Debry, R W, Adkins, R, Amrine, H M, Stanhope, M J, de Jong, W W, & Springer, M S. 2001. Parallel adaptive radiations in two major clades of placental mammals. *Nature*, **409**(6820), 610–614.

- Mank, J E, Promislow, D E L, & Avise, J C. 2006. Evolution of alternative sex-determining mechanisms in teleost fishes. *Biological Journal of the Linnean Society*, **87**(1), 83–93.
- Marais, G A B. 2007. The Hill-Robertson effects extend from nucleus to mitochondria. *Heredity*, **99**(4), 357–358.
- Margulis, L. 1967. On origin of mitosing cells. *Journal of Theoretical Biology*, **14**(3), 225–274.
- Margulis, L. 1981. *Symbiosis in cell evolution : life and its environment on the early earth*. San Francisco : W H Fremann.
- Martin, A P. 1995. Metabolic rate and directional nucleotide substitution in animal mitochondrial DNA. *Molecular Biology and Evolution*, **12**(6), 1124–1131.
- Martin, A P. 1999. Substitution rates of organelle and nuclear genes in sharks : implicating metabolic rate (again). *Molecular Biology and Evolution*, **16**(7), 996–1002.
- Martin, A P, & Palumbi, S R. 1993. Body size, metabolic-rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America*, **90**(9), 4087–4091.
- Martin, A P, Naylor, G J, & Palumbi, S R. 1992. Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature*, **357**(6374), 153–155.
- Martin, W. 2003. Gene transfer from organelles to the nucleus : frequent and in big chunks. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(15), 8612–8614.
- Martin, W, & Muller, M. 1998. The hydrogen hypothesis for the first eukaryote. *Nature*, **392**(6671), 37–41.
- Mason, P A, & Lightowers, R N. 2003. Why do mammalian mitochondria possess a mismatch repair activity?. *FEBS Letters*, **554**(1-2), 6–9.
- Maynard-Smith, J M, & Haigh, J. 1974. Hitch-hiking effect of a favorable gene. *Genetical Research*, **23**(1), 23–35.
- Mc Cutcheon, J P, & Moran, N A. 2007. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, **104**(49), 19392–7.
- Mc Donald, J H, & Kreitman, M. 1991. Adaptive protein evolution at the adh locus in drosophila. *Nature*, **351**(6328), 652–654.
- Mercer, J M, & Roth, V L. 2003. The effects of Cenozoic global change on squirrel phylogeny. *Science*, **299**(5612), 1568–1572.

- Michaux, J R, Chevret, P, Filippucci, M-G, & Macholan, M. 2002. Phylogeny of the genus *Apodemus* with a special emphasis on the subgenus *Sylvaemus* using the nuclear IRBP gene and two mitochondrial markers : cytochrome b and 12S rRNA. *Molecular Phylogenetics and Evolution*, **23**(2), 123–136.
- Milbury, Coren A, & Gaffney, Patrick M. 2005. Complete mitochondrial DNA sequence of the eastern oyster *Crassostrea virginica*. *Mar Biotechnol (NY)*, **7**(6), 697–712.
- Mindell, D P, Knight, A, Baer, C, & Huddleston, C J. 1996. Slow rates of molecular evolution in birds and the metabolic rate and body temperature hypotheses. *Molecular Biology and Evolution*, **13**(2), 422–426.
- Mooers, A O, & Harvey, P H. 1994. Metabolic rate, generation time, and the rate of molecular evolution in birds. *Molecular Phylogenetics and Evolution*, **3**(4), 344–350.
- Moritz, C, Dowling, T E, & Brown, W M. 1987. Evolution of animal mitochondrial-DNA - relevance for population biology and systematics. *Annual Review of Ecology and Systematics*, **18**, 269–292.
- Mower, J P, Touzet, P, Gummow, J S, Delph, L F, & Palmer, J D. 2007. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. *BMC Evolutionary Biology*, **7**, 135–135.
- Murphy, W J, Eizirik, E, Johnson, W E, Zhang, Y P, Ryder, O A, & O'brien, S J. 2001. Molecular phylogenetics and the origins of placental mammals. *Nature*, **409**(6820), 614–618.
- Murphy, W J, Pevzner, P A, & O'brien, S J. 2004. Mammalian phylogenomics comes of age. *Trends in Genetics*, **20**(12), 631–639.
- Müller, J, & Reisz, Robert R. 2005. Four well-constrained calibration points from the vertebrate fossil record for molecular clock estimates. *BioEssays*, **27**(10), 1069–1075.
- Müller, M, & Martin, W. 1999. The genome of *Rickettsia prowazekii* and some thoughts on the origin of mitochondria and hydrogenosomes. *BioEssays*, **21**(5), 377–381.
- Nabholz, B, Mauffrey, J-F, Bazin, E, Galtier, N, & Glémin, S. 2008a. Determination of mitochondrial genetic diversity in mammals. *Genetics*, **178**(1), 351–361.
- Nabholz, B, Glémin, S, & Galtier, N. 2008b. Mitochondrial DNA evolutionary dynamics : what can we learn from the bird/mammal comparison ?. *BMC Evolutionary Biology*. (en révision).
- Nabholz, B, Glémin, S, & Galtier, N. 2008c. Strong variations of mitochondrial mutation rate across mammals-the longevity hypothesis. *Molecular Biology and Evolution*, **25**(1), 120–130.

- Nachman, M W. 1998. Deleterious mutations in animal mitochondrial DNA. *Genetica*, **102-103**(1-6), 61–69.
- Nei, M, & Graur, D. 1984. Extent of protein polymorphism and the neutral mutation theory. *Evolutionary Biology*, **17**, 73–118.
- Neumann, K, Michaux, J, Lebedev, V, Yigit, N, Colak, E, Ivanova, N, Poltoraus, A, Surov, A, Markov, G, Maak, S, Neumann, S, & Gattermann, R. 2006. Molecular phylogeny of the Cricetinae subfamily based on the mitochondrial cytochrome b and 12S rRNA genes and the nuclear vWF gene. *Molecular Phylogenetics and Evolution*, **39**(1), 135–148.
- Nevo, E, Beiles, A, & Ben-Shlomo, R. 1984. The evolutionary significance of genetic diversity : Ecological, demographic and life-history correlates.. *Lecture Notes in Biomathematics 53 : Evolutionary Dynamics of Genetic Diversity*.
- Nielsen, R, & Weinreich, D M. 1999. The age of nonsynonymous and synonymous mutations in animal mtDNA and implications for the mildly deleterious theory. *Genetics*, **153**(1), 497–506.
- Nunn, G B, & Stanley, S E. 1998. Body size effects and rates of cytochrome b evolution in tube-nosed seabirds. *Molecular Biology and Evolution*, **15**(10), 1360–1371.
- Ohta, T. 1973. Slightly deleterious mutant substitutions in evolution. *Nature*, **246**(5428), 96–98.
- Ohta, T. 1987. Very slightly deleterious mutations and the molecular clock. *Journal of Molecular Evolution*, **26**(1-2), 1–6.
- Ohta, T. 1992. The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics*, **23**, 263–286.
- Ohta, T. 1993. An examination of the generation-time effect on molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **90**(22), 10676–10680.
- Ohta, T, & Gillespie, J H. 1996. Development of neutral and nearly neutral theories. *Theoretical Population Biology*, **49**(2), 128–142.
- Ohta, T, & Kimura, M. 1971. On the constancy of the evolutionary rate of cistrons. *Journal of Molecular Evolution*, **1**(1), 18–25.
- Otto, S P, & Whitlock, M C. 1997. The probability of fixation in populations of changing size. *Genetics*, **146**(2), 723–733.
- Pace, N R, Olsen, G J, & Woese, C R. 1986. Ribosomal-RNA phylogeny and the primary lines of evolutionary descent. *Cell*, **45**(3), 325–326.
- Paland, S, & Lynch, M. 2006. Transitions to asexuality result in excess amino acid substitutions. *Science*, **311**(5763), 990–992.

- Papetti, C, Liò, P, Rüber, L, Patarnello, T, & Zardoya, R. 2007. Antarctic fish mitochondrial genomes lack ND6 gene. *Journal of Molecular Evolution*, **65**(5), 519–528.
- Paradis, E. 2007. *Analysis of Phylogenetics and Evolution with R*. Springer.
- Pasquet, E, Pons, J-M, Fuchs, J, Cruaud, C, & Bretagnolle, V. 2007. Evolutionary history and biogeography of the drongos (Dicruridae), a tropical Old World clade of corvoid passerines. *Molecular Phylogenetics and Evolution*, **45**(1), 158–167.
- Pereira, S L, & Baker, A J. 2006. A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. *Molecular Biology and Evolution*, **23**(9), 1731–1740.
- Pesole, G, Gissi, C, de Chirico, A, & Saccone, C. 1999. Nucleotide substitution rate of mammalian mitochondrial genomes. *Journal of Molecular Evolution*, **48**(4), 427–434.
- Piganeau, G, & Eyre-Walker, A. 2003. Estimating the distribution of fitness effects from DNA sequence data : implications for the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(18), 10335–10340. 260.
- Piganeau, G, & Eyre-Walker, A. in prep.. Variation in the Effective Population Size of Mitochondrial DNA.
- Piganeau, G, Gardner, M, & Eyre-Walker, A. 2004. A broad survey of recombination in animal mitochondria. *Molecular Biology and Evolution*, **21**(12), 2319–2325.
- Popadin, K, Polishchuk, L V, Mamirova, L, Knorre, D, & Gunbin, K. 2007. Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. *Proceedings of the National Academy of Sciences of the United States of America*, **104**(33), 13390–13395.
- Prasad, A B, Allard, M W, Program, Nisc Comparative Sequencing, & Green, E D. 2008. Confirming the Phylogeny of Mammals by Use of Large Comparative Sequence Datasets. *Molecular Biology and Evolution*, in press—in press.
- Päckert, M, Martens, J, Tietze, D T, Dietzen, C, Wink, M, & Kvist, L. 2007. Calibration of a molecular clock in tits (Paridae)-Do nucleotide substitution rates of mitochondrial genes deviate from the 2% rule?. *Molecular Phylogenetics and Evolution*, **44**(1), 1–14.
- Pörtner, H O, Peck, L, & Somero, G. 2007. Thermal limits and adaptation in marine Antarctic ectotherms : an integrative view. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, **362**(1488), 2233–2258.
- Race, H L, Herrmann, R G, & Martin, W. 1999. Why have organelles retained genomes?. *Trends in Genetics*, **15**(9), 364–370.

- Rand, D M, & Kann, L M. 1996. Excess amino acid polymorphism in mitochondrial DNA : contrasts among genes from drosophila, mice, and humans. *Molecular Biology and Evolution*, **13**(6), 735–748.
- Rand, D M, & Kann, L M. 1998. Mutation and selection at silent and replacement sites in the evolution of animal mitochondrial DNA. *Genetica*, **102-103**(1-6), 393–407.
- Rannala, B, & Yang, Z. 2007. Inferring speciation times under an episodic molecular clock. *Systematic Biology*, **56**(3), 453–466.
- Ranwez, V, Delsuc, F, Ranwez, S, Belkhir, K, Tilak, M-K, & Douzery, E J P. 2007. OrthoMaM : a database of orthologous genomic markers for placental mammal phylogenetics. *BMC Evolutionary Biology*, **7**, 241–241.
- Ratnasingham, S, & Hebert, P D N. 2007. BOLD : the barcode of life data system (www.Barcodinglife.Org). *Molecular Ecology Notes*, **7**(3), 355–364.
- Repenning, C A. 2001. Beringian climate during intercontinental dispersal : a mouse eye view. *Quaternary Science Reviews*, **20**(1-3), 25–40.
- Rich, T D, Beardmore, C J, Blancher, P J, Bradstreet, M S W, Butcher, G S, Demarest, D W, Dunn, E H, Hunter, W C, Iñigo Elias, E E, Kennedy, J A, Martell, A M, Panjabi, A O, Pashley, D N, Rosenberg, K V, Rustay, C M, Wendt, J S, & Will, T C. 2004. *Partners in Flight North American landbird conservation plan*. Cornell Lab of Ornithology.
- Rivera, M C, & Lake, J A. 2004. The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature*, **431**(7005), 152–155.
- Rohland, N, Malaspinas, A-S, Pollack, J L, Slatkin, M, Matheus, P, & Hofreiter, M. 2007. Proboscidean mitogenomics : chronology and mode of elephant evolution using mastodon as outgroup. *PLoS Biol*, **5**(8), e207–e207.
- Rousset, S, Alves-Guerra, M C, Mozo, J, Miroux, B, Cassard-Doulicier, A M, Bouillaud, F, & Ricquier, D. 2004. The biology of mitochondrial uncoupling proteins. *Diabetes*, **53**(Feb), S130–S135.
- Rowe, K C, Heske, E J, Brown, P W, & Paige, K N. 2004. Surviving the ice : Northern refugia and postglacial colonization. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(28), 10355–10359.
- Samuels, D C. 2004. Mitochondrial dna repeats constrain the life span of mammals. *Trends in Genetics*, **20**(5), 226–229.
- Sanderson, M J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution*, **14**(12), 1218–1231.
- Sanderson, M J. 2002. Estimating absolute rates of molecular evolution and divergence times : a penalized likelihood approach. *Molecular Biology and Evolution*, **19**(1), 101–109.

- Santos, C, Montiel, R, Sierra, B, Bettencourt, C, Fernandez, E, Alvarez, L, Lima, M, Abade, A, & Aluja, M P. 2005. Understanding differences between phylogenetic and pedigree-derived mtDNA mutation rate : a model using families from the Azores Islands (Portugal). *Molecular Biology and Evolution*, **22**(6), 1490–1505.
- Sasaki, T, Nikaido, M, Hamilton, H, Goto, M, Kato, H, Kanda, N, Pastene, L A, Cao, Y, Fordyce, R E, Hasegawa, M, & Okada, N. 2005. Mitochondrial phylogenetics and evolution of mysticete whales. *Systematic Biology*, **54**(1), 77–90.
- Schwarze, S R, Lee, C M, Chung, S S, Roecker, E B, Weindruch, R, & Aiken, J M. 1995. High levels of mitochondrial DNA deletions in skeletal muscle of old rhesus monkeys. *Mechanisms of Ageing and Development*, **83**(2), 91–101.
- Selosse, M A, Albert, B R, & Godelle, B. 2001. Reducing the genome size of organelles favours gene transfer to the nucleus. *Trends in Ecology & Evolution*, **16**(3), 135–141.
- Sibley, C G, & Ahlquist, J E. 1990. *Phylogeny and classification of birds : A study in molecular evolution*. Yale University press.
- Signorovitch, A Y, Buss, L W, & Dellaporta, S L. 2007. Comparative genomics of large mitochondria in placozoans. *PLoS Genet*, **3**(1), e13–e13.
- Silva, M, Brimacombe, M, & Downing, J A. 2001. Effects of body mass, climate, geography, and census area on population density of terrestrial mammals. *Global Ecology and Biogeography*, **10**(5), 469–485.
- Slack, K E, Jones, C M, Ando, T, Harrison, G L A, Fordyce, R E, Arnason, U, & Penny, D. 2006. Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. *Molecular Biology and Evolution*, **23**(6), 1144–1155.
- Slack, K E, Delsuc, F, Mclenachan, P A, Arnason, U, & Penny, D. 2007. Resolving the root of the avian mitogenomic tree by breaking up long branches. *Molecular Phylogenetics and Evolution*, **42**(1), 1–13.
- Smith, M F, & Patton, J L. 1991. Variation in mitochondrial cytochrome b sequence in natural populations of South American akodontine rodents (Muridae : Sigmodontinae). *Molecular Biology and Evolution*, **8**(1), 85–103.
- Smith, S A, & Donoghue, M J. 2008. Rates of molecular evolution are linked to life history in flowering plants. *Science*, **322**(5898), 86–89.
- Speakman, J R. 2005. Body size, energy metabolism and lifespan. *Journal of Experimental Biology*, **208**(9), 1717–1730.
- Speakman, John R, Talbot, Darren A, Selman, Colin, Snart, Sam, McLaren, Jane S, Redman, Paula, Krol, Ela, Jackson, Diane M, Johnson, Maria S, & Brand, Martin D. 2004. Uncoupled and surviving : individual mice with high metabolism have greater mitochondrial uncoupling and live longer. *Aging Cell*, **3**(3), 87–95.

- Spielman, D, Brook, B W, & Frankham, R. 2004. Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(42), 15261–15264.
- Springer, M S, Debry, R W, Douady, C, Amrine, H M, Madsen, O, de Jong, W W, & Stanhope, M J. 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Molecular Biology and Evolution*, **18**(2), 132–143.
- Springer, M S, Murphy, W J, Eizirik, E, & O'brien, S J. 2003. Placental mammal diversification and the cretaceous-tertiary boundary. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(3), 1056–1061.
- Stanley, S E, & Harrison, R G. 1999. Cytochrome b evolution in birds and mammals : an evaluation of the avian constraint hypothesis. *Molecular Biology and Evolution*, **16**(11), 1575–1585. 222.
- Steppan, S J, Adkins, R M, & Anderson, J. 2004. Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. *Systematic Biology*, **53**(4), 533–553.
- Sun, J, Folk, D, Bradley, T J, & Tower, J. 2002. Induced overexpression of mitochondrial Mn-superoxide dismutase extends the life span of adult *Drosophila melanogaster*. *Genetics*, **161**(2), 661–672.
- Tabuce, R, Marivaux, L, Adaci, M, Bensalah, M, Hartenberger, J L, Mahboubi, M, Mebrouk, F, Tafforeau, P, & Jaeger, J J. 2007. Early tertiary mammals from north africa reinforce the molecular afrotheria clade. *Proceedings Of The Royal Society B-biological Sciences*, **274**(1614), 1159–1166.
- Tajima, F. 1989a. Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**(3), 585–595.
- Tajima, F. 1989b. The effect of change in population-size on dna polymorphism. *Genetics*, **123**(3), 597–601. Isi :a1989ax26700019 photo.
- Teeling, E C, Springer, M S, Madsen, O, Bates, P, O'brien, S J, & Murphy, W J. 2005. A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science*, **307**(5709), 580–584.
- Thomas, J A, Welch, J J, Woolfit, M, & Bromham, L. 2006. There is no universal molecular clock for invertebrates, but rate variation does not scale with body size. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(19), 7366–7371.
- Thorne, J L, & Kishino, H. 2002. Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology*, **51**(5), 689–702.
- Thorne, J L, Kishino, H, & Painter, I S. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution*, **15**(12), 1647–1657.

- Trifunovic, A, Wredenberg, A, Falkenberg, M, Spelbrink, J N, Rovio, A T, Bruder, C E, Bohlooly-Y, M, Gidlof, S, Oldfors, A, Wibom, R, Tornell, J, Jacobs, H T, & Larsson, N G. 2004. Premature ageing in mice expressing defective mitochondrial dna polymerase. *Nature*, **429**(6990), 417–423.
- Turrens, J F. 2003. Mitochondrial formation of reactive oxygen species. *Journal Of Physiology-london*, **552**(2), 335–344.
- Urbina, D, Tang, B, & Higgs, P G. 2006. The response of amino acid frequencies to directional mutation pressure in mitochondrial genome sequences is related to the physical properties of the amino acids and to the structure of the genetic code. *Journal of Molecular Evolution*, **62**(3), 340–361.
- Waddell, P J, Kishino, H, & Ota, R. 2007. Phylogenetic methodology for detecting protein interactions. *Molecular Biology and Evolution*, **24**(3), 650–659.
- Wallace, D C. 2005. A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer : a dawn for evolutionary medicine. *Annual Review of Genetics*, **39**, 359–407.
- Weir, J T, & Schluter, D. 2004. Ice sheets promote speciation in boreal birds. *Proc Biol Sci*, **271**(1551), 1881–1887.
- Weir, J T, & Schluter, D. 2007. The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science*, **315**(5818), 1574–1576.
- Weir, J T, & Schluter, D. 2008. Calibrating the avian molecular clock. *Molecular Ecology*, **17**(10), 2321–2328.
- Welch, J J, & Waxman, D. 2008. Calculating independent contrasts for the comparative study of substitution rates. *Journal of Theoretical Biology*, **251**(4), 667–678.
- Welch, J J, Bininda-Emonds, O R P, & Bromham, L. 2008. Correlates of substitution rate variation in mammalian protein-coding sequences. *BMC Evolutionary Biology*, **8**, 53–53.
- Whittle, C-A, & Johnston, M O. 2003. Broad-scale analysis contradicts the theory that generation time affects molecular evolutionary rates in plants. *Journal of Molecular Evolution*, **56**(2), 223–233.
- Wilson, D, & Reeder, D. 2005. *Mammal Species of the World*. Johns Hopkins University Press.
- Woese, C R, & Fox, G E. 1977. Phylogenetic structure of prokaryotic domain - primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America*, **74**(11), 5088–5090.
- Wolfe, K H, Li, W H, & Sharp, P M. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America*, **84**(24), 9054–9058.

- Wright, S I, Ness, R W, Foxe, J P, & Barrett, S C H. 2008. Genomic consequences of outcrossing and selfing in plants. *International Journal of Plant Sciences*, **169**(1), 105–118.
- Wu, C I, & Li, W H. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proceedings of the National Academy of Sciences of the United States of America*, **82**(6), 1741–1745.
- Xu, W, Jameson, D, Tang, B, & Higgs, P G. 2006. The relationship between the rate of molecular evolution and the rate of genome rearrangement in animal mitochondrial genomes. *Journal of Molecular Evolution*, **63**(3), 375–392.
- Yang, Z. 2007. *Computational Molecular Evolution*. Oxford University Press.
- Yang, Z, & Rannala, B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution*, **23**(1), 212–226.
- Zhang, P, Zhou, H, Liang, D, Liu, Y-F, Chen, Y-Q, & Qu, L-H. 2005. The complete mitochondrial genome of a tree frog, *Polypedates megacephalus* (Amphibia : Anura : Rhacophoridae), and a novel gene organization in living amphibians. *Gene*, **346**(Feb, 14), 133–143.
- Zukerkandl, E, & Pauling, L. 1965. *Evolving Genes and Proteins*. Academic Press Inc. 177. Chap. Evolutionary Divergence and Convergence in Proteins, pages 97–166.

Cinquième partie

Annexes

Determination of Mitochondrial Genetic Diversity in Mammals

Benoit Nabholz,^{*,1} Jean-François Mauffrey,[†] Eric Bazin,^{*,‡} Nicolas Galtier^{*} and Sylvain Glemin^{*}

^{*}CNRS UMR 5554–Institut des Sciences de l'Evolution, Université Montpellier 2, 34095 Montpellier, Cedex 5, France, [†]Laboratoire Population-Environnement-Développement, UMR D151, Université de Provence-IRD, 13331 Marseille, Cedex 3, France and

[‡]School of Biological Sciences, University of Reading, Whiteknights, Reading RG6 6BX, United Kingdom

Manuscript received March 14, 2007

Accepted for publication October 19, 2007

ABSTRACT

Mitochondrial DNA (mtDNA) is one of the most popular population genetic markers. Its relevance as an indicator of population size and history has recently been questioned by several large-scale studies in animals reporting evidence for recurrent adaptive evolution, at least in invertebrates. Here we focus on mammals, a more restricted taxonomic group for which the issue of mtDNA near neutrality is crucial. By analyzing the distribution of mtDNA diversity across species and relating it to allozyme diversity, life-history traits, and taxonomy, we show that (i) mtDNA in mammals does not reject the nearly neutral model; (ii) mtDNA diversity, however, is unrelated to any of the 14 life-history and ecological variables that we analyzed, including body mass, geographic range, and The World Conservation Union (IUCN) categorization; (iii) mtDNA diversity is highly variable between mammalian orders and families; (iv) this taxonomic effect is most likely explained by variations of mutation rate between lineages. These results are indicative of a strong stochasticity of effective population size in mammalian species. They suggest that, even in the absence of selection, mtDNA genetic diversity is essentially unpredictable, knowing species biology, and probably uncorrelated to species abundance.

MITOCHONDRIAL DNA (mtDNA) is by far the most widely used population genetic marker in animals (AVISE *et al.* 1987). The reasons typically invoked to justify the choice of mtDNA are its high level of variability, its clonal (maternal) inheritance, and its supposed nearly neutral mode of evolution. In a population at mutation/drift equilibrium, the expected level of genetic diversity of a neutral locus is proportional to the effective population size (and to the locus mutation rate WRIGHT 1931). mtDNA diversity is therefore typically assumed to reflect demographic effects, *i.e.*, variations in population size between species or populations, which makes it a popular tool for conservation purposes (*e.g.*, HARRISON 1989; ROMAN and PALUMBI 2003).

Several reports, however, have recently questioned the relevance of mtDNA as a marker of population size and history (BALLARD and WHITLOCK 2004; HURST and JIGGINS 2005; BAZIN *et al.* 2006). Among these, the BAZIN *et al.* (2006) study proposed that patterns of mtDNA diversity in animals are largely influenced by adaptive evolution. First, BAZIN *et al.* (2006) showed that the average level of within-species mtDNA diversity is not correlated to factors presumably influencing the effective population size: invertebrate species, for example, are not more polymorphic than vertebrates, marine mollusks are not more polymorphic than continental

ones, and small, planktonic crustaceans are not more polymorphic than large, benthic ones. Nuclear loci, in contrast, behave essentially as expected under the nearly neutral model. Second, BAZIN *et al.* (2006) reported a higher fixation rate of amino-acid substitutions in species with a large population size, while the opposite would be expected under the nearly neutral theory. These results were interpreted as reflecting the recurrent fixation of adaptive changes in mtDNA, consistent with GILLESPIE's (2000, 2001) genetic draft theory. It was proposed that selective sweeps associated to the fixation of favorable mutations (MAYNARD SMITH and HAIGH 1974) lead to frequent drops of mtDNA variability, removing any detectable influence of effective population size on the mitochondrial diversity (BAZIN *et al.* 2006).

By questioning the near-neutrality assumption, this and other studies cast doubts on the reliability of mtDNA as a marker of species demography: under the genetic draft model, the mitochondrial diversity of a given species is primarily determined by the time since the last selective sweep, irrespective of population size. If confirmed, this result may have important practical consequences and would call for a reappraisal of many population genetic studies based on mtDNA. Further assessment of this strong claim therefore appears worthwhile.

Several limitations of the BAZIN *et al.* (2006) study stress the need for additional investigations. First, this study was done at a relatively high taxonomic level and

¹Corresponding author: CNRS UMR 5554–Institut des Sciences de l'Evolution Université Montpellier 2–CC64, Place E. Bataillon 34095, Montpellier, Cedex 5, France. E-mail: bnabholz@univ-montp2.fr

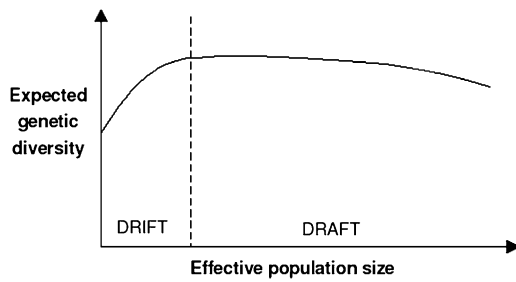


FIGURE 1.—Gillespie's genetic draft model. In small populations at mutation-drift equilibrium, the genetic diversity is proportional to the effective population size (drift zone). In large-enough populations, adaptive mutations impact the relationship by resulting in recurrent drops of diversity (draft zone). Compensation between the two forces leads to a merely flat relationship for a wide range of population sizes (redrawn after GILLESPIE 2001).

therefore involved comparing the genetic diversity of highly divergent species (*e.g.*, mammals and mollusks). Life-history traits are obviously not comparable between organisms that differ, so the link between the genetic diversity and the biology of species, if any, is difficult to make. The basic properties of the mitochondrial genome, furthermore, might vary between distantly related taxa. The BAZIN *et al.* (2006) approach, for example, did not properly account for potential variations in mitochondrial and nuclear mutation rates across animal phyla (MARTIN and PALUMBI 1993; LYNCH *et al.* 2006), although several of their results are robust to this nuisance parameter.

Second, and most importantly, it is unclear from BAZIN *et al.* (2006) whether adaptive mitochondrial evolution is effective in all animal phyla. Gillespie's model predicts a merely flat relationship between genetic diversity and effective population size (Figure 1): small populations undergo genetic drift, and large populations undergo genetic draft. The population adaptive mutation rate is obviously higher in large populations, which compensates for the decrease in genetic drift. The results of BAZIN *et al.* (2006) strongly suggest that the mtDNA diversity of invertebrate species is reduced by the recurrent fixation of advantageous mutations: invertebrates probably belong to the "draft zone" (Figure 1). But the location of vertebrate species in Gillespie's scheme is unclear. Although they do not exclude mitochondrial adaptation in vertebrates as well, the results of BAZIN *et al.* (2006) do not appear incompatible with an evolutionary process dominated by genetic drift in this group. This is an important issue because a large fraction of the mtDNA-based population and conservation genetics literature focuses on vertebrates. The question has been preliminarily addressed by MULLIGAN *et al.* (2006). Reanalyzing the Bazin *et al.* data, these authors reported a positive correlation between the average mtDNA nucleotide diversity and allozyme heterozygosity of eight mammalian orders, suggesting that mammalian mtDNA could belong to the "drift zone" (see Figure 1).

To improve our understanding of the determinism of within-species mtDNA diversity, we focus here on a more recent model taxon, namely mammals (as in MULLIGAN *et al.* 2006). Mammals were chosen because (i) they share a common body plan and physiology; (ii) life-history traits in this group are well documented, fairly variable, and comparable across species; (iii) their mitochondrial genomes share a common structure, gene order, and replication mode; (iv) their phylogeny is reasonably well known; and (v) they belong to vertebrates, for which the issue of the near-neutrality of mtDNA is still pending and crucial.

Our primary goal was to reproduce the BAZIN *et al.* (2006) approach to check whether evidence for mitochondrial adaptive evolution is detected at this lower taxonomic level. Does mtDNA diversity correlate to nuclear diversity at this scale? Is there an excess of fixation of nonsynonymous changes? Then we planned to investigate whether the mtDNA diversity of mammalian species can be predicted from life-history and ecological variables, as would be expected if the effective population size was a strong determinant of the mtDNA diversity. Are small mammals more polymorphic, on average, than large ones? Are carnivorous predator species less diverse than herbivorous prey? The relationship between life-history traits and genetic diversity in mammals was explored decades ago using allozymes (RYMAN *et al.* 1980; WOOTEN and SMITH 1985; SAGE and WOLFF 1986) and yielded equivocal results. A reanalysis using sequence data would appear timely.

MATERIALS AND METHODS

Mitochondrial data: An mtDNA data set was built from Polymorphix (BAZIN *et al.* 2005), a within-species sequence variation database. Polymorphix contains within-species homologous sequence alignments (families) built from EMBL/GenBank thanks to suitable similarity and bibliographic criteria. To obtain a homogeneous data set, a single mitochondrial gene was used, namely the cytochrome *b*. This marker represents 76% of mammalian coding sequence mitochondrial families in Polymorphix. We extracted from Polymorphix every mammalian family for which any >100-bases-long cytochrome *b* fragment was available in six individuals or more. Sequence alignments were inspected by eye and corrected when required. Dubious sequences were manually removed. This includes potential nuclear pseudogenes and sequences with many gaps or undetermined nucleotides. We obtained a data set of 277 species.

Having observed through a preliminary analysis that the level of nucleotide diversity varies between regions of the cytochrome *b* gene, we decided to analyze a common fragment in all species. For this purpose, sequences were aligned using ClustalW (THOMPSON *et al.* 1994) and widely available fragments in the database were selected. Three data sets exploring various ratios of sequence length/number of species were built:

- Data set 1: 89 species, 1140 nucleotides (complete cytochrome *b*).
- Data set 2: 138 species, 401 nucleotides (positions 1–401).
- Data set 3: 176 species, 180 nucleotides (positions 146–325).

These data sets include members of all major orders of placental mammals. Data set 2, for example, spans 32 families and 9 orders, among which four contain >10 species: Rodentia (66), Chiroptera (24), Eulipotyphla (13), and Cetartiodactyla (11). A good representation of the morphological and ecological diversity of mammals is achieved (data available at <http://kimura.univ-montp2.fr/data>). The taxonomic sampling, for example, spans five orders of magnitude for body mass. Species names were taken from Polymorphix, and higher taxonomic information is from WILSON and REEDER (2005).

Allozyme data: Estimates of allozyme heterozygosity in 184 mammalian species were obtained from the NEVO *et al.* (1984) compilation. Thirty species are represented in both the mtDNA and allozyme data set. Allozyme heterozygosities were averaged over >10 (typically 20–30) loci, and the most popular allozyme loci were used in many species. Allozyme heterozygosities are therefore fairly comparable between species.

Sampling quality control: Polymorphix families are extracted from GenBank and are therefore potentially heterogeneous. In particular, no control of the sampling strategy is made. To address this problem, we checked 60 species randomly chosen from our mtDNA data set. Among the 46 species for which we could access the corresponding publication, 37 (80%) were sampled over their entire geographic range. This proportion was unrelated to taxonomy or body mass. This suggests that these data essentially measure diversity at the species level, as intended.

Polymorphism sequence data analyses: Two measures of molecular genetic variability were used, namely the nucleotide diversity π (TAJIMA 1983) and Watterson's statistics θ_w (WATTERSON 1975). In the case of the haploid, maternally transmitted mtDNA, both statistics are unbiased estimates of the $N_{ef}\mu$ product under the assumption of neutrality and mutation/drift equilibrium, where N_{ef} is the effective population size of females and μ is the locus mutation rate. π and θ_w were calculated from the total length of the analyzed fragments and expressed in per-site level of diversity (after being divided by sequence length).

The neutrality index (NI; RAND and KANN 1996) was calculated for data set 1 when outgroups were available. This index aims at comparing the ratio of nonsynonymous (*i.e.*, amino acid changing) to synonymous (silent) changes within species (P_n/P_s) and between species (d_n/d_s): The NI is 1 when evolution is neutral, >1 under purifying selection, and <1 in case of adaptation. The significance was assessed using the McDonald–Kreitman test (MCDONALD and KREITMAN 1991). Estimates of π , θ_w , and NI were obtained using the Bio++ library (DUTHEIL *et al.* 2006).

Ecological and life-history variables: For each species in the data set, five life-history traits, eight ecological variables, and The World Conservation Union (IUCN)-threatened species categories (IUCN, 2001) were compiled from multiple published sources (see Table 1 for explanation of variables and <http://kimura.univ-montp2.fr/data> for references). The order, family, and species taxonomical levels were assigned to conform to recent taxonomical assessment (WILSON and REEDER 2005), and a fourth level, namely a supra-order, was assigned following recent molecular phylogenies (MURPHY *et al.* 2001; SPRINGER *et al.* 2004). The five life-history variables were (i) body mass, (ii) sexual maturity (age of female sexual maturity in months), (iii) reproduction frequency (time between two litters in months), (iv) maximal longevity (estimating from record longevity in years), and (v) fecundity (number of young per litter). Body mass is defined as the mean of the male and female adult body weights. All the maximal longevity values were retrieved from the AnAge database (DE MAGALHAES *et al.* 2005). We do not have any information about population

structure (*i.e.*, details on how many populations have been sampled) for all species in our data set. Therefore we used the BERLIN *et al.* (2007) data set. BERLIN *et al.* (2007) computed F_{ST} on the basis of cytochrome *b* polymorphism data for 30 genera, among which 17 genera match with our data set.

To control our life-history variables, we compared them to existing databases. The body mass data of SMITH *et al.* (2003) and the sexual maturity data of the AnAge database are highly correlated with ours (body mass: $n = 128$, $r^2 = 0.9905$, $P < 0.0001$; maturity: $n = 42$, $r^2 = 0.9081$, $P < 0.0001$). Moreover, the levels of correlation detected between our life-history variables—*e.g.*, sexual maturity *vs.* body mass ($n = 79$, $r^2 = 0.4679$, $P < 0.0001$), longevity *vs.* body mass ($n = 62$, $r^2 = 0.4771$, $P < 0.0001$), reproduction frequency *vs.* body mass ($n = 89$, $r^2 = 0.3675$, $P < 0.0001$)—are comparable with other studies (*e.g.*, MILLAR and ZAMMUTO 1983).

Statistical analyses: Genetic diversity measures were arc-sine transformed (SOKAL and ROHLF 1981) and analyzed under the general linear model assumptions using R (IHAKA and GENTLEMAN 1996). Quantitative life-history variables were log transformed. Because of the large number of factors, we could not perform a single full analysis. First, the effect of each ecological and life-history variable on genetic diversity was tested using independent regression or one-way ANOVA. Then we tested all combinations of two factors with interaction except for population structure because of the weak sample size (two-way ANOVA, multiple regression, ANCOVA).

Phylogenetic control: A significant correlation between variables across species can arise just because of their shared phylogenetic history, even if the variables evolve independently (FELSENSTEIN 1985). Phylogeny must therefore be regressed out of the analyses using the so-called comparative methods. Before doing this it must be shown that the variable of interest—in this study, genetic diversity—exhibits significant phylogenetic inertia (*e.g.*, ASHTON 2001). For this purpose, we tested for a taxonomic effect by using supra-order, order, and family as explanatory factors in one-way ANOVAs. Having detected a significant taxonomic effect, we performed a nested-hierarchical ANOVA to optimally decompose the variance across taxonomic levels. We also measured the phylogenetic autocorrelation of genetic diversity using Moran's I statistics (GITTLEMAN and KOT 1990, implemented in APE package, PARADIS *et al.* 2004).

When significant phylogenetic inertia was detected, we controlled for phylogeny in two ways. First, we used the residuals of the nested-hierarchical ANOVA for reanalyzing the effects of ecological and life-history traits. Second, we employed the phylogenetic eigenvector regression (PVR) method to correct for phylogenetic inertia (DINIZ-FILHO *et al.* 1998; DESDEVEISES *et al.* 2003). This method can be applied to both continuous and discrete variables. The method starts by performing a principal coordinate analysis on a matrix of between-species patristic distances (implemented in ADE4-R package, CHESSEL *et al.* 2004). Following DESDEVEISES *et al.* (2003), all the principal coordinates that were significantly correlated to the dependent variable were selected. Traits under analysis were then regressed on the retained eigenvectors. The estimated values express phylogenetic trends in the data, and residuals express independent evolution of each species. The proportion of genetic diversity variance explained exclusively by the ecological and life-history variables was quantified using multiple linear regressions, and its significance was assessed by using a regular parametric *F*-test (the permutation test used by DESDEVEISES *et al.* 2003 gave similar results). Finally, the contrast method (FELSENSTEIN 1985) was also used on continuous variables, yielding qualitatively equivalent results (not shown).

TABLE 1
Discrete variables and categories

Variables				
Biogeography	Environment	Geographic range		Diet
Categories				
Nearctic	Terrestrial	Limited ^a		Vegetalivorous ^b
Neotropical	Marine	Regional ^c		Grazer
Ethiopian		Extended ^d		Folivorous
Palaearctic		Worldwide		Nectarivorous
Oriental				Frugivorous
Australian				Carnivorous
Atlantic				Insectivorous
Pacific				Plankton eating
Indian				Omnivorous
Arctic				
Antarctic				
Cosmopolitan ^e				

Variables				
Habitat	Way of life	Mating system	Migrant ^f	Threatened status ^g
Categories				
Aquatic	Subterranean ^h	Monogamous	Migrant	Critically endangered
Forest	Fossorial ⁱ	Polygamous	Not migrant	Endangered
Desert	Terrestrial	Variable		Vulnerable
Other ^j	Arboreal			LR/CD ^k
	Aquatic			LR/NT ^l
	Flying			Least concern
				Data deficient
				Not evaluated

^a Endemic or/and very narrow range size.
^b Eat different vegetal elements (leaf, graze, fruit, root, bulb).
^c Part of a biogeographic province.
^d Full biogeographic province.
^e Species are present in two or more biogeographic provinces.
^f Regular seasonal movement between a reproductive area and a distinct nonreproductive area.
^g IUCN–The World Conservation Union Red List threatened species categorization system; see <http://www.redlist.org> for details.
^h Never eat on the ground.
ⁱ Burrow well, but can also eat on the ground.
^j Other habitats (savannas, meadow, bush like, or nonspecialized).
^k Lower risk/conservation dependent.
^l Lower risk/near threatened.

The phylogenetic tree was manually built according to the literature (MURPHY *et al.* 2001; HUCHON *et al.* 2002; ADKINS *et al.* 2003; HASSANIN and DOUZERY 2003; SPRINGER *et al.* 2004; STEPPAN *et al.* 2004; TEELING *et al.* 2005; tree available as supplemental material at <http://www.genetics.org/supplemental/>). One sequence per species was randomly chosen to estimate branch lengths using the maximum-likelihood method (PHYML; GUINDON and GASCUEL 2003, model TN93 + Γ).

Substitution rate analysis: The phylogenetic tree was also used to measure lineage-specific evolutionary rates. We performed a molecular clock analysis using the Multidivtime program (THORNE *et al.* 1998). To get sufficiently accurate estimates, we used only species for which the complete cytochrome *b* was available (data set 1, 89 species). The Bayesian approach in Multidivtime simultaneously estimates the rate and age for each node of a given tree, assuming a prior distribution of evolutionary rates along branches, and, option-

ally, constraints on the age of certain internal nodes. We used five fossil calibration points with upper and lower bounds: Cervidae/Bovidae (16.4–23.8 MY, HASSANIN and DOUZERY 2003), Feliformia/Caniformia (23.8–42.8 MY, BENTON and DONOGHUE 2007), Primates/rodents (61.5–95.5 MY, BENTON and DONOGHUE 2007), Pan/Homo (6.5–10.0 MY, BENTON and DONOGHUE 2007), Mus/Rattus (10.0–13.0 MY, BENTON and DONOGHUE 2007). For each species, we stored the relative rate estimated in the corresponding terminal branch of the tree.

The mitochondrial DNA of mammals is characterized by a notoriously high amount of homoplasy. This effect decreases the efficiency of mitochondrial markers for phylogenetic analyses (see SPRINGER *et al.* 2001; GALEWSKI *et al.* 2006) and presumably also affects substitution rate estimates. To limit the effect of saturation, we chose to focus on transversions only, which are much less prone to generate homoplasy than transitions.

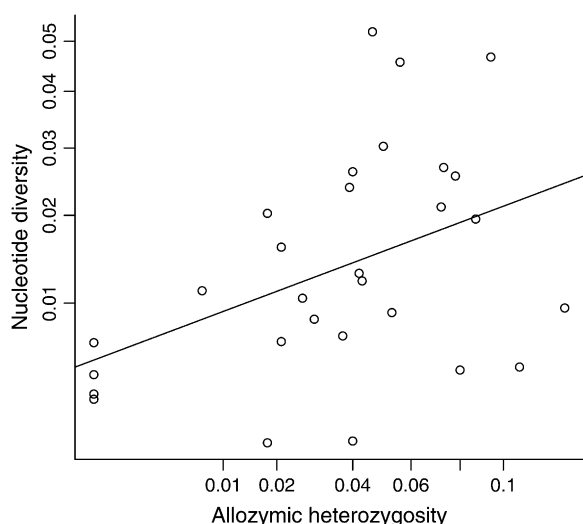


FIGURE 2.—mtDNA genetic diversity and allozyme heterozygosity (29 species). Nucleotide diversity (π) and allozyme heterozygosity are arc-sine transformed.

RESULTS

Testing the neutrality of mtDNA in mammals: A significant, positive correlation was found between the allozyme heterozygosity and mtDNA nucleotide diversity of 29 mammalian species for which both measures are available (Figure 2, $r^2 = 0.187$, P -value = 0.0191). This result contrasts with the absence of correlation between these two variables at the Metazoa level reported by BAZIN *et al.* (2006). It indicates that adaptive evolution, if any, is not frequent enough to remove the influence of demographic effects, *i.e.*, population size variations.

MULLIGAN *et al.* (2006) also reported a positive correlation between the average mtDNA and allozyme diversity of eight mammalian orders. We did not, however, detect such a significant correlation at the order level with our data set (Kendall test, $T = 3$, P -value = 0.3333). The discrepancy between the two results is due to minor differences between the data sets. MULLIGAN *et al.* (2006) used all the mammalian data points from BAZIN *et al.* (2006) (46 species), including genes other than cytochrome *b*, and various fragments of the cytochrome *b*, whereas our data set is standardized (see MATERIALS AND METHODS). If only the cytochrome *b* gene is kept in the MULLIGAN *et al.* (2006) data set, the significance of the effect is strongly reduced (Kendall test, $T = 0.73$, P -value = 0.0555; P -value = 0.0027 with all data). The correlation at the species level, finally, is only marginally significant when the whole Mulligan *et al.* data set is used ($n = 46$, $T = 0.198$, P -value = 0.0552). Although they differ in minor aspects, the two studies essentially agree with the fact that allozyme heterozygosity and mtDNA diversity are (weakly) correlated in mammals.

The mtDNA neutrality index was calculated for 76 species and varied between 0.216 and 26.7. Two-thirds of

the mtDNA data sets analyzed showed a NI value >1 , which is indicative of a predominant effect of purifying selection. NI was not correlated to any of the life-history and ecological variables analyzed here or to allozyme heterozygosity or taxonomy. No clear pattern emerged that could distinguish the 25 species showing a NI value <1 from a random sample of the data set. This contrasts again with the BAZIN *et al.* (2006) study in which a significantly lower average NI was reported in (presumably more abundant) invertebrates than (presumably less abundant) vertebrates, in agreement with the genetic draft model. Only two species, *Pteronotus parnellii* (Chiroptera, P -value = 0.007) and *Microcebus rufus* (Primates, P -value = 0.024) showed a significantly <1 NI, which is not more than expected just by chance, given the number of independent tests performed. Thirteen species, however, showed a significantly >1 NI. It should be noted that this result is robust to a potential underestimation of the synonymous divergence due to multiple hits; in the case of too distant outgroups, underestimating d_s leads to an underestimate of NI.

Overall, these results do not support a prominent role of adaptive evolution in the mitochondrial genome of mammals. Although adaptation of course may occasionally occur, overall patterns of mtDNA diversity across taxa appear compatible with a nearly neutral mode of evolution, in which demographic effects are the major determinants of mtDNA variation. Mammalian species may well belong to Gillespie's "drift zone" (see Figure 1).

Neither life-history traits nor ecology correlate with mtDNA diversity: Assuming that mtDNA diversity is governed mainly by population size variations, it is expected that the most abundant species would tend to be, on average, the genetically most diverse. Direct estimates of population size are available for a few mammalian species only and involve very large confidence intervals (NEI and GRAUR 1984). We can, however, make use of biological descriptors of species potentially correlated to population size. A total of 14 variables were used to explore as widely as possible the potential determinism of mtDNA diversity in mammals.

Analyses using nucleotide diversity (π), Watterson's statistic (θ_w), or synonymous nucleotide diversity (π_s) and any of the three mtDNA data sets gave similar results. Except when indicated, we present only the results obtained from data set 2 (401 nucleotides, 138 species) using the nucleotide diversity (other results are available on request). The mean per-site nucleotide diversity is 0.025 ± 0.022 (mean synonymous nucleotide diversity: 0.097 ± 0.090). The most variable species is the Southern pocket gopher (*Thomomys umbrinus*, Geomyidae, Rodentia, $\pi = 0.106$) and the least variable one is the gray-brown mouse lemur (*Microcebus griseorufus*, Cheirogaleidae, Primates, $\pi = 0.0004$).

We first analyzed every explanatory variable separately (Table 2). Among continuous variables, body mass,

TABLE 2
Effects of life-history and ecological variables on mtDNA diversity

	<i>N</i>	No phylogenetic control			Taxonomic control ^a		PVR ^b	
		Slope	<i>r</i> ²	<i>P</i> -value	<i>r</i> ²	<i>P</i> -value	<i>r</i> ²	<i>P</i> -value
Body mass	107	-0.0049	0.0566	<i>0.0136</i>	0.0165	0.1877	0.0273	0.0802
Sexual maturity	59	-0.0181	0.1252	<i>0.0059</i>	0.0185	0.3050	0.0018	0.6676
Reproduction	71	-0.0318	0.0334	0.1271				
Longevity	53	-0.0289	0.1260	<i>0.0091</i>	0.0650	0.0655	0.05	0.0841
Fecundity	87	0.0178	0.0080	0.4105				
Biogeography	138		0.1154	0.0504				
Diet	107		0.1947	<i>0.0005</i>	0.0533	0.4641	0.0894	0.1033
Geographic range	117		0.0446	0.1590				
Habitat	108		0.0010	0.7927				
Way of life	112		0.1058	<i>0.0346</i>	0.0616	0.2335	0.0896	0.0703
Mating system	71		0.0453	0.2064				
Migrant	94		0.0020	0.6629				
Environment	111		0.0065	0.4014				
Threatened status	138		0.0277	0.8119				

P-values less than the 0.05 threshold are in italics.

^a Residuals of nested ANOVA on taxonomy.

^b Phylogenetic eigenvector regression control.

sexual maturity, and longevity had a significant effect. In spite of the great variability of body mass in the data set (from 3.25 g for the common shrew, *Sorex cinereus*, to 666.5 kg for the yak, *Bos grunniens*), this variable explained only 6% of mtDNA polymorphism variability. This result is similar to WOOTEN and SMITH's (1985) allozyme study ($r^2 = 0.04$, P -value = 0.034). Two other continuous variables had a significant effect: sexual maturity ($r^2 = 0.1252$, P -value = 0.0059) and longevity ($r^2 = 0.1260$, P -value = 0.0091). These variables are strongly correlated to body mass ($r^2 = 0.4254$, P -value < 0.0001 for sexual maturity and $r^2 = 0.4548$, P -value < 0.0001 for longevity). The effects of body mass, sexual maturity, and longevity were removed when phylogenetic inertia was controlled for (see Table 2), indicating that the influence of life-history variables on mtDNA genetic diversity, if any, is very low.

Only two ecological variables had a significant effect, namely diet and "way of life." Diet is the ecological variable best correlated to mtDNA polymorphism ($r^2 = 0.1947$, P -value = 0.0005). The diet effect is essentially caused by the low genetic diversity of Carnivora. The "way of life" effect comes from the high genetic diversity of arboreal species, a result for which we have no obvious biological interpretation. As for life-history traits, ecological effects appear to be largely due to the phylogenetic inertia. Diet is strongly linked to taxonomy (*e.g.*, most of the carnivores belong to the Carnivora order). Not surprisingly, phylogenetic controls removed the diet effect (Table 2, P -value = 0.1033) and the "way of life" effect (Table 2, P -value = 0.070).

We detected a positive correlation between genetic diversity and F_{ST} (P -value = 0.02, Kendall no parametric test). This correlation, however, is largely explained by a

higher average F_{ST} in rodents. It was removed by a taxonomic control (P -value = 0.1263). Moreover, we did not find any correlation when we used the total BERLIN *et al.* (2007) data set (including genera missing in our data set, $N = 30$).

Then we considered every two-factor combination of the explanatory variables. We present the result only for several models (Table 3; detailed results available on request). In two-way ANOVAs, no significant effect was found, except the body mass effect in one analysis, which did not resist the phylogenetic control (see Table 2).

These results contradict many of our intuitive expectations. Figure 3 shows that the size of species geographic range does not influence mtDNA diversity: endemic species are not less polymorphic than cosmopolitan ones. Another surprising result is the absence of influence of IUCN categorization, since the system of categorization is partly linked to an evaluation of the population size. The effects of the few variables significantly influencing mtDNA diversity in one-way ANOVA are weak, disappear after the phylogenetic control (Table 2), and decrease when several variables are combined (Table 3). Overall, our results contradict the prediction that ecological and life-history traits influence the mitochondrial genetic diversity in mammals. Small, widely distributed, short-lived species are not more diverse than large, geographically restricted, threatened ones.

A strong taxonomic effect: In a surprising way, taxonomy explained a large part of the variability in cytochrome *b* diversity across species, with highly significant effects (Table 4). One-way ANOVAs indicated that most of the variation (44%) is explained by the family level. The order level explains 18% of the variability. A nested

TABLE 3
Effects of life-history and ecological variables on mtDNA diversity: two-way ANOVAs

Model	N	r ²	P-value				PVR ^a : Model
			No phylogenetic control				
			Model	Variable 1	Variable 2	Interaction	
Body mass + diet	95	0.255	<i>0.012</i>	0.223	<i>0.016</i>	<i>0.004</i>	0.1243
Body mass + way of life	102	0.209	<i>0.023</i>	<i>0.019</i>	0.090	0.282	0.1059
Body mass + range	100	0.176	<i>0.035</i>	<i>0.032</i>	0.476	0.201	0.1316
Body mass + threatened	107	0.098	0.163	<i>0.020</i>	0.604	0.055	
Threatened + range	117	0.083	0.389	0.617	0.133	0.560	

P-values less than the 0.05 threshold are in italics.
^a Phylogenetic eigenvector regression control.

ANOVA gave similar results (Table 4). Rodentia is the most polymorphic order (average $\pi = 0.033$). Carnivora is the least polymorphic order represented by more than one species in the data set (average $\pi = 0.008$, Figure 4). Supra-order average levels of diversity are also significantly different from each other, Euarchontoglires species being more diverse than Laurasiatheria. Similarly, we calculated the phylogenetic autocorrelation using Moran's I statistic (GITTLEMAN and KOT 1990) and the phylogenetic inertia using PVR (DINIZ-FILHO *et al.* 1998; DESDEVEISES *et al.* 2003). Genetic diversity had a highly significant autocorrelation ($I = -0.01943$, P -value < 0.0001) and the PVR method indicated a strong phylogenetic inertia on this variable ($r^2 = 0.319$, P -value < 0.0001 , after 10,000 random permutations). This explains why the few significant effects of ecolog-

ical and life-history traits that we detected vanished when phylogenetic inertia was controlled for.

The taxonomic effect is explained by lineage-specific variations of mutation rate: The above analyses show that neither life-history traits nor ecology influence the mitochondrial polymorphism, but reveal an unexpected link between taxonomy and genetic diversity. How to explain such a relationship? Variations of mtDNA mutation rate between mammalian lineages have been reported and could explain our results (GISSI *et al.* 2000; SPRADLING *et al.* 2001). We do not expect mutation rate to be a good predictor of species genetic diversity at mutation/drift equilibrium: between-species variations of mutation rates are probably weaker than variations of population sizes. In nonequilibrium conditions, however, mutation rates could matter. Mutation rates predominantly influence the speed of polymorphism recovery after a sudden loss of variability due to, for example, population bottlenecks (NEI and GRAUR 1984). If the stochasticity of mammalian effective population sizes was such that the mutation/drift equilibrium is rarely or never reached, then the mutation rate could be a determinant of species genetic diversity.

This hypothesis predicts that the average genetic diversity of taxonomic groups should be correlated with their current mutation rates. The mutation rate is usually unknown, but can be approached by estimating lineage-specific substitution rates; under neutrality, the two are proportional. We found that species' genetic

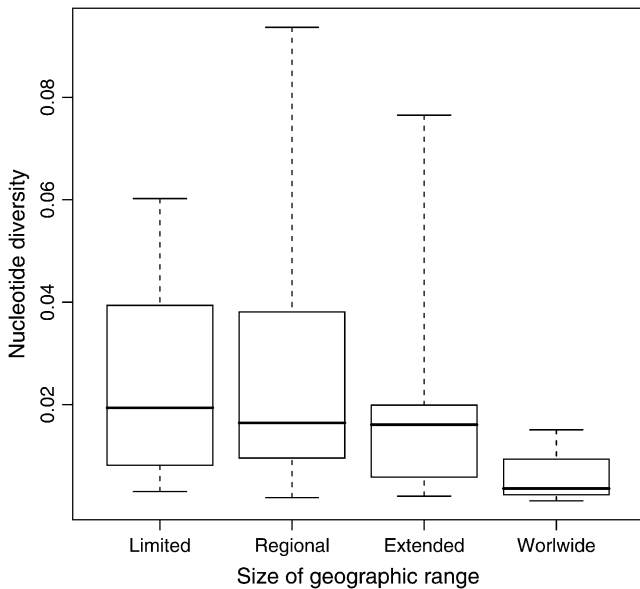


FIGURE 3.—mtDNA diversity and geographic range (117 species) and nucleotide diversity (π). The median, quartiles, and extrema of the distribution of π are given for four classes of geographic range.

TABLE 4
Taxonomic effect on mtDNA diversity

Taxonomic level	One-way ANOVA			Nested ANOVA:
	N	r ²	P-value	Variance component
Supra-order	138	0.1001	0.0009	0.050
Order	138	0.1780	0.0011	0.103
Family	138	0.4398	<0.0001	0.273

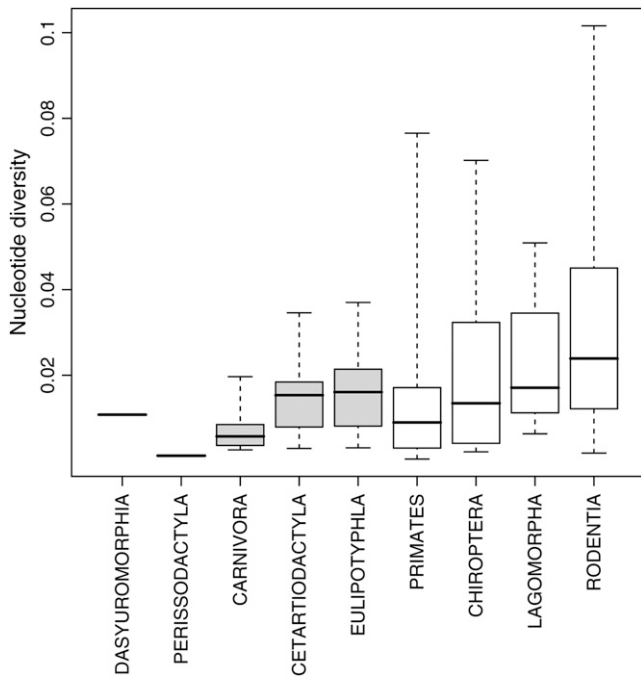


FIGURE 4.—mtDNA diversity and taxonomy (138 species) and nucleotide diversity (π). The median, quartiles, and extrema of the distribution of π are given for nine mammalian orders. Dasyuromorphia, Marsupialia. Shaded bars, Laurasiatheria; open bars, Euarchontoglires.

diversity was positively correlated to their relative transversion substitution rate as estimated with Multidivtime ($r^2 = 0.110$, $P = 0.0016$; Figure 5). This result is robust to the removal of non-neutral first and second codon positions ($r^2 = 0.044$, $P = 0.048$) and to a taxonomic control—*i.e.*, the average family rate is correlated to the average family diversity ($N = 22$, $P = 0.04$, Kendall test). Our mtDNA substitution rate estimator explains only 11% of the variance of mtDNA diversity across species, which can be explained by the high variability in genetic diversity resulting from demographic (or selective) stochasticity and/or by inaccurate estimation of evolutionary rates due to saturation and incomplete fossil information.

DISCUSSION

A detailed analysis of mtDNA diversity across a large sample of mammalian species revealed that (i) mtDNA in mammals does not reject the nearly neutral model, in contrast to a similar study conducted at the Metazoa level (BAZIN *et al.* 2006); (ii) mtDNA diversity, however, is unrelated to any of the 14 life-history and ecological variables that we analyzed, including body mass, geographic range, and IUCN categorization; (iii) mtDNA diversity is highly variable between mammalian orders and families; and (iv) this variation is most likely explained by variations of mutation rate between lineages.

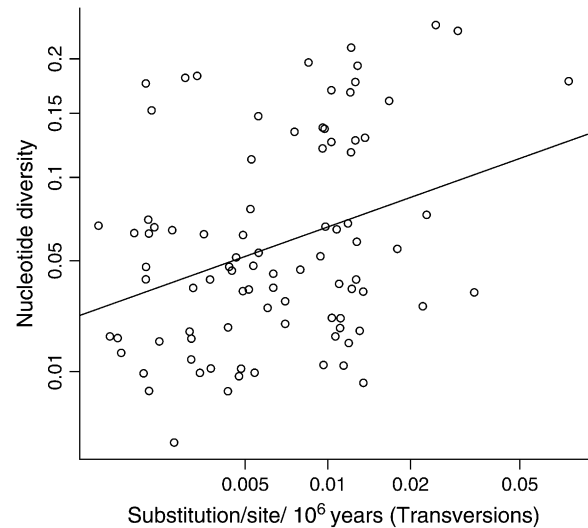


FIGURE 5.—mtDNA diversity and transversion substitution rate (89 species). Nucleotide diversity (π) is arc-sine transformed; substitution rate is log transformed.

The claim of a prominent role for adaptive evolution in animal mtDNA is therefore not supported at the Mammalia level: mtDNA diversity is correlated to allozyme diversity, and neutrality indices mostly support purifying selection. Genetic drift, not genetic draft, is probably the main determinant of species' genetic diversity in this group. This can be considered good news for users of mtDNA as a demographic marker in mammals. Although the occurrence of occasional adaptive events cannot be excluded, mammalian mtDNA appears little affected by frequent selective sweeps. The question is difficult to assess in other vertebrate taxa because their ecology and life history have not been studied as deeply as in mammals. Fish, for example, in which larger population sizes are expected and for which the mtDNA diversity apparently does not correlate to the nuclear one (BAZIN *et al.* 2006), could be worth investigating. Birds are also interesting because of their weaker average mtDNA diversity (as compared to mammals), perhaps due to the linkage between mtDNA and the W chromosome (BERLIN *et al.* 2007).

Given this result, it came as a surprise that we could not detect an effect of potential population size indicators on mtDNA genetic diversity. Of course, it is possible that the life-history and ecological variables that we used are not appropriate markers of species effective population size. But this would mean that species abundance is not correlated to body size, geographic range, IUCN categorization, or to any combination of these and other variables, which seems quite unlikely. The relationship between body mass and abundance, for example, is well supported in many groups (WHITE *et al.* 2007), including mammals (DAMUTH 1987). BAZIN *et al.* (2006) detected a significant effect of indicators of population size on allozyme heterozygosity using much

smaller data sets and arguably less sophisticated variables (marine *vs.* terrestrial mollusks, “large” *vs.* “small” crustaceans).

Similarly, the results obtained by MULLIGAN *et al.* (2006) are difficult to explain by a population size effect. We agree that the order showing the greatest mtDNA and allozyme variability (Rodentia) is one with large expected populations, whereas the least polymorphic order (Carnivora) is predicted to include species with smaller populations (higher-order predators). The ranking of orders with intermediate levels of diversity, however, is difficult to reconcile with intuition about population sizes: Artiodactyla and Cetacean species are more polymorphic, on average, than Chiroptera and Eulipotyphla. It is interesting to note, finally, that Rodentia is the fastest-evolving mammalian order and Carnivora one of the slowest evolving (Multidivtime analysis; results not shown), which is in agreement with our mutation-effect hypothesis. Finally, BERLIN *et al.* (2007) reported a significant correlation between body mass and mitochondrial diversity in birds and mammals. They did not, however, perform any phylogenetic control. When we did so using their data, the body mass effect was removed, as in this study.

The detection of a mutation rate effect but no population size effect suggests that this pattern is due to a strong stochasticity of population size in mammals. Generalizing NEI and GRAUR (1984), IIZUKA *et al.* (2002) showed that when population size randomly fluctuates between two distinct values, the mutation rate becomes the major determinant of genetic diversity (for a small-enough minimal population size). A metapopulation model with local extinction and recolonization also predicts that diversity becomes independent from effective population size when the extinction rate is higher than the migration rate (PANNELL and CHARLESWORTH 1999). The literature provides examples of population size fluctuations caused by global climatic changes, for example, during the past 3 million years (HEWITT 2000), or by pandemic diseases (HARDING *et al.* 2002). Our results suggest that stochastic variations in effective population size are probably a general process of mammalian species dynamics, strong enough to remove any effect of (current) species abundance on mtDNA diversity.

This conclusion is apparently in conflict with previous reports of ecological/life history correlates of nuclear genetic diversity. NEVO *et al.* (1984) conducted a meta-analysis of allozyme data and reported a significant effect of six variables (including range size, habitat, and longevity) on allozyme heterozygosity in mammals. These are only a few of the many ecological and life-history variables that they used, however. The percentages of explained variance were low and similar to those that we report in this study, and no phylogenetic control was performed. We reanalyzed the allozyme data controlling for taxonomy. Of the 21 variables considered by NEVO *et al.* (1984), 6 significantly influence allozyme

diversity, among which 5 are still significant after taxonomic control: geographic range size, specialist *vs.* generalist, aridity, species size, and young dispersal; only 1 effect (longevity) was removed. The allozyme results appear more robust to taxonomical control than mitochondrial ones. We see two potential explanations for this pattern. First, allozyme heterozygosity is obtained by averaging over several loci, thus reducing the variance of the coalescent process. Second, mammalian mitochondrial DNA could occasionally experience selective sweeps, although we did not detect any NI effect. Although NEVO *et al.* (1984) put the emphasis on the detected correlations, our view is that this study is largely consistent with ours in suggesting a weak, if any, influence of ecology and life history on genetic diversity in mammals.

One conflict between mtDNA and nuclear data that we consider significant is the IUCN status effect. SPIELMAN *et al.* (2004) reported that the nuclear heterozygosity of threatened taxa is, on average, lower than that of nonthreatened relatives. The effect is rather strong in mammals, in which 84% of endangered species are less diverse than their nonendangered relatives. This is true for both allozymes and microsatellite markers, and this contrasts with our mtDNA results. Even by applying the pairwise approach used by SPIELMAN *et al.* (2004), we failed to recover any effect of the IUCN status: <50% of the endangered species in our mtDNA data set are less polymorphic than their nonendangered relatives. This could be due to a lack of power: our data set includes only 22 endangered species (63 in SPIELMAN *et al.* 2004) and relies on a single locus, the mitochondrial one, while nuclear estimates of diversity are obtained by averaging over several loci, thus reducing the variance of the coalescent process. The mtDNA mutation rate effect that we detect, moreover, might add some noise and further decrease the power of the analysis.

So, even in the absence of strong adaptive effects, mtDNA diversity in mammals is largely unpredictable knowing species biology, and probably reflects in the first place the product of the mitochondrial mutation rate by the time elapsed since the last event of diversity reduction. This does not preclude the use of mtDNA to infer past demographic history, but still suggests that mtDNA diversity is a poor indicator of current population size and health.

We thank J. Dutheil and J. Claude for helpful discussions on phylogenetic control methods, S. Berlin for sharing published data, and two anonymous reviewers for helpful comments. This work was supported by the Centre National de la Recherche Scientifique, Université Montpellier II, and project MITOSYS of Agence Nationale pour la Recherche. This is publication ISE-M 2007-082.

LITERATURE CITED

ADKINS, R. M., A. H. WALTON and R. I. HONEYCUTT, 2003 Higher-level systematics of rodents and divergence time estimates based

- on two congruent nuclear genes. *Mol. Phylogenet. Evol.* **26**: 409–420.
- ASHTON, K.G., 2001 Body size variation among mainland populations of the western rattlesnake (*Crotalus viridis*). *Evolution* **55**: 2523–2533.
- AVISE, J. C., J. ARNOLD, R. M. BALL, E. BERMINGHAM, T. LAMB *et al.*, 1987 Intraspecific phylogeography: the mitochondrial-DNA bridge between population-genetics and systematics. *Annu. Rev. Ecol. Syst.* **18**: 489–522.
- BALLARD, J. W. O., and M. C. WHITLOCK, 2004 The incomplete natural history of mitochondria. *Mol. Ecol.* **13**: 729–744.
- BAZIN, E., L. DURET, S. PENEL and N. GALTIER, 2005 Polymorphix: a sequence polymorphism database. *Nucleic Acids Res.* **33**: D481–D484.
- BAZIN, E., S. GLÉMIN and N. GALTIER, 2006 Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**: 570–572.
- BENTON, M. J., and P. C. DONOGHUE, 2007 Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* **24**: 889–891.
- BERLIN, S., D. TOMARAS and B. CHARLESWORTH, 2007 Low mitochondrial variability in birds may indicate Hill-Robertson effects on the W chromosome. *Heredity* **99**: 389–396.
- CHESSSEL, D., A. B. DUFOUR and J. THIOULOUSE, 2004 The ADE4 package—I: One-table methods. *RNews* **4**(1): 5–10 (<http://cran.r-project.org/doc/Rnews/>).
- DAMUTH, J., 1987 Interspecific allometry of population-density in mammals and other animals: the independence of body-mass and population energy use. *Biol. J. Linn. Soc.* **31**: 193–246.
- DE MAGALHAES, J. P., J. COSTA and O. TOUSSAINT, 2005 HAGR: the Human Ageing Genomic Resources. *Nucleic Acids Res.* **33**: D537–D543.
- DESDEVEISES, Y., P. LEGENDRE, L. AZOUZI and S. MORAND, 2003 Quantifying phylogenetically structured environmental variation. *Evolution* **57**: 2647–2652.
- DINIZ-FILHO, J. A., C. E. R. DE SANT'ANA and L. M. BINI, 1998 An eigenvector method for estimating phylogenetic inertia. *Evolution* **52**: 1247–1262.
- DUTHEIL, J., S. GAILLARD, E. BAZIN, S. GLÉMIN, V. RANWEZ *et al.*, 2006 Bio++: a set of C++ libraries for sequence analysis, phylogenetics, molecular evolution and population genetics. *BMC Bioinformatics* **7**: 188.
- FELSENSTEIN, J., 1985 Phylogenies and the comparative method. *Am. Nat.* **125**: 1–15.
- GALEWSKI, T., M. TILAK, S. SANCHEZ, P. CHEVRET, E. PARADIS *et al.*, 2006 The evolutionary radiation of Arvicolinae rodents (voles and lemmings): relative contribution of nuclear and mitochondrial DNA phylogenies. *BMC Evol. Biol.* **6**: 80.
- GILLESPIE, J. H., 2000 Genetic drift in an infinite population: the pseudohitchhiking model. *Genetics* **155**: 909–919.
- GILLESPIE, J. H., 2001 Is the population size of a species relevant to its evolution? *Evolution* **55**: 2161–2169.
- GISSI, C., A. REYES, G. PESOLE and C. SACCONI, 2000 Lineage-specific evolutionary rate in mammalian mtDNA. *Mol. Biol. Evol.* **17**: 1022–1031.
- GITTLEMAN, J. L., and M. KOT, 1990 Adaptation statistics and a null model for estimating phylogenetic effects. *Syst. Zool.* **39**: 227–241.
- GUINDON, S., and O. GASCUEL, 2003 A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**: 696–704.
- HARDING, K. C., T. HARKONEN and H. CASWELL, 2002 The European seal plague: epidemiology and population consequences. *Ecol. Lett.* **5**: 727–732.
- HARRISON, R. G., 1989 Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends Ecol. Evol.* **4**: 6–11.
- HASSANIN, A., and E. J. P. DOUZERY, 2003 Molecular and morphological phylogenies of Ruminantia and the alternative position of the Moschidae. *Syst. Biol.* **52**: 206–228.
- HEWITT, G., 2000 The genetic legacy of the quaternary ice ages. *Nature* **405**: 907–913.
- HUCHON, D., O. MADSEN, M. SIBBALD, K. AMENT, M. J. STANHOPE *et al.*, 2002 Rodent phylogeny and a timescale for the evolution of glires: evidence from an extensive taxon sampling using three nuclear genes. *Mol. Biol. Evol.* **19**: 1053–1065.
- HURST, G. D. D., and F. M. JIGGINS, 2005 Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proc. Biol. Sci.* **272**: 1525–1534.
- HAHA, R., and R. GENTLEMAN, 1996 R: a language for data analysis and graphics. *J. Comp. Stat.* **5**: 299–314.
- IZUKA, M., H. TACHIDA and H. MATSUDA, 2002 A neutral model with fluctuating population size and its effective size. *Genetics* **161**: 381–388.
- LYNCH, M., B. KOSKELLA and S. SCHAACK, 2006 Mutation pressure and the evolution of organelle genomic architecture. *Science* **311**: 1727–1730.
- MARTIN, A. P., and S. R. PALUMBI, SR., 1993 Body size, metabolic rate, generation time, and the molecular clock. *Proc. Natl. Acad. Sci. USA* **90**: 4087–4091.
- MAYNARD SMITH, J., and J. HAIGH, 1974 The hitch-hiking effect of a favorable gene. *Genet. Res.* **23**: 23–35.
- MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652–654.
- MILLAR, J. S., and R. M. ZAMMUTO, 1983 Life histories of mammals: an analysis of life tables. *Ecology* **64**: 631–635.
- MULLIGAN, C. J., A. KITCHEN and M. M. MIYAMOTO, 2006 Comment on “population size does not influence mitochondrial genetic diversity in animals.” *Science* **314**: 1390.
- MURPHY, W. J., E. EIZIRIK, S. J. O'BRIEN, O. MADSEN, M. SCALLY *et al.*, 2001 Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* **294**: 2348–2351.
- NEI, M., and D. GRAUR, 1984 Extent of protein polymorphism and the neutral mutation theory. *Evol. Biol.* **17**: 73–118.
- NEVO, E., A. BEILES and R. BEN-SHLOMO, 1984 The evolutionary significance of genetic diversity: ecological, demographic and life-history correlates, pp. 13–295 in *Evolutionary Dynamics of Genetic Diversity* (Lecture Notes in Biomathematics, Vol. 53), edited by G. S. MANI. Springer-Verlag, Berlin.
- PANNELL, J. R., and B. CHARLESWORTH, 1999 Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution* **53**: 664–676.
- PARADIS, E., J. CLAUDE and K. STRIMMER, 2004 APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**: 289–290.
- RAND, D. M., and L. M. KANN, 1996 Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol. Biol. Evol.* **13**: 735–748.
- ROMAN, J., and S. PALUMBI, 2003 Whales before whaling in the north Atlantic. *Science* **301**: 508–510.
- RYMAN, N., C. REUTERWALL, K. NYGREN and T. NYGREN, 1980 Genetic-variation and differentiation in Scandinavian moose (*Alces alces*): Are large mammals monomorphic? *Evolution* **34**: 1037–1049.
- SAGE, R. D., and J. O. WOLFF, 1986 Pleistocene glaciations, fluctuating ranges, and low genetic-variability in a large mammal (*Ovis dalli*). *Evolution* **40**: 1092–1095.
- SMITH, F. A., S. K. LYONS, S. K. M. ERNEST, K. E. JONES, D. M. KAUFMAN *et al.*, 2003 Body mass of late quaternary mammals. *Ecology* **84**: 3403.
- SOKAL, R., and F. ROHLF, 1981 *Biometry*. W. H. Freeman, San Francisco.
- SPIELMAN, D., B. W. BROOK and R. FRANKHAM, 2004 Most species are not driven to extinction before genetic factors impact them. *Proc. Natl. Acad. Sci. USA* **101**: 15261–15264.
- SPRADLING, T., M. HAFNER and J. DEMASTES, 2001 Differences in rate of cytochrome-*b* evolution among species of rodents. *J. Mammal.* **82**: 65–80.
- SPRINGER, M., R. DEBRY, C. DOUADY, H. AMRINE, O. MADSEN *et al.*, 2001 Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Mol. Biol. Evol.* **18**: 2132–2143.
- SPRINGER, M. J., M. J. STANHOPE, O. MADSEN and W. W. DE JONG, 2004 Molecules consolidate the placental mammal tree. *Trends Ecol. Evol.* **19**: 430–438.
- STEPHAN, S. J., R. M. ADKINS and J. ANDERSON, 2004 Phylogeny and divergence-date estimates of rapid radiations in murid rodents based on multiple nuclear genes. *Syst. Biol.* **53**: 533–553.
- TAJIMA, F., 1983 Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**: 437–460.

- TEELING, E. C., M. S. SPRINGER, O. MADSEN, P. BATES, S. J. O'BRIEN *et al.*, 2005 A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* **307**: 580–584.
- THOMPSON, J. D., D. G. HIGGINS and T. J. GIBBSON, 1994 Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- THORNE, J. L., H. KISHINO and I. S. PAINTER, 1998 Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* **15**: 1647–1657.
- WATTERSON, G. A., 1975 Number of segregating sites in genetic models without recombination. *Theor. Popul. Biol.* **7**: 256–276.
- WHITE, E. P., S. K. M. ERNEST, A. J. KERKHOFF and B. J. ENQUIST, 2007 Relationships between body size and abundance in ecology. *Trends Ecol. Evol.* **22**: 323–330.
- WILSON, D., and D. REEDER, 2005 *Mammal Species of the World*. Smithsonian Institution Press, Washington, DC.
- WOOTEN, M. C., and M. H. SMITH, 1985 Large mammals are genetically less variable. *Evolution* **3**: 210–212.
- WRIGHT, S., 1931 Evolution in Mendelian populations. *Genetics* **16**: 97–159.

Communicating editor: D. CHARLESWORTH

Strong Variations of Mitochondrial Mutation Rate across Mammals—the Longevity Hypothesis

Benoit Nabholz, Sylvain Glémin, and Nicolas Galtier

Institut des Sciences de l'Evolution—CC64, Centre National de la Recherche Scientifique—Université Montpellier 2, Montpellier, France

Mitochondrial DNA (mtDNA) is the most popular marker of molecular diversity in animals, primarily because of its elevated mutation rate. After >20 years of intensive usage, the extent of mitochondrial evolutionary rate variations across species, their practical consequences on sequence analysis methods, and the ultimate reasons for mtDNA hypermutability are still largely unresolved issues. Using an extensive cytochrome *b* data set, fossil data, and taking advantage of the decoupled dynamics of synonymous and nonsynonymous substitutions, we measure the lineage-specific mitochondrial mutation rate across 1,696 mammalian species and compare it with the nuclear rate. We report an unexpected 2 orders of magnitude mitochondrial mutation rate variation between lineages: cytochrome *b* third codon positions are renewed every 1–2 Myr, in average, in the fastest evolving mammals, whereas it takes >100 Myr in slow-evolving lineages. This result has obvious implications in the fields of molecular phylogeny, molecular dating, and population genetics. Variations of mitochondrial substitution rate across species are partly explained by body mass, longevity, and age of female sexual maturity. The classical metabolic rate and generation time hypothesis, however, do not fully explain the observed patterns, especially a stronger effect of longevity in long-lived than in short-lived species. We propose that natural selection tends to decrease the mitochondrial mutation rate in long-lived species, in agreement with the mitochondrial theory of aging.

Introduction

Mitochondrial DNA (mtDNA) has been the most widely used tool for reconstructing population and species histories, presumably in the first place for practical reasons: mtDNA is relatively easy to amplify, not duplicated, typically nonrecombining, supposedly nearly neutral, and highly variable between and within species. Several potential pitfalls associated with this marker, however, have been discussed in recent years, including evidence for recombination, adaptive evolution, and mutation hot spots (Eyre-Walker et al. 1999; Ballard and Whitlock 2004; Hurst and Jiggins 2005; Bazin et al. 2006; Galtier et al. 2006). Obviously, understanding the forces that drive the molecular evolution of mtDNA is necessary for a proper use of this molecule as a population genetic, phylogeographic, and phylogenetic marker. In this paper, we investigate one particularly intriguing feature of the animal mitochondrial genome, namely its hypermutability. The mitochondrial mutation rate is orders of magnitude higher than the nuclear one (Lynch et al. 2006). This hypermutability largely explains the high level of within-species mtDNA heterozygosity (Bazin et al. 2006) and the high amount of homoplasy detected in animal mtDNA phylogenetic analyses (Springer et al. 2001; Delsuc et al. 2003; Galewski et al. 2006). mtDNA hypermutability also causes numerous diseases in human (Florentz et al. 2003; Wallace 2005). The question of why such an elevated mutation rate has evolved in animals is open and debated (Rand 1994; Palmer et al. 2000; Hellberg 2006).

The mtDNA mutation rate is high on average and, additionally, appears variable between animal lineages, as indicated by evolutionary analyses (Martin et al. 1992; Martin and Palumbi 1993; Rand 1994; Mindell et al. 1996). Variations of amino acid substitution rates have been reported—a

well-known example is the accelerated evolution of the mitochondrial proteins of anthropoid primates (Grossman et al. 2004). Protein evolution, however, is partly governed by natural selection, so that such studies cannot be interpreted directly in terms of mutation rate. Focusing on synonymous positions would appear a better approach—under the assumption of neutral evolution, the substitution rate is equal to the mutation rate (Kimura 1968). Assessing the variation of synonymous substitution rates between lineages, however, is technically problematic because of saturation: the high mtDNA mutation rate can lead to the frequent occurrence of multiple hits at the same site, rapidly dissolving the phylogenetic and evolutionary signal (Spradling et al. 2001; Springer et al. 2001; Galewski et al. 2006), especially for fast-evolving lineages.

Achieving an accurate and unbiased estimation of the neutral mitochondrial substitution rate and its variations would benefit to many fields of evolutionary biology (Mindell and Thacker 1996). In phylogenetic studies, a correct assessment of the level of saturation in mtDNA data sets would help deal with homoplasy and long-branch attraction (Spradling et al. 2001; Springer et al. 2001; Galewski et al. 2006). An accurate estimation of the mtDNA mutation rate is also obviously required for molecular dating. The famous but rather rough “2% per million year” calibration of the molecular clock, proposed 3 decades ago for primate mtDNA (Brown et al. 1979), is still sometimes used for various mammalian or even bird taxa (e.g., Garcia-Moreno 2004; Weir and Schluter 2007). It is important to appraise whether such gross approximations are acceptable or not. Knowledge about the variations of the mtDNA mutation rate, finally, would also help interpret patterns of within-species diversity across animals, a topic recently debated in the literature (Bazin et al. 2006; Lynch et al. 2006; Nabholz et al. Forthcoming).

Besides these practical aspects, analyzing the variations of mtDNA mutation rate across species might shed light onto the fundamental issue of its determinism—and perhaps onto the ultimate reasons for mtDNA hypermutability. Several theories have been proposed to explain differences in rates of molecular evolution between lineages.

Key words: mutation rate, mitochondrial DNA, theory of aging, mammals.

E-mail: bnabholz@univ-montp2.fr.

Mol. Biol. Evol. 25(1):120–130. 2008

doi:10.1093/molbev/msm248

Advance Access publication November 12, 2007

© The Author 2007. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.

For permissions, please e-mail: journals.permissions@oxfordjournals.org

The generation time hypothesis (Li et al. 1987; Ohta 1993; Mooers and Harvey 1994; Bromham et al. 1996) expresses that species with a shorter generation time experience a higher number of DNA replication rounds per year and thus a higher number of replication errors per time unit. This model applies to replication-dependent mutations only and assumes a constant replication error rate across species. It should affect the nuclear and organellar genomes with the same magnitude. It predicts an inverse relationship between generation time and neutral substitution rate.

The metabolic rate hypothesis (Martin et al. 1992; Martin and Palumbi 1993; Martin 1995; Nunn and Stanley 1998) links the mtDNA mutation rate to the by-product of mutagenic free radicals—the so-called reactive oxidative species (ROS)—during mitochondrial respiration (Wallace 2005). This hypothesis was recently put forward when Gillooly et al. (2005) proposed and tested a model explaining the relationship between mass-specific metabolic rate, which varies with body mass and temperature, and the mitochondrial substitution rate (Gillooly et al. 2005). This model applies to replication-independent mutations. It predicts a linear relationship between metabolic rate and neutral substitution rate.

Another body of literature, mainly from the biomedical field, suggests a distinct interpretation of the link between free radicals production and mutation rate, which we now formalize. In the mid 1950s, Denham Harman (1957) introduced the mitochondrial theory of aging, which states that the free radicals generated throughout the life span of an organism contribute to, and even cause, its senescence. This theory has been supported by the report of an inverse relationship between oxidative damage marker concentration and maximum life span in mammals (Barja and Herrero 2000). The mitochondrial theory of aging posits that somatic mutations occurring on mtDNA, by affecting the efficiency of respiratory proteins, result in an increase of the rate of free radical leakage, which in turn increases the mutation rate—a positive feedback loop eventually resulting in aging (Kujoth et al. 2007). It is therefore tempting to propose a selective hypothesis explaining the variations of the mtDNA mutation rate: the selective pressure for a reduced mutation rate should be higher in long-lived species than in short-lived ones. This model predicts a negative correlation between mutation rate and longevity.

This study, therefore, has 2 connected goals. First, we want to estimate as accurately as possible, and on a relatively wide taxonomical range, the lineage-specific neutral mtDNA substitution rates, taken as estimates of the mutation rate. This means overcoming the problem of saturation (i.e., frequent multiple hits) at synonymous positions in fast-evolving lineages and calibrating divergence times. Then we want to progress in our understanding of the determinism of mtDNA substitution rate variations by distinguishing between the 3 hypothesis evoked above, namely the generation time hypothesis, the metabolic rate hypothesis, and the longevity hypothesis.

We used mammals as our model taxon for 3 reasons. First, mtDNA sequence data are available in a very large number of mammalian species—more than one-third of the total number of living mammals. We chose to focus on 1 gene, cytochrome *b*, because it covers most of the spe-

cies for which mtDNA sequence data are available in GenBank and because its evolution appears representative of the whole mitochondrial genome (Waddell et al. 2007). The second advantage of mammals is that life history traits in this group are highly variable and well documented. Body mass was taken as an indicator of the metabolic rate—at constant temperature, the metabolic rate scales in quarter-power allometric relation with body mass (Gillooly et al. 2001, 2005). Generation time was estimated by the age of female sexual maturity. Life span, finally, was measured by the maximal recorded longevity. Mammals, finally, is an important and popular taxon for biodiversity and conservation studies, in which the mitochondrial marker has been extensively used. Properly characterizing the evolutionary dynamics of mtDNA in this group would appear a valuable and useful achievement.

Materials and Methods

Sequence Data

We extracted from National Center for Biotechnology Information/GenBank all the mammalian sequences for which the complete cytochrome *b* sequence was available. We selected 1 sequence per species. Accession number is available at <http://kimura.univ-montp2.fr/~benoit/data.html>. This was also done for 4 nuclear genes: interphotoreceptor retinoid-binding protein exon 1, (IRBP, 664 species), α adrenergic receptor 2B (151 species), breast cancer 1 (168 species), and von Willebrand factor (294 species). All 4 genes yielded qualitatively similar results but not always easy to compare because of differences in the number of available sequences. This is why we only provide the IRBP results.

Substitution Rate Estimation

The whole cytochrome *b* data set was split in groups within which sequence divergence is moderate. To achieve this, the GenBank taxonomic classification was traversed recursively, starting from mammalian orders and moving toward lower levels. For each traversed taxonomic group, we 1) gathered the corresponding third codon position sequences, 2) aligned sequences using ClustalW (Thompson et al. 1994), 3) built a maximum likelihood phylogenetic tree using PHYML (Guindon and Gascuel 2003), general time reversible + Gamma model of nucleotide evolution, and 4) calculated the pairwise patristic distances for every pair of species (defined as the sum of branch lengths in the path connecting the 2 species in the tree). When the median pairwise distance between species was lower than 0.4, the current taxonomic group was selected for further analysis, and the taxonomic traversal was stopped. Otherwise, the above procedure was applied to underlying taxonomic groups. The procedure was stopped at the genus level; even those genera that did not satisfy the divergence criterion were selected. This might lead to a slight underestimation of the evolutionary rate in the most fast-evolving lineages. The 0.4 threshold was obtained empirically by examining by eye the plot between uncorrected and maximum likelihood-estimated pairwise divergences—the relationship is more

Table 1
Fossil Calibration Dates Used in this Study (Myr)

Divergence	Maximal Date	Minimal Date	Reference
Monotrema/Theria	191.1	162.5	Benton and Donoghue (2007)
Didelphimorphia/Australidelphian	71.2	61.5	Benton and Donoghue (2007)
Diprotodontian/Dasyuromorphia	74	54	Drummond et al. (2006)
<i>Phascogale/Dasyurus</i>	10	24	Drummond et al. (2006)
<i>Bradypus/Dasybus</i>	Free	58	Delsuc et al. (2004)
Feliformia/Caniformia	63.8	42.8	Benton and Donoghue (2007)
<i>Equus/Rhinoceros</i>	58	54	Springer et al. (2003)
<i>Hippopotamus/Cetacea</i>	Free	52	Springer et al. (2003)
Pteropodidae/Rhinolophidae	60	43	Springer et al. (2003)
Soridae/Erinaceidae	Free	63	Springer et al. (2003)
<i>Bos/Sus</i>	65	48	Springer et al. (2003)
<i>Bos/Cervus</i>	23.8	16.4	Hassanin and Douzery (2003)
<i>Pan/Homo</i>	10	6.5	Benton and Donoghue (2007)
<i>Macaca/Homo</i>	33.9	23	Benton and Donoghue (2007)
Primates/Rodentia	100.5	61.5	Benton and Donoghue (2007)
<i>Apodemus sylvaticus/Apodemus mystacinus</i>	Free	7	Michaux et al. (2002)
<i>Gerbillus/Gerbilliscus</i>	9	7	Neumann et al. (2006)
<i>Xerus/Sciurus</i>	Free	36	Mercer and Roth (2003)
<i>Cavia/Octodon</i>	31	Free	Huchon and Douzery (2001)
<i>Mus/Rattus</i>	12.3	10	Benton and Donoghue (2007)
<i>Bolomys/Akodon</i>	5.3	3.5	Smith and Patton (1991)
<i>Neofiber/Ondatra</i>	4.3	2.2	Repenning (2001)

or less linear for corrected divergence values lower than 0.4. Sequence management and GenBank taxonomic exploration were achieved using homemade C++ programs based on the Bio++ libraries (Dutheil et al. 2006).

Within each selected group, the species-specific relative neutral substitution rate was estimated by applying MULTIDIVTIME (Thorne et al. 1998) to third codon positions (model F84). The Monte Carlo Markov Chain was run for 1 million generations after a burn-in of 200,000 generations that achieved stationarity. The rates assigned to terminal branches were taken as species-specific substitution rate estimates. This relative rate estimation was also performed using amino acid sequences for comparative purposes.

Groups were dated by applying MULTIDIVTIME to amino acid sequences (model mtREV + Gamma). The Monte Carlo Markov Chain was run for 200,000 generations after a burn-in of 200,000 generations. The 22 fossils calibration points we used are given in table 1. For computational reasons, the amino acid data set was made of 2 selected sequences per group—the whole data set was too large. The 2 representative species of each group were randomly chosen, with the constraint that their most recent common ancestor had to be the ancestral node of the considered group, thus avoiding the use of too closely related species. We manually generated the phylogenetic tree connecting these taxa according to the molecular phylogeny literature (Springer et al. 2003; Steppan et al. 2004; Teeling et al. 2005). The amino acid analysis provided 1 date estimate per group, allowing the conversion of relative lineage-specific substitution rates into absolute lineage-specific substitution rates.

We replicated the dating analysis 10 times using different pairs of representative sequences. The correlation of rate estimates between replicates was very high (R^2 between 0.84 and 0.98), indicating a good robustness to species sampling. Results were also essentially unchanged when we modified the prior distribution of the ancestral rate (mean and vari-

ance), or of the autocorrelation parameter (mean and variance), or when we used the MCMCTREE (Rannala and Yang 2007) implementation of the autocorrelated model of rate evolution. An analysis of mitochondrial whole-genome sequences in Cetacea and Carniformia, for which a good sample of complete sequences is available, yielded results very similar to the cytochrome *b* analysis.

Life History Data

Body mass, age of female sexual maturity, gestation duration, and longevity data were obtained from the AnAge database (de Magalhaes et al. 2005). Measuring the generation time is not straightforward because it depends on the age structure of species (Charlesworth 1994), about which data are lacking for most mammals. We took either female sexual maturity or the sum of female sexual maturity and gestation duration as rough approximations of the generation time. Both gave similar results. Life history data and rate estimation are available at <http://kimura.univ-montp2.fr/~benoit/data.html>.

Statistical Analyses

The relationships between life history traits and substitution rate variations were investigated using the general linear model implemented in R (R Development Core Team 2004). In order to assess the relative contribution of each variable, we performed these analyses using the species for which the 3 traits are documented ($N = 410$). Results obtained using the maximal number of species available for each life history trait were qualitatively unchanged (available at <http://kimura.univ-montp2.fr>). Substitution rates and life history traits were log transformed. We used the type II sum of squares in 2-way analysis of variance (ANOVA), as implemented in the CAR package (Fox

2002). This analysis is independent of the order in which factors are introduced; it is equivalent to testing 1 variable against all other variables in the type I ANOVA. Phylogenetic control was achieved using the phylogenetic contrast method (Felsenstein 1985) as implemented in the APE package (Paradis 2007).

Results

Estimating Mammalian mtDNA Third Codon Position Substitution Rates

The rate of evolution of a molecule is the average number of nucleotide substitutions occurring per time unit. To estimate this variable, say, for a pair of species, one should simply divide the number of changes having occurred during the divergence of the 2 species by twice the date of this divergence, typically obtained from the fossil record. In practice, we face several problems: 1) we want to consider the divergence of an arbitrary number of species, not just pairs of species; 2) the substitution rate can vary between lineages; 3) the fossil record is scarce, so that many divergence dates are unknown; and 4) estimating the number of changes having occurred in various lineages is difficult for high levels of sequence divergence, when saturation obscures the signal. Several methods have been developed to cope with these problems (Sanderson 1997; Yoder and Yang 2000). The sophisticated Bayesian methods (Thorne et al. 1998; Drummond et al. 2006; Rannala and Yang 2007) take into account the phylogeny of the analyzed species, model rate variations across lineages, incorporate constraints from the fossil record, and aim at correcting for multiple substitutions thanks to the use of Markov models of nucleotide evolution. In the case of mtDNA third codon positions, however, divergence levels can be so high that applying such methods to the whole data set would presumably lead to a severe underestimate of the fastest rates.

To address this specific problem, we used a 3-step method making use of the decoupled nonsynonymous (i.e., amino acid) and synonymous mtDNA evolutionary rates. At step 1, we partitioned the data set in monophyletic groups of sequences such that the typical divergence within a group is lower than some threshold. At step 2, we assigned a divergence date to each group. This was done by building a data set made of the amino acid sequences of 2 species per group and applying the MULTIDIVTIME program (Thorne et al. 1998) with appropriate fossil calibration. At step 3, we estimated the species-specific neutral nucleotide substitution rate. This was achieved by applying MULTIDIVTIME to third codon positions within each group, using the divergence dates inferred at step 2 as calibrations. So the relatively slow-evolving protein sequences were used to date the mammalian tree by propagating fossil information to every node of the tree, then these dates allowed us to estimate the substitution rates of fast-evolving third codon positions using recent, saturation-free divergences only.

The same strategy was applied to a nuclear marker for comparative purposes. The IRBP was chosen because it is the only nuclear gene available in a number of species comparable with the cytochrome *b* data set.

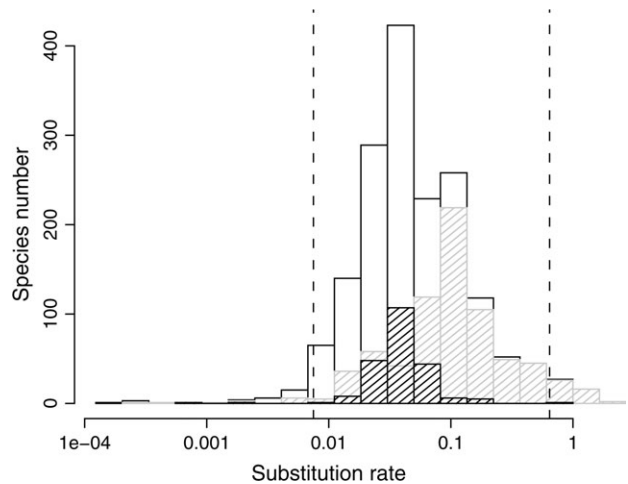


FIG. 1.—Distribution of the cytochrome *b* neutral substitution rate across 1,696 mammalian species. The substitution rate is log transformed. Black: whole data set; shading black: Chiroptera ($n = 222$); shading gray: Rodentia ($n = 734$); and vertical dotted lines: 2.5% and 97.5% percentiles. Substitution rates are in unit of substitution per third codon position per million year.

Extreme Variation of mtDNA Neutral Substitution Rate across Mammalian Species

We retrieved from GenBank the 1,938 mammalian species for which the complete cytochrome *b* sequence was available. They were partitioned in 175 groups, within which the median divergence between sequences was lower than 0.4 substitution per third codon position. These groups corresponded to 103 genera, 15 subfamilies, 43 families, 4 suborders, and 10 orders. The heterogeneity of these taxonomic levels is in itself suggestive of a large variation of substitution rate between lineages. Groups include a total of 1,696 species (species with no close relative having been removed), that is, nearly one-third of living mammals. Orders represented by more than 100 species are Rodentia (734), Cetartiodactyla (237), Chiroptera (222), Carnivora (131), Primates (113), and Eulipotyphla (104).

The mean estimated mammalian cytochrome *b* substitution rate at third codon positions was 0.098 substitution/site/Myr and the median 0.0427. The distribution of estimated mitochondrial substitution rates across species is approximately log normal and shows a very large variance (fig. 1, black histogram): the 2.5% and 97.5% percentiles are 0.007 and 0.6434 substitution/site/Myr, respectively. So even if we consider the 5% most extreme species as unreliable outliers, the estimated rates span 2 orders of magnitude, revealing an unexpectedly high level of variation of mtDNA third codon position substitution rate across mammalian lineages. To make this result more tangible, table 2 provides the detailed results for 6 pairs of closely related species for which the divergence date is known with good confidence—the first 3 are slowly evolving, the last 3 fast evolving. The Balushistan gerbil (*Gerbillus nanus*) evolves at a rate of 0.741 substitution/site/Myr (confidence interval [CI] 0.446–1.074), that is, >100 times faster than the Bryde's Whale (*Balaenoptera borealis*, 0.007 substitution/site/Myr, CI: 0.003–0.012). Cytochrome *b* third codon positions are renewed every 1–2 Myr, in average, in the

Table 2
Substitution Rate of Particularly Slow-Evolving (top) and Fast-Evolving (bottom) Species of the Data Set

Species 1	Species 2	Rate 1 ^a	Rate 2 ^a	Divergence	Reference
<i>Balaenoptera borealis</i>	<i>Balaenoptera brydei</i>	0.007 (0.003–0.012)	0.007 (0.003–0.012)	6.9 ± 2.2	Sasaki et al. (2005)
<i>Megaptera novaeangliae</i>	<i>Balaenoptera physalus</i>	0.008 (0.005–0.020)	0.008 (0.005–0.020)	15.3 ± 3.0	Sasaki et al. (2005)
<i>Homo sapiens</i>	<i>Pan Troglodytes</i>	0.011 (0.003–0.009)	0.008 (0.004–0.010)	8.25 ± 1.75	Benton and Donoghue (2007)
<i>Mus indutus</i>	<i>Rattus norvegicus</i>	0.223 (0.113–0.413)	0.098 (0.021–0.195)	11.5 ± 0.8	Benton and Donoghue (2007)
<i>Apodemus sylvaticus</i>	<i>Apodemus mystacinus</i>	0.220 (0.082–0.500)	0.258 (0.087–0.542)	7.9 ± 0.3	Michaux et al. (2002)
<i>Gerbillus nigeriae</i>	<i>Gerbillus nanus</i>	0.728 (0.087–1.170)	0.741 (0.446–1.074)	6.0 ± 0.92	Neumann et al. (2006) and
				4.12 ± 0.9	Chevret and Dobigny (2005)

NOTE.—Divergence dates were taken from the cited references.

^a Rates are in unit of substitution per third codon position per million years. Confidence intervals are given within parenthesis.

fastest evolving mammals (e.g., mice, gerbils, and voles), whereas it takes >100 Myr in slow-evolving lineages (e.g., whales, kangaroos, and hominids).

A hierarchical ANOVA shows that the order taxonomic level explains 21% of the variance of substitution rate across species. The fastest evolving order, Rodentia, averages 0.176 substitution/site/Myr, which is 20 times faster than in Cetacea (average 0.010 substitution/site/Myr). Figure 2 shows the variation between orders. The family level, with 75% of explained variance, is the taxonomical level best explaining variations of substitution rate. Rodentia, for instance, include not only very fast-evolving families (e.g., Muridae) but also slow-evolving ones (e.g., Bathyergidae).

The average mammalian cytochrome *b* neutral substitution rate is nearly 60 times higher than the nuclear one, as estimated from IRBP (mean IRBP third codon position rate:

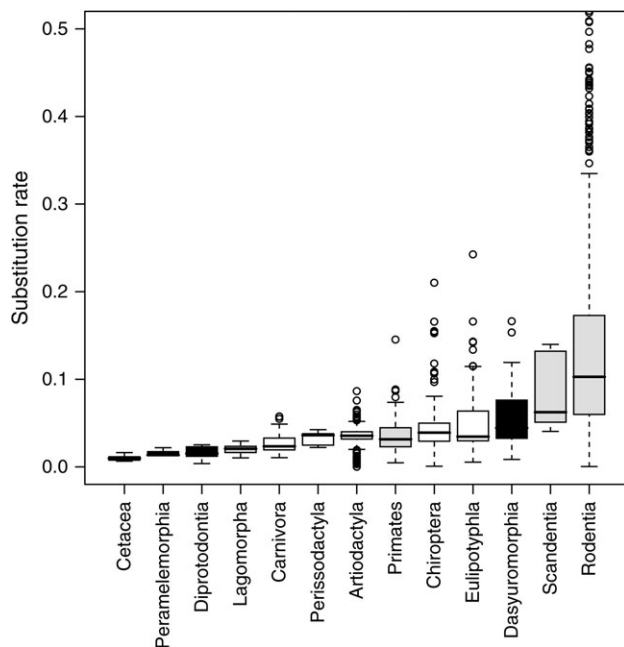


FIG. 2.—Within-order distributions of the cytochrome *b* neutral substitution rate. Orders represented by more than 5 species are shown. Light gray: Euarchontoglires; white: Laurasiatheria; and dark gray: Metatheria. Horizontal bars give the median of substitution rate distributions; boxes give the quartiles; whiskers extend to 1.5 times the interquartile range; and circles are for outliers. The Rodentia distribution includes values higher than 0.5, not shown for clarity. Substitution rates are in unit of substitution per third codon position per million year.

0.00083 substitution/site/Myr, 664 species). IRBP, in addition, does not show the extreme level of variation detected in mtDNA. Without paying any attention to the accuracy of estimates and the removal of outliers, the ratio of the fastest to the slowest estimated rate is only 9 as far as IRBP is concerned. The cytochrome *b* and IRBP rates are significantly correlated across species ($R^2 = 0.265$, P value < 0.0001, $N = 364$, not shown), but the relationship is not linear. In fast-evolving mammals, the mitochondrial substitution rate can be 100 times higher than the nuclear one (e.g., 136 in Muridae, Rodentia), whereas this ratio is lower than 20 in slow-evolving mammals (e.g., 15 in Hominidae, Primates).

Mitochondrial lineage-specific substitution rates were estimated from amino acid sequences for comparative purposes. Nonsynonymous and third codon position substitution rates were well correlated ($R^2 = 0.264$, P value < 0.0001). The median estimated rate was 0.003 amino acid substitution per site per million year. The Rodentia average was 0.088, that is, ~50 times faster than in Carnivora (0.0015). This shows that even nonsynonymous substitution rates are highly variable between mammalian lineages and that the saturation problem also applies at the amino acid level in fast-evolving groups.

Determinants of Substitution Rate Variations

Having resolved the problem of saturation, and estimated with reasonable accuracy the neutral substitution rate, we examined the possible causes of the detected variations by examining 3 life history traits: age of sexual maturity, body mass, and longevity. The cytochrome *b* substitution rate at third codon positions is strongly negatively correlated to each of these variables in log–log linear relationship (fig. 3, table 4). These relationships are in agreement with the 3 explanatory hypotheses evoked above. Female sexual maturity explains the highest fraction of substitution rate variance across species ($R^2 = 0.261$). Combining it with gestation duration—that is, using female sexual maturity + gestation duration as an estimate of generation time—did not improve significantly the relationship (results not shown). Body mass was the life history trait that explains the lowest amount of variance ($R^2 = 0.247$). We failed, moreover, to recover the slope predicted by Gillooly et al. (2005). These authors predicted that, at constant internal organism temperature, body mass and substitution rate should follow an allometric relationship with exponent equal to $-1/4$. Their analysis of 14 pairs of vertebrate

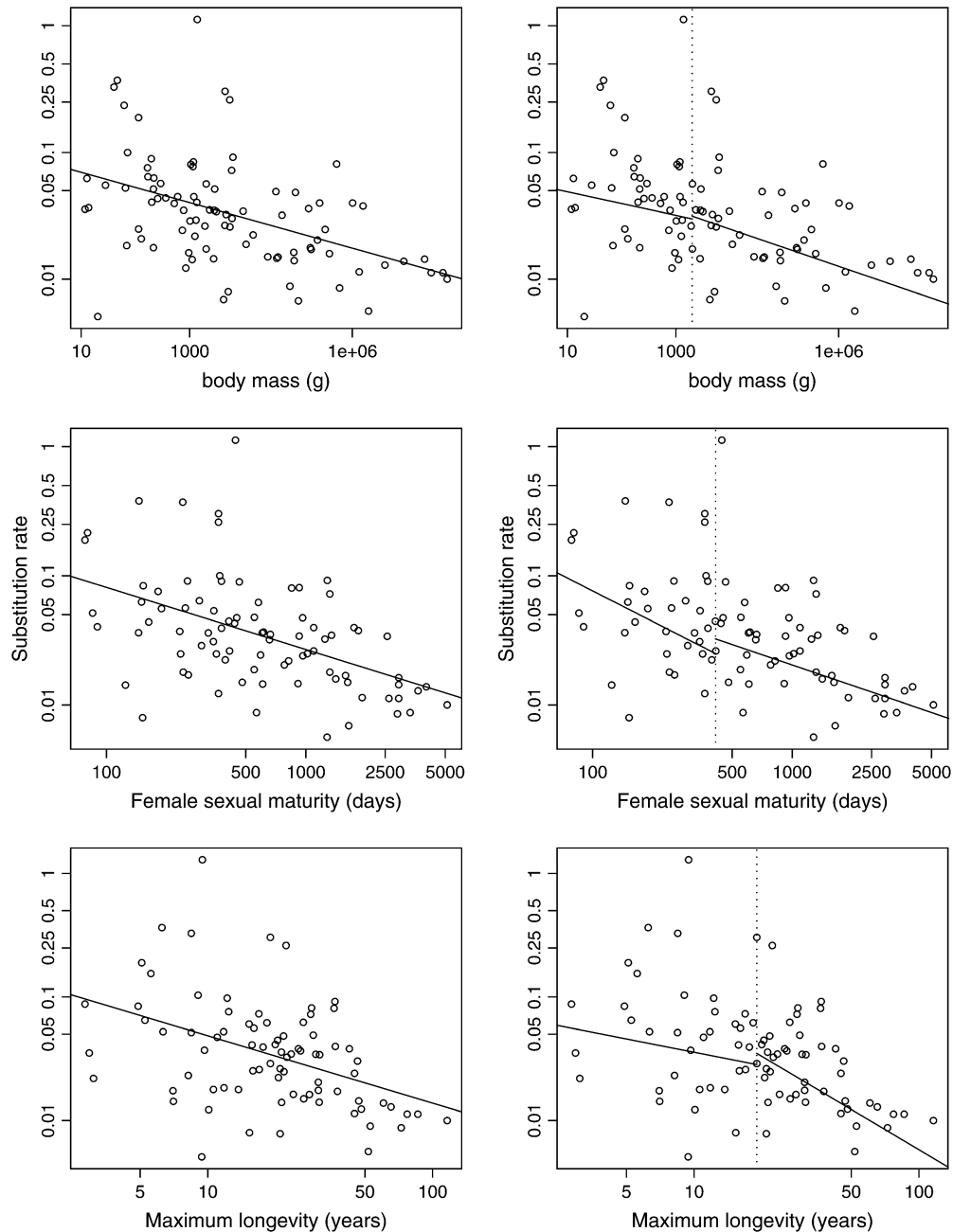


FIG. 3.—Relationship between substitution rate and life history traits. Each dot is for one family. Left panels: a single regression. Right panels: 2 independent regressions for families below versus above the median life history trait (vertical dashed lines). Substitution rates are in unit of substitution per third codon position per million year.

species was consistent with this prediction. Our 643-species data set yielded a slope of -0.132 ± 0.010 (type 1 regression) or -0.139 ± 0.010 (orthogonal regression) between log substitution rate and log body mass.

The 3 life history variables are highly positively correlated with each other, making the distinction between the relative contribution of each factor problematic (Bromham et al. 1996). We performed multiple regressions of substitution rate against every combination of 2 variables with interaction (see supplementary materials). The model best explaining substitution rate variations was $\log(\text{sexual maturity}) + \log(\text{longevity}) + \log(\text{body mass})$ ($R^2 = 0.296$,

P value < 0.0001). So despite the strong correlation between these 3 life history traits, they appear to have distinctive and cumulative effects on the determinism of substitution rate variations.

We checked whether these results were robust to the removal of phylogenetic effects. Phylogenetic inertia is suspected to be high in our data set because taxonomy explains a substantial fraction of mtDNA substitution rate variations and because the life history traits we use are known to be linked to the phylogeny. We therefore used the phylogenetic contrast method (Felsenstein 1985) to regress phylogeny out of the analyses. The effects of the 3 life history traits

on substitution rates were still significant after phylogenetic control (table 4). Their relative strengths, however, were different: the sexual maturity effect was decreased and even disappeared when combined with longevity in a multiple regression. Body mass was the variable explaining the largest percentage of substitution rate variance after phylogenetic control.

The longevity-related model we formalized (see Introduction) invokes strong selective pressure for reduced mtDNA mutation rate in long-lived mammals, for which too high a mutation rate could result in premature senescence, but relaxed selection in short-lived mammals, for which a larger range of mutation rates would be acceptable. To test this prediction we split the data set in 2 subsets of equal size according to species longevity (median: 19.8 years) and performed simple and multiple regressions separately for the 2 subsets (table 5). In agreement with our prediction, the effect of longevity was much stronger for long-lived ($R^2 = 0.198$) than short-lived ($R^2 = 0.098$) mammals, whereas the effects of body mass and sexual maturity were comparable on both sides of the longevity threshold (table 5). Figure 3 also shows a substantial shift in the slope of the substitution rate/longevity relationship between short-lived and long-lived species, whereas this effect is not detected for sexual maturity and body mass. Highly similar results were obtained when we split the data according to body mass or sexual maturity medians as threshold (data not shown). Interestingly, multiple regressions revealed that the major determinant of substitution rate variations is female sexual maturity in short-lived species but longevity in long-lived species (table 5).

The shift in slope and correlation coefficient for longevity between short-lived and long-lived mammals was robust to the phylogenetic control, to the removal of species with a longevity higher than 50 years, and to the removal of the particularly slow-evolving Cetaceans (results not shown). This result is consistent with the hypothesis of an optimized mitochondrial mutation rate in long-lived mammals—and uneasy to explain otherwise. The longevity-related selective hypothesis is congruent with the significant decrease of between-species variance in substitution rate with longevity observed in figure 3—the long-lived species would be under a stronger selective pressure according to this model. This effect is also revealed by introducing the squared longevity as an additional explanatory factor. Squared longevity explained the highest fraction of variance of substitution rate across species ($R^2 = 0.266$) when analyzed solely and withdrew the effect of longevity in a 2-way regression (data not shown), indicating that the effect of longevity is not linear. Finally, the effect of longevity on substitution rates was not higher, and was even lower, in long-lived ($R^2 = 0.0003$) than in short-lived ($R^2 = 0.1057$) species as far as the nuclear IRBP gene was concerned. So this result appears specific to mtDNA.

Relative Part of Transitions and Transversions in Substitution Rate Variation

Mammalian mtDNA undergoes much more transitions (A ↔ G, C ↔ T) than transversions. We investigated whether the variations in substitution rate we report are

due only to changes of the transition rate or to both transitions and transversions. We estimated the third codon position transversion rate variation by recoding the data set as purines and pyrimidines before applying the methodology described above. The average rate was 0.0059 transversion/site/Myr; transversions represent only 6% of all the substitutions. The transversion rate was correlated to the transition rate ($N = 1595$, $R^2 = 0.692$, P value < 0.0001). The variation of transition and transversion rates appears comparable. The house mouse *Mus musculus*, for instance, evolves ~30 times faster than the blue whale *Balaenoptera musculus* as far as mitochondrial transversions are concerned. Consequently, significant correlations between the transversion rate and life history traits were detected (sexual maturity, $R^2 = 0.280$; body mass, $R^2 = 0.136$; longevity, $R^2 = 0.262$).

Discussion

Reliability of Divergence Date Estimates

The estimation of substitution rates heavily relies on the accuracy of date estimates. We compared our estimated divergence dates with previous molecular studies which combined the information from several genes, including nuclear ones (Hasegawa et al. 2003; Springer et al. 2003; Stepan et al. 2004; Drummond et al. 2006). Our cytochrome *b* date estimates are largely congruent with published ones (table 3), especially for recent divergences (e.g., *Homo*—*Hylobates*, Hominidae—Cercopithecidae). Some older divergence dates appear overestimated by the cytochrome *b* analysis, however, like the primary divergence of Chiroptera or Eulipotyphla (table 3). It should be noted that the oldest divergences are not used to estimate substitution rates because third codon positions are only analyzed within groups of relatively recent origin. Standard deviations around the estimated dates are significantly larger in this study than in multigene analyses, presumably because a single gene is used. We incorporated as many fossil calibration points as possible (table 1) and by this way obtained reasonable estimates for recent dates.

A Highly Variable mtDNA Substitution Rate

According to our analysis, the mtDNA neutral substitution rate varies by 2 orders of magnitude across mammalian species. Previous studies had noticed that the mtDNA rate varies across lineages, but none detected the extreme levels of variations we report, for various reasons: smaller data set (Martin et al. 1992; Martin and Palumbi 1993), lack of a fossil information (Mindell et al. 1996; Bromham et al. 1996), or inappropriate treatment of saturation, even at the amino acid level (Gissi et al. 2000; Bininda-Emonds 2007). Strong substitution rate variations were revealed in this study because the estimation of dates and rates was obtained from essentially independent sources, namely amino acids and third codon positions, respectively. When we performed a single MULTIDIVTIME run on third codon positions using the whole data set and the same fossil calibration points, the ratio between the most extreme estimated substitution rates was only 10, that is, severely underestimated (data not shown).

Table 3
Comparison of Cytochrome *b* and Multilocus Divergence Date Estimates (Myr)

Divergence	Cytochrome <i>b</i>	Multilocus	Reference
Hominidae—Cercopithecidae	27.9 (23.2–33.5)	34.6 (33–36.2)	Springer et al. (2003)
<i>Homo</i> — <i>Hylobates</i>	16.1 (10.4–22.8)	21.7 (20.7–22.7)	Springer et al. (2003)
<i>Homo</i> — <i>Pan</i>	8.2 (6.6–9.9)	7.4 (6.7–8.1)	Springer et al. (2003)
<i>Mus</i> — <i>Rattus</i>	11.7 (10.4–12.3)	16.2 (15.8–16.6)	Springer et al. (2003)
Eulipotyphla ancestor	78.0 (63.9–101.6)	61 (57.9–64.1)	Springer et al. (2003)
Odontoceti—Mysticeti	34.4 (8.0–53.5)	25.5 (23.5–27.5)	Springer et al. (2003)
<i>Ceratotherium</i> — <i>Rhinoceros</i>	24.36 (12.6–39.5)	26.1 (23.8–28.4)	Springer et al. (2003)
Chiroptera ancestor	67.2 (52.6–83.5)	65.2 (62.3–68.1)	Springer et al. (2003)
Feliformia—Caniformia	53.0 (43.4–63.0)	49 (45.3–51.7)	Springer et al. (2003)
<i>Ochotona</i> — <i>Lepus</i>	45.5 (36.1–57.2)	51 (48–54)	Springer et al. (2003)
<i>Phascogale</i> — <i>Dasyurus</i>	20.2 (13.3–23.8)	14.4 (9.7–18.9)	Drummond et al. (2006)
<i>Dasyuridae</i> — <i>Diprotodontia</i>	59.0 (54.2–66.3)	64.2 (52.2–70.9)	Drummond et al. (2006)

This result has important practical implications in population genetics, molecular evolution, phylogeny, and molecular dating. Cytochrome *b* third codon positions should not be used for phylogenetic analyses in rodents, for instance, even at the family or subfamily level, whereas they can safely be trusted in Cetaceans. Molecular dating studies relying on mtDNA require a cautious calibration step. In the absence of fossil data, one should refrain from estimating divergence dates based on “generic” estimates such as 2% per million year. This can easily lead to an underestimation or an overestimation of divergence dates by a factor of 10. This and other studies, finally, highlight the necessity to account for variations of mutation rate when interpreting within-species diversity patterns. For instance, we would expect the average rodent species to be 10 times more polymorphic than the average primate species irrespective of population size, population structure, and selective effects.

It should be recalled, finally, that the substitution rate we estimated is an average over all cytochrome *b* third codon positions. Mutation hot spots have been detected in mammalian mitochondrial coding sequences, and the position of these hot spots appear highly variable in time and between species (Galtier et al. 2006). The dynamics of mtDNA mutation is only partly described by the lineage-specific averages we provide. Full-genome analyses of the between-sites and between-lineages variations would appear necessary for a better understanding of the molecular mechanisms underlying the frequent changes of mtDNA mutation rate we detect.

Understanding the Determinants

Having properly estimated the species-specific mtDNA substitution rate at third codon positions and detected an unexpectedly high level of variation across lineages, we examined possible determinants of this variation.

Table 4
Effects of Life History Variables on mtDNA Substitution Rate

	Without Phylogenetic Control				With Phylogenetic Control			
	Slope	<i>F</i>	<i>R</i> ²	<i>P</i> Value	Slope	<i>F</i>	<i>R</i> ²	<i>P</i> Value
Body mass	−0.127	134.1	0.2474	<0.0001	−0.137	81.85	0.1674	<0.0001
Maturity	−0.441	144	0.2609	<0.0001	−0.342	59.82	0.1281	<0.0001
Longevity	−0.610	142.7	0.2592	<0.0001	−0.527	88.42	0.1322	<0.0001

P values under the 0.05 threshold are bold-faced.

These changes being presumably neutral, we considered models relevant to the mtDNA mutation process, namely the metabolic rate hypothesis, the generation time hypothesis, and the longevity-related selected mutation rate hypothesis. These hypotheses are not mutually exclusive. They are, furthermore, uneasy to distinguish formally because the relevant life history traits—body mass, age of sexual maturity, and longevity—are strongly correlated across mammalian species.

It should be noted that life history traits have been measured on currently living specimen, whereas our evolutionary estimates of the mutation rate represent an average across the length of terminal branches of the tree—this is why using many species, and therefore reducing terminal branch lengths, is so important. It needs some time, furthermore, for a change of mutation rate to impact sequence divergence. Very recent changes of mutation rate have not yet substantially affected branch lengths and are not detectable by a phylogenetic approach. Consequently, when we restricted our correlation analyses to specific taxonomic groups, results were hardly ever significant at the genus or subfamily level, and significance increased when families, suborders or orders were analyzed. All these caveats mean that the absolute levels of correlation we report between life history traits and substitution rate probably underestimate the real effects.

Overall, our statistical analyses do not allow to refute either of the 3 explanatory hypothesis. Concerning the metabolic rate effect, we failed to recover the −0.25 power relationship reported by Gillooly et al. (2005) from a 14-species pairs analysis and predicted by their model. The implication of ROS in mtDNA point mutations in somatic cells is unquestioned (Richter et al. 1988), but only germ line mutations are relevant from an evolutionary point of view. It is unclear to us whether ROS production in female germ cells is proportional to the whole-body metabolic

Table 5
Effects of Life History Variables on mtDNA Substitution in Short-Lived (less than 19 years) versus Long-Lived (more than 19 years) Species

Variable 1	Variable 2	Short-lived				Long-lived			
		Slope	R ²	P Value 1	P Value 2	Slope	R ²	P Value 1	P Value 2
Maturity		-0.398	0.111	<0.0001		-0.435	0.168	<0.0001	
Body mass		-0.150	0.151	<0.0001		-0.080	0.104	<0.0001	
Longevity		-0.530	0.098	<0.0001		-0.812	0.198	<0.0001	
Maturity	Longevity		0.1281	0.0088	0.0495		0.213	0.0497	0.0008
Maturity	Body mass			0.0217	0.0001			<0.0001	0.054
Body mass	Longevity			0.2504	0.0002			0.1346	<0.0001

P values under the 0.05 threshold are bold-faced.

rate—experiments would be helpful here (see Lanfear et al. 2007). The fact that transition mutations are primarily involved in substitution rate variations could appear in agreement with the metabolic rate hypothesis—most mutations occurring through oxidative DNA damage are transitions (Wang et al. 1998). Replication-dependent mutations due to errors of the mtDNA polymerase, however, are mostly transitions as well (Kujoth et al. 2007), so the argument is weak.

Generation time is the strongest predictor of mtDNA substitution rate variations when variables are considered separately. One strong prediction of the generation time hypothesis, however, is that the effect should apply similarly to the nuclear and mitochondrial genomes. This is obviously not the case: the mtDNA (cytochrome *b*) mutation rate appears much more variable than the nuclear (IRBP) one, apparently rejecting the generation time hypothesis. One doubt remains, however, because of the distinct transmission modes of mitochondrial and nuclear DNA. If nuclear germ line mutations primarily occurred in males (see Li et al. 2002) and if the number of germ-line replications per generation was less variable across species in males than in females, then distinct mtDNA and nuclear patterns could be expected. Embryologic data about the male and female germ line pathways in various mammals would be required to clarify this point. Another argument speaking against generation time as the unique factor controlling substitution rate is the existence of a strong variation of base composition between mammalian lineages, especially the fastest evolving ones (Gibson et al. 2005, Urbina et al. 2006). Obviously, the synonymous substitution process is affected by other causes than just a per year/per generation scaling.

Our analysis, finally, brings support to the hypothesis of a selected decrease of substitution rate in long-lived species, in agreement with the evidence for a causal role of mtDNA mutations in aging (for a recent review see Kujoth et al. 2007). The mtDNA substitution rate is less variable, and better correlated to longevity, in long-lived than in short-lived mammalian species. Neither the metabolic rate nor the generation time hypothesis can explain these results. An independent argument for this model was provided by Samuels (2004) and Khaidakov et al. (2006). These authors showed that the total number of direct repeats in the mammalian mitochondrial genome was negatively correlated to species longevity. Direct repeats are known to contribute to somatic mtDNA damage, so this result was interpreted as reflecting a selective pressure for a decreased mtDNA so-

matic mutation rate in long-lived species. A group of particular interest for addressing the link between mutation rate and lifetime is Chiroptera (bats). Chiroptera live longer than nonflying eutherian mammals of comparable size (Brunet-Rossini and Austad 2004). The average body mass of Chiroptera species in our data set is 6.6 times lower than in Rodentia, but their average longevity is 2.4 times higher. It is interesting to note that Chiroptera evolve more than twice as slow as Rodentia, on average (fig. 1). Increasing the life history data set for Chiroptera would appear appropriate to progress with this comparison. Please note, finally, that this analysis, as most analyses of longevity in animals, relies on maximal life spans observed in captivity. Estimates from wild animals would be preferable, but such data are available in too few species.

The hypothesis of a selective constraint for a decreased mtDNA mutation rate in long-lived species is attractive, consistent with current knowledge about mitochondrial physiology and aging, and supported by our analysis. Additional arguments in favor of this model would be welcome. One natural perspective would be the analysis of the genetic systems controlling mtDNA mutation, that is, genes involved in mtDNA replication or oxidative stress reduction. According to our model, such genes should be under stronger selective constraint in long-lived than in short-lived mammals.

Acknowledgments

This work was supported by the Agence Nationale de la Recherche (project MITOSYS) and the Centre National de la Recherche Scientifique. We are grateful to Julien Dutheil for his help with C++ programming. We thank Paul Higgs and one anonymous reviewer for helpful comments. This is publication 2007-128 of Institut des Sciences de l'Evolution de Montpellier.

Literature Cited

- Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Mol Ecol.* 13:729–744.
- Barja G, Herrero A. 2000. Oxidative damage to mitochondrial DNA is inversely related to maximum life span in the heart and brain of mammals. *FASEB J.* 14:312–318.
- Bazin E, Glemin S, Galtier N. 2006. Population size does not influence mitochondrial genetic diversity in animals. *Science.* 312:570–572.
- Benton MJ, Donoghue P. 2007. Paleontological evidence to date the tree of life. *Mol Biol Evol.* 24:26–53.

- Bininda-Emonds ORP. 2007. Fast genes and slow clades: comparative rates of molecular evolution in mammals. *Evol Bioinformatics*. 3:59–85.
- Bromham L, Rambaut A, Harvey PH. 1996. Determinants of rate variation in mammalian DNA sequence evolution. *J Mol Evol*. 43:610–621.
- Brown WM, George M, Wilson A. 1979. Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci USA*. 76:1967–1971.
- Brunet-Rossini AK, Austad S. 2004. Ageing studies on bats: a review. *Biogerontology*. 5:211–222.
- Charlesworth B. 1994. Evolution in age-structured populations. Cambridge (UK): Cambridge University Press.
- Chevret P, Dobigny G. 2005. Systematics and evolution of the subfamily Gerbillinae (Mammalia, Rodentia, Muridae). *Mol Phylogenet Evol*. 35:674–688.
- Delsuc F, Stanhope M, Douzery E. 2003. Molecular systematics of armadillos (Xenarthra, Dasypodidae): contribution of maximum likelihood and Bayesian analyses of mitochondrial and nuclear genes. *Mol Phylogenet Evol*. 28:261–275.
- Delsuc F, Vizcaíno S, Douzery E. 2004. Influence of tertiary paleoenvironmental changes on the diversification of South American mammals: a relaxed molecular clock study within xenarthrans. *BMC Evol Biol*. 4:11.
- de Magalhães JP, Costa J, Toussaint O. 2005. HAGR: the human ageing genomic resources. *Nucleic Acids Res*. 33:D537–D543.
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 4:e88.
- Dutheil J, Gaillard S, Bazin E, Glemin S, Ranwez V, Galtier N, Belkhir K. 2006. Bio++: a set of C++ libraries for sequence analysis, phylogenetics, molecular evolution and population genetics. *BMC Bioinformatics*. 7:188.
- Eyre-Walker A, Smith N, Smith J. 1999. How clonal are human mitochondria? *Proc Biol Sci*. 266:477–483.
- Felsenstein J. 1985. Phylogenies and the comparative method. *Am Nat*. 125:1–15.
- Florentz C, Sohm B, Tryoen-Toth P, Putz J, Sissler M. 2003. Human mitochondrial tRNAs in health and disease. *Cell Mol Life Sci*. 60:1356–1375.
- Fox J. 2002. An R and S-PLUS companion to applied regression. London: Sage Publications.
- Galewski T, Tilak M, Sanchez S, Chevret P, Paradis E, Douzery E. 2006. The evolutionary radiation of arvicolinae rodents (voles and lemmings): relative contribution of nuclear and mitochondrial DNA phylogenies. *BMC Evol Biol*. 6:80.
- Galtier N, Enard D, Radondy Y, Bazin E, Belkhir K. 2006. Mutation hot spots in mammalian mitochondrial DNA. *Genome Res*. 16:215–222.
- García-Moreno J. 2004. Is there a universal mtDNA clock for birds? *J Avian Biol*. 35:465–468.
- Gibson A, Gowri-Shankar V, Higgs PG, Rattray M. 2005. A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic methods. *Mol Biol Evol*. 22:251–264.
- Gillooly JF, Allen AP, West GB, Brown JH. 2005. The rate of DNA evolution: effects of body size and temperature on the molecular clock. *Proc Natl Acad Sci USA*. 102:140–145.
- Gillooly JF, Brown J, West G, Savage V, Charnov E. 2001. Effects of size and temperature on metabolic rate. *Science*. 293:2248–2251.
- Gissi C, Reyes A, Pesole G, Saccone C. 2000. Lineage-specific evolutionary rate in mammalian mtDNA. *Mol Biol Evol*. 17:1022–1031.
- Grossman LI, Wildman D, Schmidt T, Goodman M. 2004. Accelerated evolution of the electron transport chain in anthropoid primates. *Trends Genet*. 20:578–585.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 52:696–704.
- Harman D. 1957. Aging: a theory based on free radical and radiation chemistry. *J Gerontol*. 2:298–300.
- Hasegawa M, Thorne J, Kishino H. 2003. Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. *Genes Genet Syst*. 78:267–283.
- Hassanin A, Douzery EJP. 2003. Molecular and morphological phylogenies of ruminantia and the alternative position of the moschidae. *Syst Biol*. 52:206–228.
- Hellberg ME. 2006. No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. *BMC Evol Biol*. 6:24.
- Huchon D, Douzery EJP. 2001. From the old world to the new world: a molecular chronicle of the phylogeny and biogeography of hystricognath rodents. *Mol Phylogenet Evol*. 20:238–251.
- Hurst GDD, Jiggins FM. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proc R Soc Biol Sci Ser B*. 272:1525–1534.
- Khaidakov M, Siegel E, Reis R. 2006. Direct repeats in mitochondrial DNA and mammalian lifespan. *Mech Ageing Dev*. 127:808–812.
- Kimura M. 1968. Evolutionary rate at molecular level. *Nature*. 217:624–626.
- Kujoth GC, Bradshaw P, Haroon S, Prolla T. 2007. The role of mitochondrial DNA mutations in mammalian aging. *PLoS Genet*. 3:161–173.
- Lanfear R, Thomas J, Welch J, Brey T, Bromham L. 2007. Metabolic rate does not calibrate the molecular clock. *Proc Natl Acad Sci USA*. 104:15388–15393.
- Li WH, Tanimura M, Sharp P. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *J Mol Evol*. 25:330–342.
- Li WH, Yi S, Makova K. 2002. Male-driven evolution. *Curr Opin Genet Dev*. 12:650–656.
- Lynch M, Koskella B, Schaack S. 2006. Mutation pressure and the evolution of organelle genomic architecture. *Science*. 311:1727–1730.
- Martin AP. 1995. Metabolic rate and directional nucleotide substitution in animal mitochondrial DNA. *Mol Biol Evol*. 12:1124–1131.
- Martin AP, Naylor G, Palumbi S. 1992. Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature*. 357:153–155.
- Martin AP, Palumbi S. 1993. Body size, metabolic-rate, generation time, and the molecular clock. *Proc Natl Acad Sci USA*. 90:4087–4091.
- Mercer JM, Roth VL. 2003. The effects of cenozoic global change on squirrel phylogeny. *Science*. 299:1568–1572.
- Michaux JR, Chevret P, Filippucci MG, Macholan M. 2002. Phylogeny of the genus *Apodemus* with a special emphasis on the subgenus *Sylvaemus* using the nuclear IRBP gene and two mitochondrial markers: cytochrome b and 12S rRNA. *Mol Phylogenet Evol*. 23:123–136.
- Mindell DP, Knight A, Baer C, Huddleston C. 1996. Slow rates of molecular evolution in birds and the metabolic rate and body temperature hypotheses. *Mol Biol Evol*. 13:422–426.
- Mindell DP, Thacker C. 1996. Rates of molecular evolution: phylogenetic issues and applications. *Annu Rev Ecol Syst*. 27:279–303.
- Mooers AO, Harvey P. 1994. Metabolic rate, generation time, and the rate of molecular evolution in birds. *Mol Phylogenet Evol*. 3:344–350.

- Nabholz B, Mauffrey JF, Bazin E, Galtier N, Glémin S. Forthcoming. Determination of mitochondrial genetic diversity in mammals. *Genetics*.
- Neumann K, Michaux J, Lebedev V, et al. (12 co-authors). 2006. Molecular phylogeny of the Cricetinae subfamily based on the mitochondrial cytochrome b and 12S rRNA genes and the nuclear vWF gene. *Mol Phylogenet Evol.* 39:135–148.
- Nunn GB, Stanley S. 1998. Body size effects and rates of cytochrome b evolution in tube-nosed seabirds. *Mol Biol Evol.* 15:1360–1371.
- Ohta T. 1993. An examination of the generation-time effect on molecular evolution. *Proc Natl Acad Sci USA.* 90:10676–10680.
- Palmer JD, Adams K, Cho Y, Parkinson C, Qiu Y, Song K. 2000. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proc Natl Acad Sci USA.* 97:6960–6966.
- Paradis E. 2007. *Analysis of phylogenetics and evolution with R*. New York: Springer.
- Rand DM. 1994. Thermal habit, metabolic-rate and the evolution of mitochondrial-DNA. *Trends Ecol Evol.* 9:125–131.
- Rannala B, Yang Z. 2007. Inferring speciation times under an episodic molecular clock. *Syst Biol.* 56:453–466.
- R Development Core Team. 2004. *R: a language and environment for statistical computing*. [Internet]. Vienna (Austria): R Foundation for Statistical Computing. [cited 2006 Dec 18]. Available from: <http://www.R-project.org>.
- Repenning CA. 2001. Beringian climate during intercontinental dispersal: a mouse eye view. *Q Sci Rev.* 20:25–40.
- Richter C, Park J, Ames B. 1988. Normal oxidative damage to mitochondrial and nuclear-DNA is extensive. *Proc Natl Acad Sci USA.* 85:6465–6467.
- Samuels DC. 2004. Mitochondrial DNA repeats constrain the life span of mammals. *Trends Genet.* 20:226–229.
- Sanderson MJ. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol Biol Evol.* 14:1218–1231.
- Sasaki T, Nikaido M, Hamilton H, et al. (12 co-authors). 2005. Mitochondrial phylogenetics and evolution of mysticete whales. *Syst Biol.* 54:77–90.
- Smith MF, Patton J. 1991. Variation in mitochondrial cytochrome b sequence in natural populations of South American akodontine rodents (Muridae: Sigmodontinae). *Mol Biol Evol.* 8:85–103.
- Spradling T, Hafner M, Demastes J. 2001. Differences in rate of cytochrome-b evolution among species of rodents. *J Mammal.* 82:65–80.
- Springer MS, Debry RW, Douady C, Amrine HM, Madsen O, de Jong WW, Stanhope MJ. 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Mol Biol Evol.* 18:132–143.
- Springer MS, Murphy W, Eizirik E, O'Brien S. 2003. Placental mammal diversification and the cretaceous-tertiary boundary. *Proc Natl Acad Sci USA.* 100:1056–1061.
- Steppan SJ, Adkins RM, Anderson J. 2004. Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. *Syst Biol.* 53:533–553.
- Teeling EC, Springer MS, Madsen O, Bates P, O'Brien SJ, Murphy WJ. 2005. A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science.* 307:580–584.
- Thompson JD, Higgins DG, Gibson TJ. 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence. *Nucleic Acids Res.* 22:4673–4680.
- Thorne JL, Kishino H, Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol.* 15:1647–1657.
- Urbina D, Tang B, Higgs PG. 2006. The response of amino-acid frequencies to directional mutation pressure in mitochondrial genome sequences is related to the physical properties of the amino acids and to the structure of the genetic code. *J Mol Evol.* 62:340–361.
- Waddell PJ, Kishino H, Ota R. 2007. Phylogenetic methodology for detecting protein interactions. *Mol Biol Evol.* 24:650–659.
- Wallace DC. 2005. A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annu Rev Genet.* 39:359–407.
- Wang D, Kreutzer D, Essigmann J. 1998. Mutagenicity and repair of oxidative DNA damage: insights from studies using defined lesions. *Mutat Res.* 400:99–115.
- Weir JT, Schluter D. 2007. The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science.* 315:1574–1576.
- Yoder A, Yang ZH. 2000. Estimation of primate speciation dates using local molecular clocks. *Mol Biol Evol.* 17:1081–1090.

Ziheng Yang, Associate Editor

Accepted October 24, 2007

Mitochondrial DNA evolutionary dynamics: what can we learn from the bird/mammal comparison?

Benoit Nabholz^{*1}, Sylvain Glémin¹ and Nicolas Galtier¹

¹Université Montpellier 2 CNRS UMR 5554 - Institut des Sciences de l'Evolution Place E. Bataillon - CC064, 34095 Montpellier, France.

Email: Benoit Nabholz* - benoit.nabholz@univ-montp2.fr; Sylvain Glémin - sylvain.glemin@univ-montp2.fr ; Nicolas Galtier - nicolas.galtier@univ-montp2.fr;

*Corresponding author

Abstract

Background: During the last ten years, major advances have been made in characterizing and understanding the evolution of mitochondrial DNA, the most popular marker of molecular biodiversity. Several important results were recently reported using mammals as model organisms, including (i) the absence of relationship between mitochondrial DNA diversity and life-history or ecological variables, (ii) the absence of prominent adaptive selection, contrary to what was found in invertebrates, and (iii) the unexpectedly large variation in neutral substitution rate among lineages, revealing a possible link with species maximum life span. We propose to challenge these results thanks to the bird/mammal comparison. Direct estimates of population size are available in birds, and this group presents striking life-history trait differences with mammals (higher mass-specific metabolic rate and longevity). These properties make birds the ideal model to directly test for population size effects, and to discriminate between competing hypotheses about the causes of substitution rate variation.

Results: A phylogenetic measure of cytochrome *b* third-codon position reveals that the mitochondrial DNA mutation rate is quite variable in birds, passerines being the fastest evolving order. On average, mitochondrial DNA evolves slower in birds than in mammals of similar body size. This result is in agreement with the longevity hypothesis, and contradicts the hypothesis of a metabolic rate-dependent mutation rate. Birds show no footprint of adaptive selection on cytochrome *b* evolutionary patterns, but no link between direct estimates of population size and cytochrome *b* diversity. The mutation rate is the best predictor we have of within-species mitochondrial

diversity in birds. It partly explains the differences in mitochondrial DNA diversity patterns observed between mammals and birds, previously interpreted as reflecting Hill-Robertson interferences with the W chromosome.

Conclusions: Mitochondrial DNA biodiversity patterns in birds are strongly influenced by the wide, unexpected variation of mutation rate across species. From a fundamental point of view, these results are strongly consistent with a relationship between species maximal longevity and mitochondrial mutation rate, in agreement with the mitochondrial theory of ageing. From an applied point of view, this study reinforces and extends the message of caution previously expressed for mammals: mitochondrial data tell nothing about species population sizes, and strongly depart the molecular clock assumption.

Background

Animal mitochondrial DNA (mtDNA) evolution contrasts with nuclear evolution. Mitochondrial and nuclear genomes differ in many ways, such as total length, ploidy level, mode of inheritance, recombination rate, presence of introns, percentage of non-coding DNA, effective population size, and repair mechanisms, *e.g.* [1,2]. Among these particularities, hypermutability is one of the most striking features of animal mitochondria: the mtDNA mutation rate is typically one order of magnitude higher than the nuclear one [1,3,4]. The question of the origin and evolution of such high mutation rates is still open and debated [5,6]; and see [7,8] for exceptions.

This high mutation rate is one of the reasons why mtDNA is a very popular marker for biodiversity studies. For example, it has been massively used to investigate intraspecific to intra-ordinal evolutionary relationship, and disentangle rapid speciation events in phylogenetic studies (*e.g.* see [9,10] for a comparison of nuclear and mitochondrial markers). In recent years, mtDNA has been used to identify species using a standardized portion of the cytochrome oxidase I gene (COI), according to the so-called DNA barcoding approach [11]. The high mtDNA mutation rate, however, is also the source of frequent homoplasy, *i.e.*, phylogenetic incongruence between sites of the molecule because of multiple mutations at the same site. Homoplasy complicates the use of mtDNA in phylogenetic [9,10,12] and even population

genetic [13] studies.

We recently reported several results related to mtDNA mutation dynamics in mammals. First, we showed that mtDNA substitution rates are extremely variable between species: they differ by two orders of magnitude between slow-evolving and fast-evolving mammalian lineages [14], definitively rejecting the famous “2% per site per million year” calibration, which should not be generally trusted. We proposed that mutation rate variations are possibly linked to species longevity through the action of natural selection: too high a mutation rate would be deleterious in long-lived mammals because it could result in premature aging due to the accumulation of somatic mtDNA mutations [14], see also [15–17]. Secondly, we found that within-species cytochrome *b* (*cytb*) nucleotide diversity is correlated with the mitochondrial mutation rate, as expected, but not with any life-history or ecological variable potentially related to population size, including body size, geographic range and conservation status [18]. This surprising pattern was apparently not explained by selective effects [18] (see also [19]), contrary to results obtained at the Metazoa level [20], especially invertebrates [21]. We therefore proposed that the lack of correlation between mtDNA diversity and potential indicators of population size in mammals is due to strong demographic stochasticity. At any rate, the mutation rate is the best predictor we have of mitochondrial genetic diversity across mammalian species. Overall, these results highlight the importance of mutation rate variations in shaping mtDNA biodiversity patterns in mammals, and suggest that mutational effects should be carefully taken into account when analysing such data.

In this study, we extend our analysis of the evolutionary dynamics of mtDNA through the bird/mammal comparison. Our objective is dual. First, we want to check whether the mammalian results are specific to this group, or have any degree of generality. Secondly, we want to make use of the genetical and physiological specificities of birds to test various evolutionary hypothesis raised by previous studies. Birds are the perfect candidate to challenge the results obtained in mammals: like mammals, they include charismatic, well studied species, in which a large amount of genetic and biological data are available. Being relatively large, warm blooded vertebrates, birds are comparable to mammals in terms of physiology, ecology, and life-history. Birds, furthermore, show a number of peculiarities potentially relevant to mitochondrial evolution and diversity.

The first reason why birds are appropriate for a comparative approach is the availability of direct population size estimates obtained through global population surveys, especially for North American

species [22]. Birds therefore provide the opportunity to directly test the relationship between population size and mitochondrial diversity, whereas in mammals we had to rely on ecological and life history traits, plausibly, but only supposedly, correlated to population size. We will therefore check whether the lack of relationship between mtDNA diversity and species abundance we reported in mammals is confirmed in birds, or was due to inappropriate measure of the effective population size.

Birds, secondly, present a genetic peculiarity: female is the heterogametic sex. Because it is strictly maternally transmitted, the avian mitochondrial genome is thus in full genetic linkage with the female-specific W chromosome [23]. Berlin *et al.* [24] proposed that the mtDNA diversity could therefore be reduced by Hill-Robertson interference: selective effects applying to loci linked to the W will affect mtDNA through hitch-hiking. This clever hypothesis has provoked some comments and reactions [24–27]. Hickey [26] questioned one of the most important arguments of Berlin *et al.* [24], namely the lower synonymous diversity in bird than in mammal mtDNA, arguing that mutation rate could be a confounding effect. An accurate estimation of mtDNA substitution rate variations in birds appears necessary to correctly interpret the patterns of mtDNA diversity in the bird/mammal comparison.

Birds, finally, have quite high metabolic rates - 1.5 to 2.5 times higher than mammals of similar sizes [28] - probably because of the high energetic demand of the flying locomotion. Paradoxically, birds species are strikingly long-lived as compared to their mammalian counterparts; on average, birds live three times longer than mammals of similar sizes [28, 29]. Birds can be characterized as long-lived homeotherms, a specificity they share with bats [30]. These peculiarities are useful to understand further the impact of metabolic rate and life-span on mtDNA substitution rates.

Three main hypotheses have been proposed to explain mtDNA substitution rate variations in mammals: the generation time hypothesis [30–33], the metabolic rate hypothesis [34, 35], and the longevity hypothesis [14–17], presented above. The classical approach to test these hypotheses in a comparative framework is to correlate substitution rate variations with the relevant life history traits [14, 17, 33]: female sexual maturity, basal mass-specific metabolic rate (or body mass, taken as a proxy), and maximum longevity. However, these life-history traits are strongly correlated with each other, so that the respective contributions of the three variables are difficult to disentangle. The contrast between the avian and mammalian physiologies provides a unique opportunity to discriminate between two of these competing models, namely the longevity and the metabolic rate hypotheses. According to the longevity hypothesis,

birds should present, on average, lower mtDNA neutral substitution rates than mammals, while the reverse pattern is expected under the metabolic rate hypothesis.

To test these predictions and challenge the results we obtained in mammals, we propose to accurately estimate the lineage-specific neutral substitution rate variation of mtDNA in birds using the phylogenetic framework developed by [14]. Using this dataset, we want to check the rough 2%/site/Myr calibration. This calibration, originally estimated with an RFLP analysis on mammals by Brown *et al.* [3], was generalized to birds by a study in geese [36]. This calibration rapidly became a standard in ornithology studies, probably because of the scarcity of the fossil record (see [37,38] and [39] for a review). Recently, the reliability of this calibration was debated in various studies either supporting [40–42] or rejecting it [39,43]. We will also ask which of metabolic rate and longevity is the best predictor of mutation rate variations in warm-blooded vertebrates. Finally, we will estimate the mitochondrial genetic diversity (synonymous and non synonymous) on a wide taxonomic range in birds to test for the neutrality assumption and investigate the link between population size, mutation rate, genetic hitch-hiking, and mtDNA diversity.

Results

Substitution rate variation and its determinants

To infer the substitution rate variation of a large number of bird species, we need to estimate divergence dates, and the numbers of nucleotide substitutions which occurred during these divergences (branch lengths). The divergence dates are obtained thanks to paleontologic calibration points, which are mostly available for relatively old divergences. Nucleotide substitution numbers, however, must be estimated using little-divergent sequences because of mutational saturation. To resolve this methodological problem, we take advantage of the decoupled non-synonymous (*i.e.* amino acids) and synonymous evolutionary dynamics. We introduced a 3-step method [14], in which we first define groups of sequences showing limited divergences, then estimate the relative species-specific substitution rate at third codon positions within groups, and finally assign a divergence for each of these groups using amino acid sequences. At step 2 and 3, molecular and fossil information are combined thanks to a sophisticated Bayesian method [44].

Using the procedure, we analysed a *cytb* dataset of 1571 species, which represent $\sim 15\%$ of the total living species of birds. The average *cytb* third codon position substitution rate was 0.027 substitution/site/Myr, and the median was 0.018 substitution/site/Myr, which is close to the popular 2%. However, the distribution across species shows a large variation : the 2.5% and 97.5% percentiles are 0.003 and 0.090

substitution/site/Myr, respectively (Figure 1). Thus, even if we consider the 5% most extreme substitution rate values as outliers, the cytb third codon position substitution rate shows a thirty-fold range of variation across bird lineages. We quantified the effect of taxonomy on substitution rate variation using a hierarchical ANOVA considering super-order, order, family and genus with random effects. The order level explains 46% (63% in simple one-way ANOVA) of the variance in substitution rates across species. Considering only order with more than 20 species, Passerines are the fastest evolving order (0.038 ± 0.055 substitution/site/Myr) and Anseriformes the slowest one (0.004 ± 0.0008 substitution/site/Myr, see Figure 2). Compared with mammals, birds globally evolve 3.7 times slower on average (mammalian average: 0.098 substitution/site/Myr, Nabholz *et al.* [14]), and show a range of variation less extreme (coefficient of variation: 1,58 for birds, 1,72 for mammals, Figure 3).

We correlated the species-specific substitution rate to life-history traits in order to discriminate between the competing explanatory hypotheses. A smaller amount of life-history data are available in birds than in mammals, especially for female sexual maturity, which is documented in 30 species only. We thus focused on the effects of body mass and maximum longevity using the 196 species in which these two life-history traits are available. In agreement with both the metabolism and the longevity hypotheses, the two life-history traits are negatively linked to the substitution rate. Body mass appears to be the best predictor of the substitution rate ($R^2 = 0.39$, $p < 0.001$, $n = 196$), but maximum longevity also shows a strong level of correlation ($R^2 = 0.25$, $p < 0.001$, $n = 196$, Figure 4, Table 1).

The distribution of (log-transformed) maximum longevity is strongly bimodal in birds, passerines being substantially more short-lived than other birds (see Figure S1). We separated passerines from non-passerine and re-performed the same analyses in the two subsets separately. The results are shown in Table 2. Surprisingly, the third-codon position substitution rate is positively correlated to body mass and maximum longevity in passerines birds. In non-passerines birds, results are similar to the global analysis, body mass best explains substitution rate variations, and removes the effect of longevity in a two-way ANOVA (Table 2). This result contradicts the mammalian ones [14].

In order to check whether these results were robust to the removal of phylogenetic effects. We applied the phylogenetic contrast method [45] to regress phylogeny out of the analyses. In this analysis, we used chronogramme obtain with amino-acids cytochrome b sequences. The phylogenetic control do not

qualitatively change the results although most p-values were decreased (Table 1, Table 2).

Finally, we jointly analysed the mammalian and bird datasets (Table 3, Figure 5), including class (birds vs mammals) as an explanatory variable. The two-factor models (class + life history trait) revealed that the bird *vs.* mammal status has a strong effect on mtDNA substitution rate when body mass is taken with lesser effect and a significant but weaker effect as far as longevity is concerned (Table 3). There is, moreover, a significant interaction between body mass and class ($p < 0.001$), but not between longevity and class ($p = 0.65$).

Determination of mtDNA diversity in birds

Bird cytb polymorphism sequence datasets were retrieved from the Polymorphix database [46]. The synonymous (π_s) and non-synonymous (π_n) levels of diversity were computed for the 147 species in which more than four sequences were available. The mean per-site synonymous nucleotide diversity was 0.040 ± 0.041 . The most variable species was Shelley's Grenbul (*Andropadus masukuensis*, Pycnonotidae, Passeriformes, $\pi_s = 0.170$), and two species showed no synonymous variations (Arabian Bustard, *Ardeotis arabs*, Otididae, Gruiformes, and Snowy owl, *Bubo scandiacus*, Strigidae, Strigiformes).

To test the relationship between population size and mtDNA diversity, we first correlated nucleotide polymorphism and body mass, taken as a proxy of population size, as in Nabholz *et al.* [18] and Berlin *et al.* [24]. π_s and body mass were negatively correlated but, contrary to Berlin *et al.* [24], this relationship was not significant ($R^2 = 0.019$, $p = 0.30$, $n = 56$). Secondly, we applied a more direct analysis using population size estimates for North American breeding birds through the North American Landbird Conservation Plan [22]. Such data were available for 28 strictly North American species of our data set. We found no correlation between direct population size estimate and cytb synonymous diversity (Figure 6, $t = 0.031$, $p = 0.81$, Kendall's test). These results confirm those obtained on mammals [18], here using direct estimations of population sizes instead of life-history and ecological proxies. The dataset, however, is much smaller in the present analysis.

We tested the taxonomic effect on π_s using one-way ANOVAs on order and family level. We found that neither the family nor the order level had a significant effect. The sole taxonomic variable having a significant effect on mtDNA diversity is the passerine/non-passerine status, passerines being more diverse

than other birds (passerines: $n = 79$, non-passerines: $n = 68$, $R^2 = 0.03$, $p < 0.05$).

We compared the distribution of π_s between birds and mammals. Similarly to Berlin *et al.* [24], we found that birds have a lower average synonymous diversity than mammals ($n = 169$, $R^2 = 0.09$, $p < 0.001$, Figure S2a), but this difference is largely reduced when we controlled for the mutation rate (previously estimated above after control, $R^2 = 0.03$, $p = 0.032$, Figure S2b). Interestingly, we detected no significant difference of allozyme heterozygosity between mammals (average $H = 0.046 \pm 0.033$) and birds (0.053 ± 0.028) using the Nevo *et al.* [47] data set (birds : $n = 42$, mammals: $n = 164$, $R^2 = 0.01$, $p = 0.08$). As in mammals, we found a positive and significant correlation between species-specific cytb third-codon substitution rate and cytb synonymous diversity per genus ($n = 46$, $R^2 = 0.10$, $p = 0.02$, Figure 7). These analyses confirm that the mutation rate is a major determinant of mtDNA polymorphism in birds as well as in mammals.

Finally, we compared the ratio of non synonymous to synonymous changes within (π_n/π_s) and between species (D_n/D_s) in birds and mammals - these ratio were computed when an outgroup was available. The two groups showed a similar D_n/D_s (birds : $n = 81$, average $D_n/D_s = 0.019 \pm 0.011$; mammals: $n = 75$, average $D_n/D_s = 0.023 \pm 0.024$; $R^2 = 0.01$, $p = 0.19$), but the π_n/π_s ratio was higher in birds than in mammals (birds : average $\pi_n/\pi_s = 0.118 \pm 0.110$; mammals : average $\pi_n/\pi_s = 0.080 \pm 0.062$; $R^2 = 0.04$, $p < 0.05$), as previously reported by Berlin *et al.* [24].

Discussion

mtDNA substitution rates - support for the longevity hypothesis

We showed that the mtDNA mutation rate, measured through the neutrally-evolving third codon positions of cytb, is highly variable between bird lineages. The molecular clock hypothesis does not apply to bird mtDNA evolution, fast-evolving species being thirty times more rapid than slow-evolving ones. The method we applied is particularly prone to reveal the actual amplitude of substitution rate variation: by overcoming the problem of mutational saturation, it allowed us to use an extensive dataset encompassing the whole taxonomic diversity of birds. This study, therefore, provides a synthetic view of substitution rate variations across bird lineages, bringing an important result in the field of molecular dating, in which the rough approximation of 2% of substitution per million of years is still debated [39–43]. Our analysis suggests that the molecular clock assumption should be avoided as far as bird mtDNA is concerned. The conclusions of Pereira and Baker [43] (obtained from full-genome data), and some studies done at smaller

scale (reviewed in [39]) are here confirmed by cytochrome b third codon positions at wide taxonomic scale. Although the average rate in birds is close to this value, the 2% per site per million year calibration is a very bad summary of the whole picture: the rate is essentially (up to five times) higher than 2% in passerines, and (up to ten times) lower in non-passerines (Figure 1). Users of mtDNA as a tools for inferring divergence dates should imperatively use statistical phylogenetic methods accounting for substitution rate variation across lineages, the so-called clock-relaxed methods [44, 48–50].

Several interesting results were revealed thanks to the comparison with mammals. First, birds show a more narrow range of variation of life-history traits than mammals. For example body mass varies from 5.5 g to 3.9 kg in the bird dataset, whereas it varies from 3.7g to 138 tones in mammals. This result confirms the general influence of life-history traits on neutral mtDNA substitution rate in birds and mammals. This result is important because such a relationship was not reported in every taxonomic groups [51], and even not by some studies in mammals [52, 53]. Secondly, the neutral mtDNA substitution rate is lower in birds than in mammals, and this difference increases if body mass is regressed out of the analysis (Table 2). For example, birds evolve four times slower than mammals as far as small species (body mass < 500 g) are concerned. This result is in agreement with the longevity hypothesis, but not with the metabolic rate hypothesis. According to the latter model, birds should show a higher average mutation rate than mammals, because of their higher mass-specific metabolic rate and their lower average body mass [54].

Within birds, body mass appears to be the factor best explaining substitution rate variations, but this result is mostly caused by the odd behavior of passerine birds. Passerines show a positive relationship between longevity (or body mass) and substitution rate, which is quite surprising, and perhaps not a biologically relevant result: this relationship was not recovered when we used other methods of substitution rate estimation (see Methods). Within non-passerines birds, which are essentially long-lived species, longevity is the variable best explaining substitution rate variations; the longevity effect removes the body mass effect in multiple regression analyses (Table 1). This result parallels the one found by Nabholz *et al.* [14], who reported that longevity is the best explanatory variable in long-lived mammals.

During the course of evolution, birds have acquired adaptations to manage their high mass-specific metabolic rate, including an increased of reactive oxygen species (ROS) protection and lower ROS production [55, 56]. In the context of the mitochondrial theory of aging, which postulates that ROS production as a byproduct of mitochondrial respiration contributes to aging, these adaptations could be

interpreted as a response to the elevated longevity of birds [57]. The reduced rate of ROS production could explain the lower average mutation rate of birds mtDNA, as compared to mammals (see also [26]). It is surprising, therefore, to find a predominant influence of body mass, not longevity, in the within-birds analysis. Passerines show a positive relationship between longevity (or body mass) and substitution rate, which is quite surprising, and perhaps not biologically relevant - this relationship was not recovered when we used other methods of substitution rate estimation (see Table S4). Within non-passerines, which are essentially long-lived species, body mass is still the variable best explaining substitution rate variations (Table 2). Using DNA-DNA hybridization data (and a smaller dataset), Mooers and Harvey [32] found no support for the metabolic rate hypothesis, but reported a significant generation time effect. In this context, it would be interesting to enlarge the female sexual maturity data set, in order to test the generation time hypothesis with our extended mitochondrial dataset.

Determination of mtDNA diversity in birds

The neutral genetic diversity primarily depends on effective population size and mutation rate. Because population sizes appear so variable in time and between species, variations in genetic diversity across species have been frequently interpreted in demographic terms, *e.g.* [58,59]. Recently, the effect of population size on mtDNA diversity has been challenged [18,20]. The large variations in mutation rates we report in birds suggest that mutational effects should be carefully taken into account in studying mtDNA polymorphism in this group, as well as in mammals [18].

In general, the relationship between population size and mtDNA diversity is difficult to evaluate because of the scarcity of population size estimates for wild species. Life-history traits, like body mass, have therefore been used as proxy of population size [18,60]. Birds are an exception: thanks to the broad interest of the scientific community in bird ecology and systematics, direct estimates of species effective sizes are available for North American species [22]. We correlated these population size estimates to synonymous mtDNA diversity and found no significant relationship. The data set is rather small ($n = 28$), but we note that this relationship should theoretically be strong, given the wide range of population sizes (2 orders of magnitude). When using body mass as an indicator of population size, thus increasing much the data set, we found a negative correlation with *cytb* synonymous diversity, but this result was not significant. Berlin *et al.* [24] report a significant relationship between π_s and body mass in birds using genus averages. If we apply the same method, we also find a significant relationship ($n = 37$, $R^2 = 0.16$, $p = 0.01$), but the effect

is removed when the passerine/non-passerine status is taken into account (multiple regression, p body mass = 0.25). This result is in agreement with our previous mammalian analysis, in which the weak effect of body mass was removed by a taxonomic control [18]. So similarly to mammals, the mitochondrial genetic diversity is essentially uncorrelated to life-history traits in birds, and when it is so, it is most likely via the influence of mutation rate, not population size.

Mitochondrial selective regime in birds

To explain the lack of relationship between mtDNA diversity and proxies of population size in mammals, Nabholz *et al.* [18] invoked demographic stochasticity, while Bazin *et al.* [20] invoked recurrent hitchhiking effects for other species having larger expected population sizes (invertebrates and marine species). What about Birds? To address this question we computed the neutrality index ($NI = \pi_n/\pi_s / d_n/d_s$; [61], see Methods). $NI < 1$ indicates positive selection, while $NI > 1$ indicates purifying selection. The average NI in birds is significantly higher than in mammals ($R^2 = 0.058$, $p < 0.01$, Figure S2) and only three species of birds show a $NI < 1$ (seven in mammals). The high NI values in birds are due to their high π_n/π_s ratios - d_n/d_s ratios are similar to those of mammals. This result indicates that mtDNA evolution in birds is mainly governed by purifying selection, as previously reported [24]. It is also indicative of a strong level of constraint acting on bird cytochrome b sequences, which is known as the avian constraint hypothesis [62].

Like mammals, birds appear to have population sizes low enough to belong to the “drift domain”, rather than the “draft domain” *sensu* Gillespie [63] (see discussion in [18]). In the absence of a significant impact of positive selection, we have to invoke strong demographic stochasticity to explain the lack of relationship between population size and mtDNA diversity. Mutation rate is expected to be a major determinant of within-species genetic diversity whether populations are at mutation-drift equilibrium (standard theory) or not [64, 65]. Consistently, we report a significantly positive relationship between cytb diversity and substitution rate (Figure 7), again paralleling the mammalian situation.

In a recent paper, Berlin *et al.* [24] proposed an alternative hypothesis to explain mtDNA polymorphism patterns in birds. They suggested that Hill-Robertson effects due to linkage with the W chromosome could explain the lower synonymous mitochondrial diversity and the higher π_n/π_s observed in birds, as compared to mammals. This should also explain the lack of correlation between neutral polymorphism and population size. Outwardly, we confirm this hypothesis by showing that the lower synonymous mitochondrial diversity is not found for nuclear allozymes. But this picture is complicated by the globally

lower mtDNA substitution rate in birds than in mammals. Controlling for mutation rates actually removes the major part of the difference between birds and mammals. To quantify the effect of mutation rate variation on synonymous mtDNA diversity we combined the bird and mammal polymorphism and neutral substitution rate datasets and performed a multiple regression of substitution rate and bird/mammal status (class) on synonymous mtDNA diversity. This model explains 21% of π_s variation (birds : $n = 46$, mammals : $n = 123$, $p < 0.001$). The result also shows that the mutation rate has a strong effect on π_s ($t = 4.9$; $p < 0.001$) and considerably reduces the effect of class ($t = 2.1$, $p = 0.021$). This analysis indicates that a substantial part of the difference between birds and mammalian π_s is explained by their distinct mutation rates (see Figure S3b). The difference in mutation rates, however, cannot explain the higher level of mitochondrial π_n/π_s found in birds. Birds and mammals have apparently comparable effective population sizes, as suggested by their similar average allozymic heterozygosity. The higher π_n/π_s ratio in birds may therefore be due to Hill-Robertson interferences between the mitochondrial and W chromosomes, as suggested by Berlin *et al.* [24], although this effect is not detectable on absolute levels of π_s , when mutation rate is controlled for.

Conclusions

In this work, we achieved a comprehensive comparison of mitochondrial substitution rate variation and mitochondrial diversity between birds and mammals. Cytb neutral substitution rates are, on average, 3.9 times slower in birds than in mammals. Among-lineages variability, although substantial, is also lower in birds than in mammals. This analysis corroborates the longevity hypothesis about mutation rate determination [14, 17]. Despite their higher metabolic rate, birds undergo a lower mitochondrial mutation rate, probably because of the beneficial effect of reduced oxidative damage in long-lived species [26, 28]. We suggest that the low mtDNA mutation rate of birds might explain their exceptional mass-specific longevity.

As for mammals, we report no evidence of adaptive selection on cytb sequence evolution, but still no population size effect, even when using direct estimates of population size available for North American species. We thus invoke the same demographic instability explanation as previously proposed for mammals [18], although selective interferences with the W chromosome [24] could contribute to the noisy relationship between mtDNA diversity and population size. We also confirm the importance of mutation rate as a determinant of mtDNA diversity - the lower mutation rates in birds almost fully explain their

lower synonymous diversity.

This study confirms and extends the message of caution expressed by Nabholz *et al.* [14,18] about the usage of mtDNA as a molecular marker of biodiversity in vertebrates: (i) mtDNA diversity is not related to species abundance; (ii) mtDNA greatly departs the molecular clock hypothesis. The 2% per site per million year calibration (estimated from primate data) has no degree of generality, and should not be used for dating purposes in the absence of fossil data.

Methods

Sequence data for substitution rate analysis

Complete bird cytochrome b sequences were extracted from National Center for Biotechnology information/Genbank. One sequence per species (the longest, excluding indeterminations) was selected. Accession numbers are given in Table S2. The 5'-most 102 nucleotides were removed from the alignments, because missing in a large number of species. Total alignment length was 1043 nucleotides.

Sequence data and Allozyme data for polymorphism analysis

An mtDNA dataset was built from Polymorphix [46]. Polymorphix is a database dedicated to sequence polymorphism. It contains within-species homologous sequence families built from EMBL/GenBank under suitable similarity and bibliographic criteria. To obtain a homogeneous dataset, comparable with a previous mammalian study [18], only the protein coding cytochrome *b* gene was used. We extracted from Polymorphix every bird sequence family for which any >100-bases-long cytochrome b fragment was available in four individuals or more. Polymorphix sequence families were aligned using CLUSTALW [66], inspected by eye and corrected when required. Dubious sequences (badly aligned or including many undetermined nucleotides) were manually removed. Alignments are available at

<http://kimura.univ-montp2.fr/~benoit/data.html>.

Estimates of allozyme heterozygosity in 74 bird and 194 mammalian species were obtained from [47].

Allozyme heterozygosities were averaged using 10-30 loci. The most popular allozyme loci are shared by many species. Allozyme heterozygosities are therefore fairly comparable between species.

Substitution Rate estimation

We used the same uncoupled method as previously developed for mammals [14]. The whole cytochrome b data set was split in groups within which sequence divergence is moderate. To achieve this, the GenBank

taxonomic classification was traversed recursively, starting from birds orders and moving toward lower levels. For each traversed taxonomic group, we 1) gathered the corresponding third codon position sequences, 2) aligned sequences using ClustalW [66], 3) built a maximum likelihood phylogenetic tree using PHYML [67], general time reversible + Gamma model of nucleotide evolution, and 4) calculated the pairwise patristic distances for every pair of species (defined as the sum of branch lengths in the path connecting the 2 species in the tree). When the median pairwise distance between species was lower than 0.4, the current taxonomic group was selected for further analysis, (except for the Falconiformes and Ciconiiformes orders, for which monophyly is uncertain [68,69]), and the taxonomic traversal was stopped. Otherwise, the above procedure was applied to underlying taxonomic groups. The within group, species-specific relative neutral substitution rates were estimated using two different softwares: MULTIDIVTIME [44] and MCMCTREE [50,70] available in the PAML packages [71]. The major difference is that MULTIDIVTIME makes use of a normal approximation of the likelihood, and only implements the F84 + Γ model of sequence evolution [44], whereas MCMCTREE performs exact likelihood calculation and can be used with different models of sequence evolution - we used the HKY + Γ . The Monte Carlo Markov Chain was run for 1 million generations after a burn-in of 200,000 generations that achieved stationarity.

Bird phylogeny is still partly uncertain, particularly the basal relationships within the Neoaves clade [68,69,72–74]. We used two alternative phylogenetic trees to date divergences. The first one is conform to Ericson *et al.* [68] topology, obtained using mostly nuclear markers, and the second one is conform to the complete genome mitochondrial topology proposed by Slack *et al.* [69]. In both topologies, the relationships within passerines were made congruent to Barker *et al.* [75]. Groups were dated by applying MULTIDIVTIME to amino-acid sequences (two representative species per group, model mtREV + Γ). Twelve fossil calibrations points were used (table S1). The results were very similar whatever the method and the tree (table S2 and table S3), so we chose to present only the results obtained with MULTIDIVTIME using the mitochondrial genome topology, except when mentioned. We checked that neither the mean nor the variance of species-specific estimated substitution rates are correlated to terminal branch lengths (not shown). Sequence management and GenBank taxonomic exploration were achieved using homemade C++ programs based on the Bio++ libraries [76].

Polymorphism sequence data analyses

Two measures of molecular genetic diversity were used, namely the nucleotide diversity π [77] and Watterson's statistics θ_w [78]. In the case of the haploid, maternally transmitted mtDNA, both statistics are unbiased estimates of the $Ne\mu$ product under the assumption of neutrality and mutation/drift equilibrium, where Ne is the effective population size of females and μ the locus mutation rate. π and θ_w were calculated from the total length of the analysed fragments, and expressed in per-site level of diversity (after being divided by sequence length). We also measured the synonymous and non-synonymous nucleotide diversity (π_s, π_n), also expressed per-synonymous site and per-non synonymous site.

The Neutrality Index (NI, [61]) was calculated for dataset 1 when outgroups were available. This index aims at comparing the ratio of non-synonymous (= amino acid changing) to synonymous (silent) changes within species (π_n/π_s) and between species (d_n/d_s): NI is one when evolution is neutral, higher than one under purifying selection, and lower than one in case of adaptation. π_n and π_s were estimated as described above. d_n and d_s are the non-synonymous and synonymous pairwise sequence divergences between related taxa, calculated using the maximum likelihood method implemented in codeml in PAML 4 [71]. Excluding low frequency variants (<0.125) when estimating NI yielded similar results (not shown). The phylogenetic tree was obtained using PHYML software [67] with TN93 + Γ model of sequences evolution. We used the maximum likelihood method because of the potential underestimation of d_s in simple pairwise distance due to multiple hits. Alignments are available at <http://kimura.univ-montp2.fr/~benoit/data.html>.

Life History data and Population size estimation

Body mass, age of female sexual maturity, and maximum longevity were obtained from the AnAge database [79]. Direct estimates of population size are available for North American breeding birds in the North American Landbird Conservation Plan, Breeding Bird Survey abundance data [22], <http://www.partnersinflight.org/>. Life history data and rate estimates are available at <http://kimura.univ-montp2.fr/~benoit/data.html>.

Statistical analyses

Genetic diversity measures were arcsine-transformed [80] and analyzed under the general linear model assumptions using R [81]. Quantitative life-history variables were log-transformed. One-way and two-way ANOVA with interactions were performed on transformed variables. Fisher's tests were performed using

class II errors.

Authors contributions

All authors designed research project. BN compiled the data and performed the analyses. All authors wrote and approved the final manuscript.

Acknowledgements

This work was supported by the Agence National de la Recherche (projet MITOSYS) and the Centre National de la Recherche Scientifique. We thank the three anonymous referees for their helpful comments. We thank the ISEM bioinformatic platform for providing the computing resources. This is manuscript number ISEM-XXXX.

References

1. Ballard J, Whitlock M: **The incomplete natural history of mitochondria.** *Molecular Ecology* 2004, **13**(4):729–744.
2. Lane N: *Power, Sex, Suicide. Mitochondria and the Meaning of Life.* Oxford University Press 2005.
3. Brown WM, George M, Wilson AC: **Rapid evolution of animal mitochondrial DNA.** *Proceedings of the National Academy of Sciences of the United States of America* 1979, **76**(4):1967–1971.
4. Lynch M: **The origins of eukaryotic gene structure.** *Molecular Biology and Evolution* 2006, **23**(2):450–468.
5. Rand DM: **Thermal habit, metabolic-rate and the evolution of mitochondrial-DNA.** *Trends in Ecology & Evolution* 1994, **9**(4):125–131.
6. Palmer JD, Adams KL, Cho YR, Parkinson CL, Qiu YL, Song KM: **Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates.** *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97**(13):6960–6966.
7. Hellberg ME: **No variation and low synonymous substitution rates in coral mtdna despite high nuclear variation.** *BMC Evolutionary Biology* 2006, **6**:24–24.
8. Huang D, Meier R, Todd PA, Chou LM: **Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding.** *Journal of Molecular Evolution* 2008, **66**(2):167–174.
9. Springer M, Debry R, Douady C, Amrine H, Madsen O, de Jong W, Stanhope M: **Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction.** *Molecular Biology and Evolution* 2001, **18**(2):132–143.
10. Galewski T, Tilak M, Sanchez S, Chevret P, Paradis E, Douzery EJP: **The evolutionary radiation of arvicolinae rodents (voles and lemmings): relative contribution of nuclear and mitochondrial dna phylogenies.** *BMC Evolutionary Biology* 2006, **6**:80–80.
11. Ratnasingham S, Hebert PDN: **BOLD: the barcode of life data system (www.Barcodinglife.Org).** *Molecular Ecology Notes* 2007, **7**(3):355–364.
12. Springer MS, Amrine HM, Burk A, Stanhope MJ: **Additional support for Afrotheria and Paenungulata, the performance of mitochondrial versus nuclear genes, and the impact of data partitions with heterogeneous base composition.** *Systematic Biology* 1999, **48**:65–75.
13. Galtier N, Enard D, Radondy Y, Bazin E, Belkhir K: **Mutation hot spots in mammalian mitochondrial DNA.** *Genome Research* 2006, **16**(2):215–222.

14. Nabholz B, Glémin S, Galtier N: **Strong variations of mitochondrial mutation rate across mammals—the longevity hypothesis.** *Molecular Biology and Evolution* 2008, **25**:120–130.
15. Samuels D: **Mitochondrial DNA repeats constrain the life span of mammals.** *Trends in Genetics* 2004, **20**(5):226–229.
16. Khaidakov M, Heflich RH, Manjanatha MG, Myers MB, Aidoo A: **Accumulation of point mutations in mitochondrial DNA of aging mice.** *Mutation Research/dna Repair* 2003, **526**(1-2):1–7.
17. Welch JJ, Bininda-Emonds ORP, Bromham L: **Correlates of substitution rate variation in mammalian protein-coding sequences.** *BMC Evolutionary Biology* 2008, **8**:53–53.
18. Nabholz B, Mauffrey JF, Bazin E, Galtier N, Glémin S: **Determination of mitochondrial genetic diversity in mammals.** *Genetics* 2008, **178**:351–361.
19. Mulligan C, Kitchen A, Miyamoto M: **Comment on "population size does not influence mitochondrial genetic diversity in animals".** *Science* 2006, **314**(5804):1390–1390.
20. Bazin E, Glemin S, Galtier N: **Population size does not influence mitochondrial genetic diversity in animals.** *Science* 2006, **312**(5773):570–572.
21. Hurst G, Jiggins F: **Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts.** *Proceedings Of The Royal Society B-biological Sciences* 2005, **272**(1572):1525–1534.
22. Rich T, Beardmore C, Blancher P, Bradstreet M, Butcher G, Demarest D, Dunn E, Hunter W, Iñigo Elias E, Kennedy J, Martell A, Panjabi A, Pashley D, Rosenberg K, Rustay C, Wendt J, Will T: *Partners in Flight North American landbird conservation plan.* Cornell Lab of Ornithology 2004.
23. Berlin S, Ellegren H: **Evolutionary genetics. Clonal inheritance of avian mitochondrial DNA.** *Nature* 2001, **413**(6851):37–38.
24. Berlin S, Tomaras D, Charlesworth B: **Low mitochondrial variability in birds may indicate Hill-Robertson effects on the W chromosome.** *Heredity* 2007, **99**(4):389–396.
25. Marais GAB: **The Hill-Robertson effects extend from nucleus to mitochondria.** *Heredity* 2007, **99**(4):357–358.
26. Hickey AJR: **Avian mtDNA diversity : An alternate explanation for low mtDNA diversity in birds: an age-old solution?** *Heredity* 2008, **100**(5):443–443.
27. Lane N: **Mitochondria and the W chromosome: low variability on the W chromosome in birds is more likely to indicate selection on mitochondrial genes.** *Heredity* 2008, **100**(5):444–445.
28. Holmes DJ, Flückiger R, Austad SN: **Comparative biology of aging in birds: an update.** *Experimental Gerontology* 2001, **36**(4-6):869–883.
29. Holmes DJ, Ottinger MA: **Birds as long-lived animal models for the study of aging.** *Experimental Gerontology* 2003, **38**(11-12):1365–1375.
30. Li WH, Tanimura M, Sharp PM: **An evaluation of the molecular clock hypothesis using mammalian DNA sequences.** *Journal of Molecular Evolution* 1987, **25**(4):330–342.
31. Ohta T: **An examination of the generation-time effect on molecular evolution.** *Proceedings of the National Academy of Sciences of the United States of America* 1993, **90**(22):10676–10680.
32. Mooers AO, Harvey PH: **Metabolic rate, generation time, and the rate of molecular evolution in birds.** *Molecular Phylogenetics and Evolution* 1994, **3**(4):344–350.
33. Bromham L, Rambaut A, Harvey P: **Determinants of rate variation in mammalian DNA sequence evolution.** *Journal of Molecular Evolution* 1996, **43**(6):610–621.
34. Martin AP, Naylor GJ, Palumbi SR: **Rates of mitochondrial dna evolution in sharks are slow compared with mammals.** *Nature* 1992, **357**(6374):153–155.
35. Martin AP, Palumbi SR: **Body size, metabolic-rate, generation time, and the molecular clock.** *Proceedings of the National Academy of Sciences of the United States of America* 1993, **90**(9):4087–4091.
36. Shields GF, Wilson AC: **Calibration of mitochondrial DNA evolution in geese.** *Journal of Molecular Evolution* 1987, **24**(3):212–217.

37. Bermingham E, Rohwer S, Freeman S, Wood C: **Vicariance biogeography in the Pleistocene and speciation in North American wood warblers: a test of Mengel's model.** *Proceedings of the National Academy of Sciences of the United States of America* 1992, **89**(14):6624–6628.
38. Weir JT, Schluter D: **The latitudinal gradient in recent speciation and extinction rates of birds and mammals.** *Science* 2007, **315**(5818):1574–1576.
39. Lovette IJ: **Mitochondrial dating and mixed-support for the "2% rule" in birds.** *Auk* 2004, **121**:1–6.
40. Garcia-Moreno J: **Is there a universal mtDNA clock for birds?** *Journal of Avian Biology* 2004, **35**(6):465–468.
41. Päckert M, Martens J, Tietze DT, Dietzen C, Wink M, Kvist L: **Calibration of a molecular clock in tits (Paridae)—do nucleotide substitution rates of mitochondrial genes deviate from the 2% rule?** *Molecular Phylogenetics and Evolution* 2007, **44**:1–14.
42. Weir JT, Schluter D: **Calibrating the avian molecular clock.** *Molecular Ecology* 2008, **17**(10):2321–2328.
43. Pereira SL, Baker AJ: **A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock.** *Molecular Biology and Evolution* 2006, **23**(9):1731–1740.
44. Thorne J, Kishino H, Painter I: **Estimating the rate of evolution of the rate of molecular evolution.** *Molecular Biology and Evolution* 1998, **15**(12):1647–1657.
45. Felsenstein J: **Phylogenies and the comparative method.** *American Naturalist* 1985, **125**:1–15.
46. Bazin E, Duret L, Penel S, Galtier N: **Polymorphix: a sequence polymorphism database.** *Nucleic Acids Research* 2005, **33**:D481–D484.
47. Nevo E, Beiles A, Ben-Shlomo R: **The evolutionary significance of genetic diversity : Ecological, demographic and life-history correlates..** *Lecture Notes in Biomathematics 53 : Evolutionary Dynamics of Genetic Diversity* 1984.
48. Thorne J, Kishino H: **Divergence time and evolutionary rate estimation with multilocus data.** *Systematic Biology* 2002, **51**(5):689–702.
49. Drummond A, Ho S, Phillips M, Rambaut A: **Relaxed phylogenetics and dating with confidence.** *Plos Biol* 2006, **4**(5):e88–e88.
50. Yang Z, Rannala B: **Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds.** *Molecular Biology and Evolution* 2006, **23**:212–226.
51. Thomas J, Welch J, Woolfit M, Bromham L: **There is no universal molecular clock for invertebrates, but rate variation does not scale with body size.** *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(19):7366–7371.
52. Gissi C, Reyes A, Pesole G, Saccone C: **Lineage-specific evolutionary rate in mammalian mtDNA.** *Molecular Biology and Evolution* 2000, **17**(7):1022–1031.
53. Castresana J: **Cytochrome b phylogeny and the taxonomy of great apes and mammals.** *Molecular Biology and Evolution* 2001, **18**(4):465–471.
54. Gillooly J, Allen A, West G, Brown J: **The rate of dna evolution: effects of body size and temperature on the molecular clock.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**:140–145.
55. Ogburn CE, Austad SN, Holmes DJ, Kiklevich JV, Gollahon K, Rabinovitch PS, Martin GM: **Cultured renal epithelial cells from birds and mice: enhanced resistance of avian cells to oxidative stress and DNA damage.** *J Gerontol A Biol Sci Med Sci* 1998, **53**(4):B287–B292.
56. Herrero A, Barja G: **H2o2 production of heart mitochondria and aging rate are slower in canaries and parakeets than in mice: sites of free radical generation and mechanisms involved.** *Mechanisms of Ageing and Development* 1998, **103**(2):133–146.
57. Barja G: **Free radicals and aging.** *Trends in Neurosciences* 2004, **27**(10):595–600.
58. Frankham R: **Do island populations have less genetic variation than mainland populations?** *Heredity* 1997, **78**:311–327.

59. Spielman D, Brook B, Frankham R: **Most species are not driven to extinction before genetic factors impact them.** *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(42):15261–15264.
60. Popadin K, Polishchuk LV, Mamirova L, Knorre D, Gunbin K: **Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals.** *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**(33):13390–13395.
61. Rand D, Kann L: **Excess amino acid polymorphism in mitochondrial dna: contrasts among genes from drosophila, mice, and humans.** *Molecular Biology and Evolution* 1996, **13**(6):735–748.
62. Stanley SE, Harrison RG: **Cytochrome b evolution in birds and mammals: an evaluation of the avian constraint hypothesis.** *Molecular Biology and Evolution* 1999, **16**(11):1575–1585.
63. Gillespie J: **Is the population size of a species relevant to its evolution?** *Evolution* 2001, **55**(11):2161–2169.
64. Nei M, Graur D: **Extent of protein polymorphism and the neutral mutation theory.** *Evolutionary Biology* 1984, **17**:73–118.
65. Iizuka M, Tachida H, Matsuda H: **A neutral model with fluctuating population size and its effective size.** *Genetics* 2002, **161**:381–388.
66. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Research* 1994, **22**(22):4673–4680.
67. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Systematic Biology* 2003, **52**(5):696–704.
68. Ericson PGP, Anderson CL, Britton T, Elzanowski A, Johansson US, Kallersjö M, Ohlson JI, Parsons TJ, Zuccon D, Mayr G: **Diversification of neoaves: integration of molecular sequence data and fossils.** *Biology Letters* 2006, **2**(4):543–547.
69. Slack KE, Delsuc F, Mclenachan PA, Arnason U, Penny D: **Resolving the root of the avian mitogenomic tree by breaking up long branches.** *Molecular Phylogenetics and Evolution* 2007, **42**:1–13.
70. Rannala B, Yang Z: **Inferring speciation times under an episodic molecular clock.** *Systematic Biology* 2007, **56**(3):453–466.
71. Yang Z: *Computational Molecular Evolution.* Oxford University Press 2007.
72. Fain MG, Houde P: **Parallel radiations in the primary clades of birds.** *Evolution* 2004, **58**(11):2558–2573.
73. Gibb GC, Kardailsky O, Kimball RT, Braun EL, Penny D: **Mitochondrial genomes and avian phylogeny: complex characters and resolvability without explosive radiations.** *Molecular Biology and Evolution* 2007, **24**:269–280.
74. Morgan-Richards M, Trewick SA, Bartosch-Härlid A, Kardailsky O, Phillips MJ, Mclenachan PA, Penny D: **Bird evolution: testing the Metaves clade with six new mitochondrial genomes.** *BMC Evolutionary Biology* 2008, **8**:20–20.
75. Barker FK, Cibois A, Schikler P, Feinstein J, Cracraft J: **Phylogeny and diversification of the largest avian radiation.** *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(30):11040–11045.
76. Dutheil J, Gaillard S, Bazin E, Glemin S, Ranwez V, Galtier N, Belkhir K: **Bio++: a set of C++ libraries for sequence analysis, phylogenetics, molecular evolution and population genetics.** *BMC Bioinformatics* 2006, **7**:188–188.
77. Tajima F: **Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism.** *Genetics* 1989, **123**(3):585–595.
78. Watterson G: **Number of segregating sites in genetic models without recombination.** *Theoretical Population Biology* 1975, **7**(2):256–276.
79. de Magalhaes JP, Costa J, Toussaint O: **Hagr: the human ageing genomic resources.** *Nucleic Acids Research* 2005, **33**:D537–D543.
80. Sokal R, Rohlf F: **Biometry.** 1981.
81. Team RDC: **R: A language and environment for statistical computing.** *R Foundation for Statistical Computing* 2004.

Figures

Figure 1 - Distribution of cytochrome b neutral substitution rate in 1,562 bird species.

Distribution of cytochrome *b* neutral substitution rate in 1,562 bird species.

Figure 2 - Whithin-order distribution of the cytochrome b third codon position substitution rate.

Orders represented by more than 20 species are shown. Values above 0.1 substitution/site/106 years are not shown.

Figure 3 - Distribution of cytochrome b neutral substitution rate in 1,696 mammalian species (Data from [14]) and 1,562 birds species (this study).

Distribution of cytochrome *b* neutral substitution rate in 1,696 mammalian species (Data from [14]) and 1,562 bird species (this study).

Figure 4 - Relationship between substitution rate and life-history traits.

a) Relationship with body mass, b) Relationship with maximum longevity. Open circles are for passerine birds ($n = 88$), close circles for non-passerine birds ($n = 108$). Substitution rates are in unit of substitution per third codon position per million years.

Figure 5 - Relationship between substitution rate and life-history traits.

a) Body mass, b) Maximum longevity. Red dots are for mammals ($n = 500$), blue dots for birds ($n = 196$). Substitution rates are in unit of substitution per third codon position per million years.

Figure 6 - mtDNA synonymous diversity (π_s) vs. direct population size estimate in birds ($n = 28$).

mtDNA synonymous diversity (π_s) vs. direct population size estimate in birds ($n = 28$).

Figure 7 - mtDNA synonymous diversity (π_s) and substitution rate.

Red dots are for mammals ($n = 123$), blue dots for birds ($n = 46$). Substitution rates are in logarithmic scale and in unit of substitution per third codon position per million year and in logarithmic scale.

Figure legend text.

Tables

Table 1 - Effects of life-history variables on mtDNA substitution rate in bird species with and without phylogenetic control.

Model	Without Phylogenetic control				With Phylogenetic control			
	slope	R ²	p1	p2	slope	R ²	p1	p2
Body mass	-0.25	0.39	<0.001		-0.22	0.38	<0.001	
Maximum longevity	-0.66	0.25	<0.001		-0.46	0.19	<0.001	
Body mass + Maximum longevity		0.39	<0.001	0.42		0.38	<0.001	0.89

Table 2 - Effects of life-history variables on mtDNA substitution rate in passerines vs. other bird species.

This table is attached as separate file. This table is similar to table 1 but summary the results within passerines and non-passerines birds.

Table 2 - Effects of life-history variables on mtDNA substitution rate in passerines vs. other bird species. Effects of life-history variables and bird (n = 196) vs. mammal (n = 500) status (referred as Class) on mtDNA substitution rate..

(Body mass + Class)			
Variable	Effect	t value	p
Body mass	-0.25	-8.9	<0.001
Class (Mammals)	0.22	3.1	0.002
Body mass : Class (Mammals)	0.13	4.4	<0.001
(Maximum longevity + Class)			
Variable	Effect	t value	p
Maximum longevity	-0.65	-9.4	<0.001
Class (Mammals)	0.31	2.5	0.01
Maximum longevity : Class (Mammals)	0.04	0.5	0.65

Additional Files

Additional file 1 — Figure S1

File format: PDF

Distribution of bird maximum longevity. The maximum longevity is log transformed. Black: whole dataset; shading yellow: passerines birds (n = 88); shading blue: no passerines birds (n = 108)

Additional file 2 — Figure S2

File format: PDF Neutrality index (NI) distribution in birds (n = 81) and mammals (n = 75). NI values greater than 20 were forced to 20 for clarity.

Additional file 3 — Figure S3

File format: PDF a) mtDNA synonymous diversity (π_s) in birds (n = 46) and mammals (n = 123). The median, quartiles, and extrema of the distribution of π_s are given. b) Residuals of the relationship between mtDNA synonymous diversity (π_s) and third-codon position substitution rate in birds and mammals.

Additional file 4 — TableS1

File format: PDF Fossil calibration dates used in this study (in million years).

Additional file 5 — TableS2

File format: Text (csv) GenBank accession numbers, taxonomy, mitochondrial substitution rate and life-history traits of the 1571 species used in the substitution rate analysis.

Additional file 6 — TableS3

File format: PDF Effects of life-history variables on mtDNA substitution rate in bird species according the different combination of topologies and programs.

Additional file 7 — TableS4

File format: PDF Effects of life-history variables on mtDNA substitution rate in passerines versus other bird species according the different combination of topologies and programs.

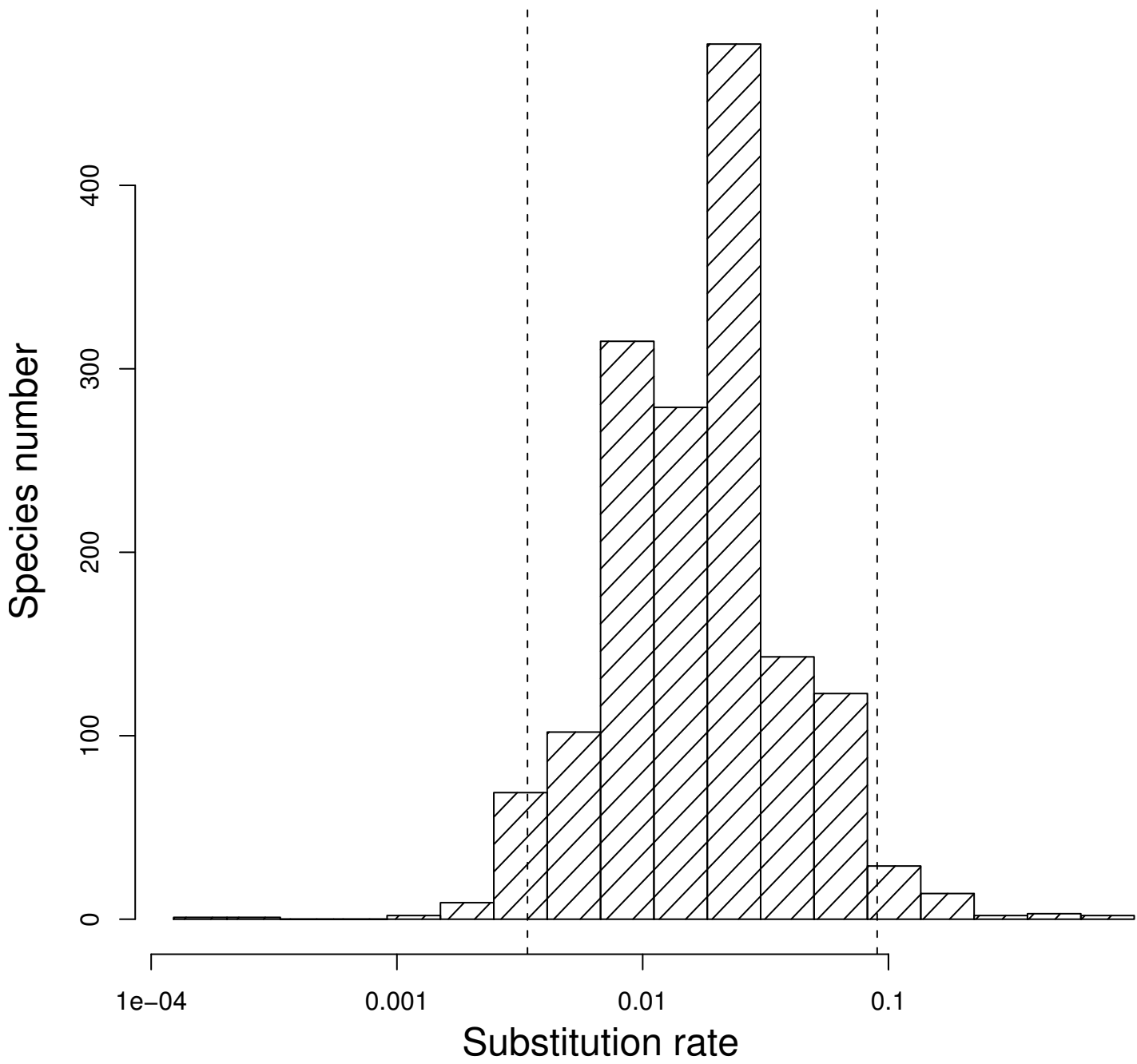


Figure 1

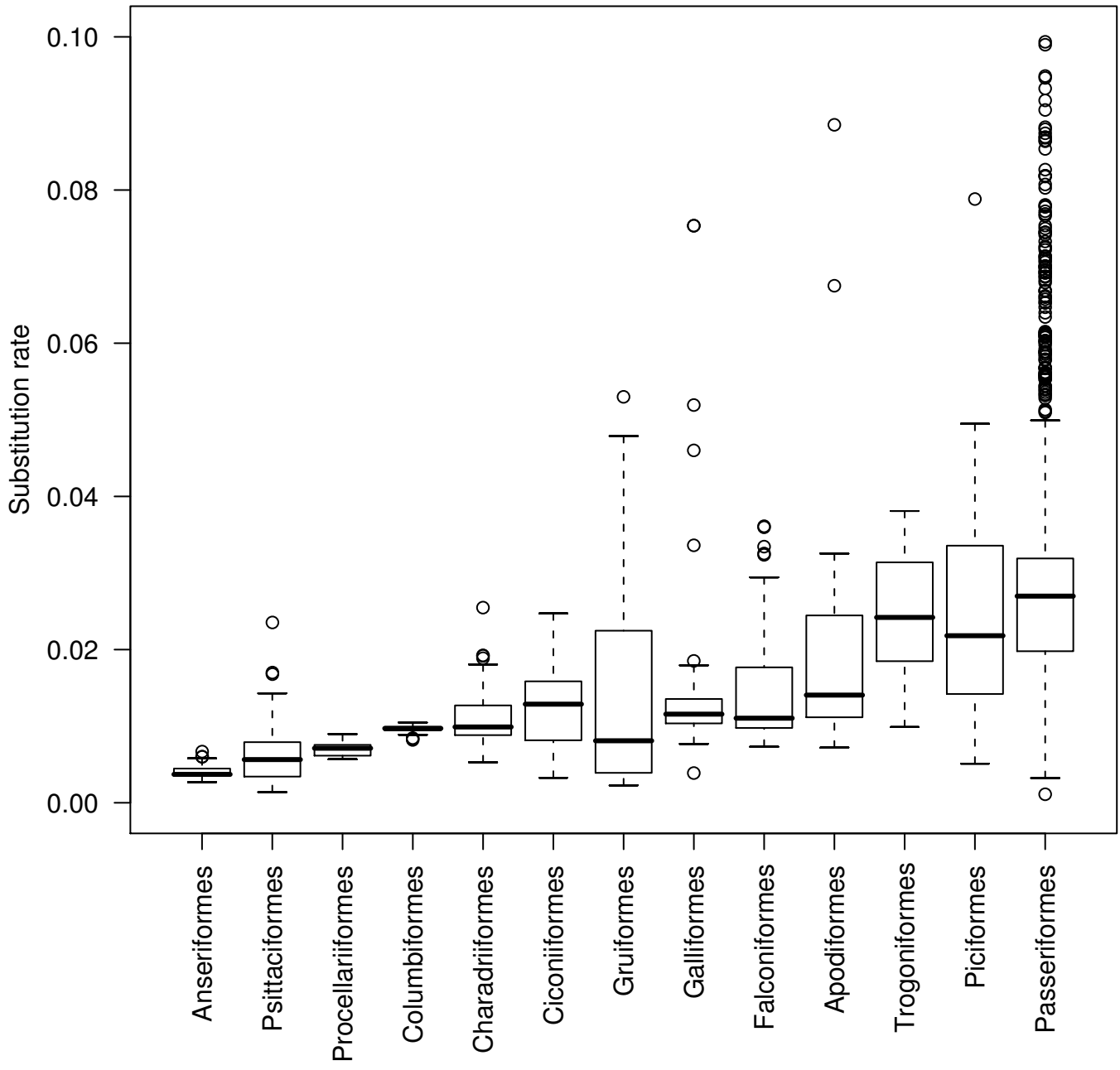


Figure 2

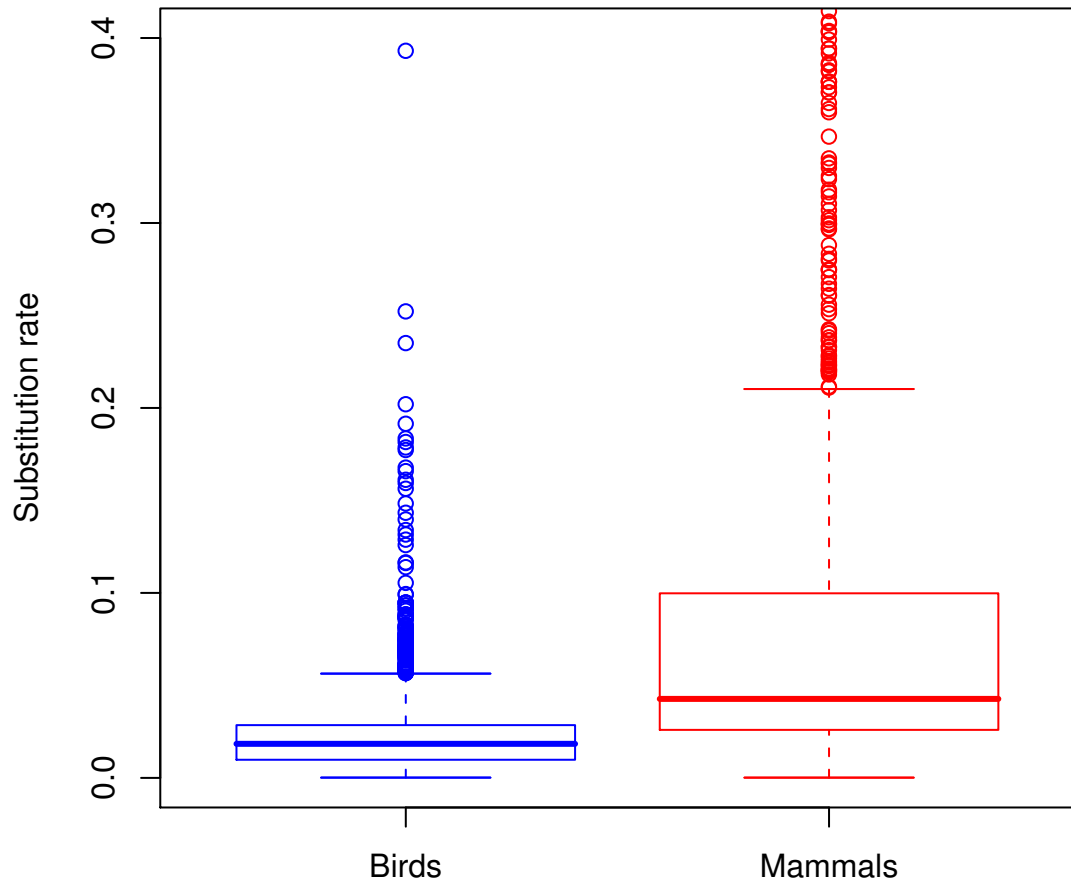


Figure 3

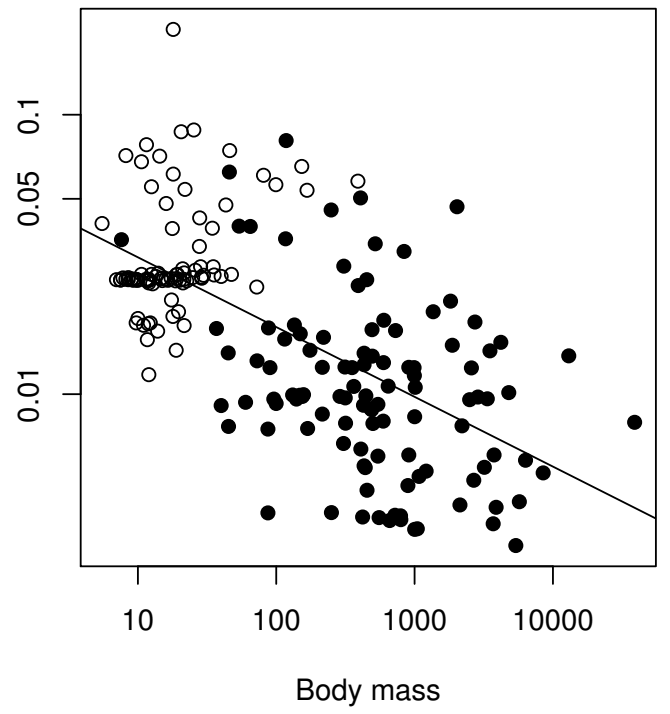
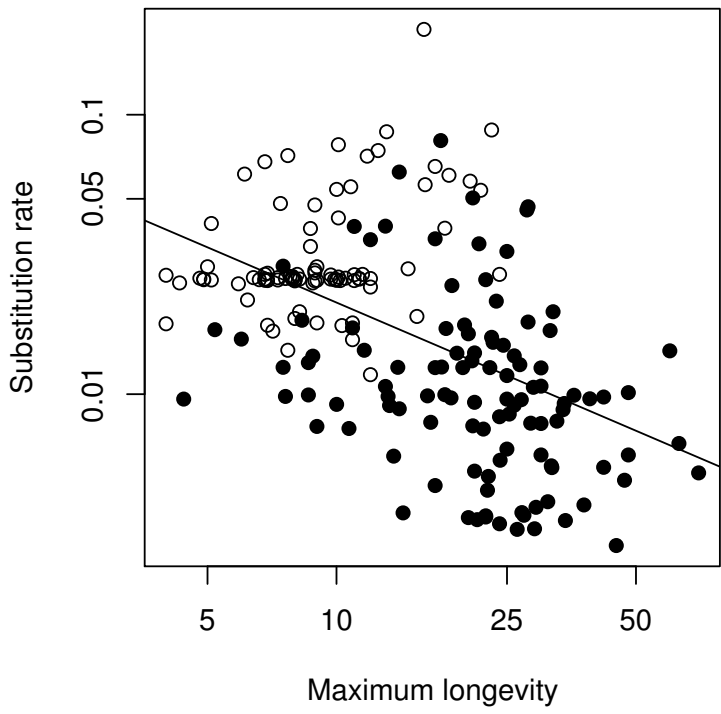


Figure 4

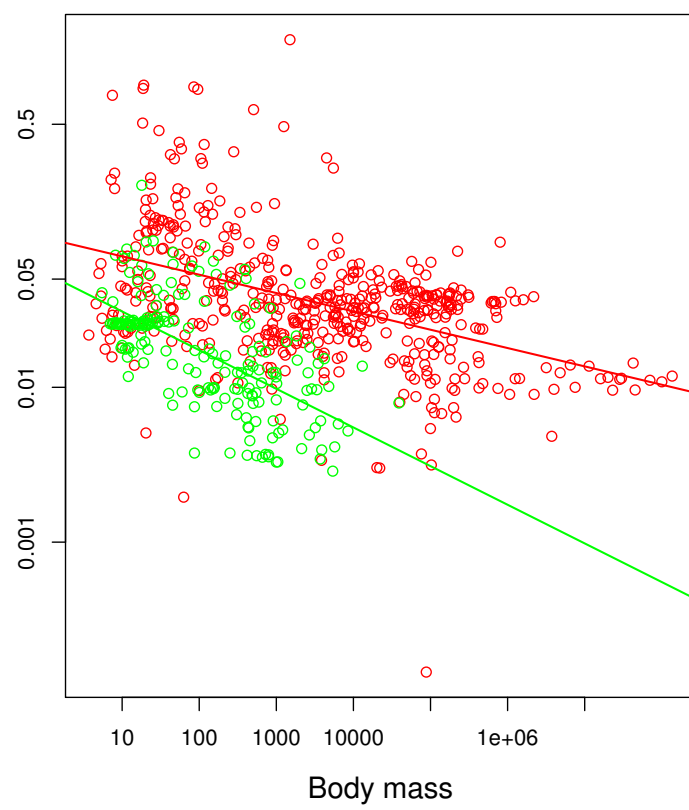
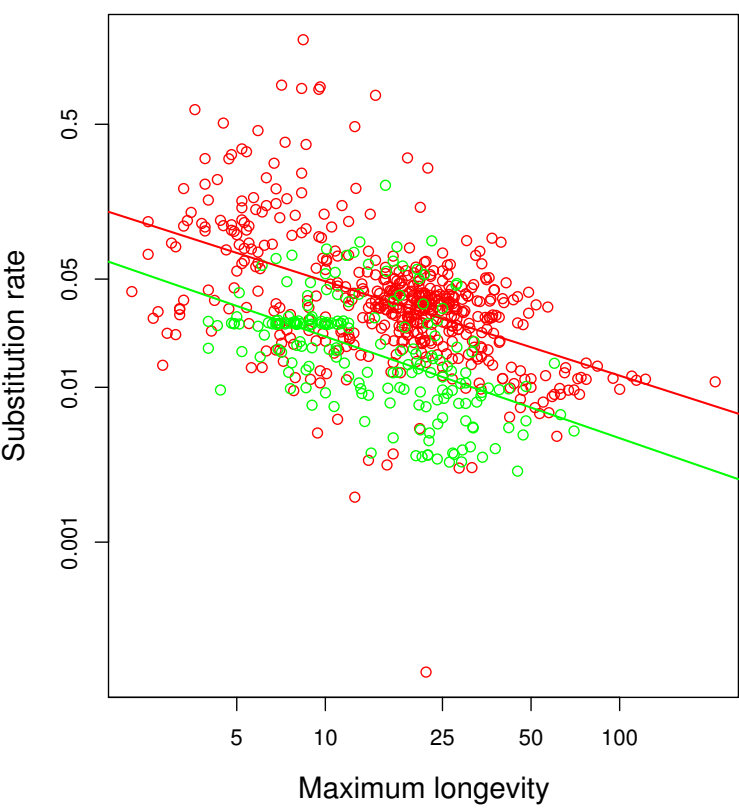


Figure 5

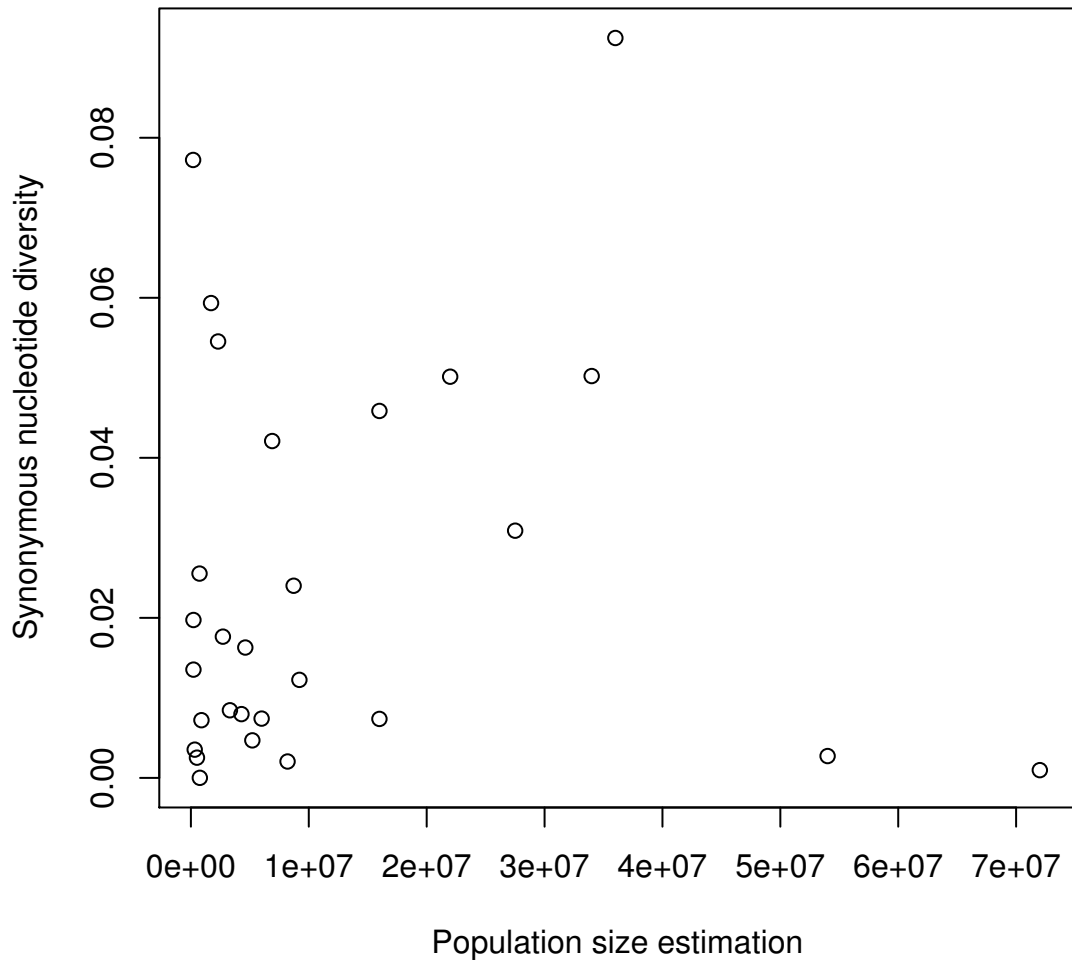


Figure 6

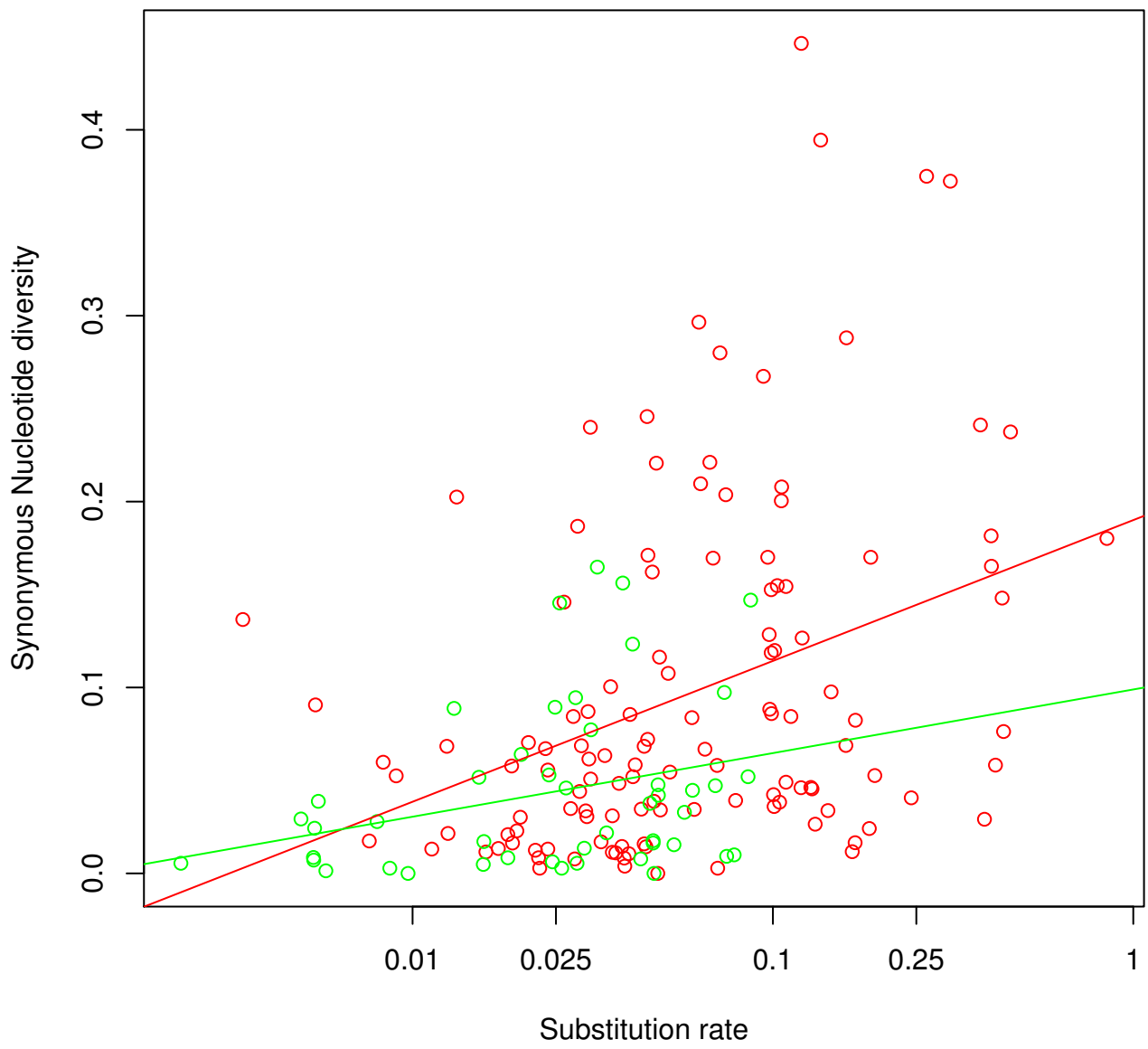


Figure 7

Additional files provided with this submission:

Additional file 1: figure.s1.pdf, 17K

<http://www.biomedcentral.com/imedia/2932935422221334/supp1.pdf>

Additional file 2: figure.s2.pdf, 3K

<http://www.biomedcentral.com/imedia/9326568012221339/supp2.pdf>

Additional file 3: figure.s3.pdf, 5K

<http://www.biomedcentral.com/imedia/1073430733222133/supp3.pdf>

Additional file 4: table.s1.pdf, 434K

<http://www.biomedcentral.com/imedia/2067286521222133/supp4.pdf>

Additional file 5: tables2.csv, 141K

<http://www.biomedcentral.com/imedia/7303897102221336/supp5.csv>

Additional file 6: table_s3.pdf, 532K

<http://www.biomedcentral.com/imedia/1331814133222133/supp6.pdf>

Additional file 7: table_s4.pdf, 533K

<http://www.biomedcentral.com/imedia/1707821832222134/supp7.pdf>

Additional file 8: table_2.pdf, 532K

<http://www.biomedcentral.com/imedia/1087426016222339/supp8.pdf>

Résumé : L'origine et l'évolution du génome mitochondrial sont fascinantes par de nombreux aspects. Aujourd'hui, ce génome représente moins de 1% de l'ensemble de l'ADN de la majorité des animaux mais il héberge quelques uns des gènes les plus essentiels aux fonctions métaboliques. L'hypermutabilité mitochondriale est, chez les animaux, l'une de ses caractéristiques les plus singulières. L'étude du déterminisme du taux de mutation mitochondrial est le premier objectif de cette thèse. Par une approche phylogénétique précise, nous avons obtenu le taux de mutation de plus d'un millier d'espèces d'oiseaux et de mammifères. Nous avons révélé une gamme de variation très importante entre espèces, cette variation allant jusqu'à couvrir deux ordres de grandeur chez les mammifères. Grâce à une étude intra-classe et par la comparaison entre les oiseaux et les mammifères, nous avons montré que cette variation pouvait être en lien avec la longévité des organismes à travers une sélection pour une réduction du taux de mutation mitochondrial chez les espèces longévives. La deuxième partie de cette thèse concerne l'action de la sélection naturelle et de la dérive génétique sur l'évolution de l'ADN mitochondrial (ADNmt). Prenant comme point de départ les récentes preuves de sélection positive dans l'évolution de l'ADNmt (notamment chez les invertébrés), nous détectons, *a contrario*, essentiellement les traces de la sélection purificatrice chez les oiseaux et les mammifères. De manière plus surprenante, nous avons montré que la taille des populations n'est apparemment pas liée au niveau de polymorphisme mitochondrial des espèces, mais influencé, en revanche, l'efficacité de la sélection, mesurée au travers de la quantité de mutations faiblement délétères qui se fixent au cours de la divergence. Nous proposons qu'une forte stochasticité temporelle de la taille des populations pourrait expliquer ce résultat. L'ensemble de ces résultats participent à la compréhension de l'évolution d'un génome singulier, et a également des implications sur les utilisateurs, particulièrement nombreux, de l'ADN mitochondrial en temps que marqueur moléculaire.

Mots-clés : Mutation, diversité génétique, génome mitochondrial, longévité, oiseaux, mammifères, taille des populations, sélection, évolution moléculaire, phylogénie.

Mitochondrial DNA evolutionary dynamics of birds and mammals : Mutation, Selection and Population size.

Abstract : The origin and evolution of mitochondrial genome is fascinating. Currently, it makes up less than 1% of the whole organism genome, but contains some of the most important genes. A particularly intriguing feature of the animal mitochondrial genome is its hypermutability. The first goal of this work is to progress in our understanding of the determinism of mitochondrial DNA (mtDNA) substitution rate variations by distinguishing between two classical hypotheses of evolutionary biology –the generation time hypothesis and the metabolic rate hypothesis– and an other hypothesis that comes from biomedecine, namely the longevity hypothesis. Using a phylogenetic approach, we obtained lineage-specific mitochondrial mutation rates across more than one thousand bird and mammalian species. This analysis reveals an unexpectedly high level of mitochondrial mutation rate variation between lineages. The bird/mammal comparison and a within-class analysis suggest that this variation could be linked to species longevity through a (direct or indirect) selective pressure reducing the mitochondrial mutation rate in long-lived species. In the second part of this work, we address the impact of natural selection and genetic drift on mtDNA. Recent evidence of positive selection acting on mtDNA (mostly in invertebrates) was used as a starting point. We showed that, contrary to invertebrates species, bird and mammal mtDNA evolution is mainly under purifying selection. Surprisingly, even in the absence of positive selection, population size variations have no effect on mtDNA genetic diversity, but influence the rate of non-synonymous substitutions. This result could be explained by strong stochasticity of population sizes. All these results contribute to increase our understanding of an unusually evolving genome, and also have implications for the numerous users of mtDNA as a tool to reconstruct population and species history.

Keywords : Mutation, genetic diversity, mitochondrial genome, longevity, birds, mammals, population size, selection, molecular evolution, phylogeny.