



**HAL**  
open science

# Biologie Moléculaire et Phylogénèse des Lémuriformes (Primates de Madagascar)

Daniel Montagnon

► **To cite this version:**

Daniel Montagnon. Biologie Moléculaire et Phylogénèse des Lémuriformes (Primates de Madagascar). Biochimie [q-bio.BM]. Université Louis Pasteur - Strasbourg I, 2001. Français. NNT : . tel-00002829

**HAL Id: tel-00002829**

**<https://theses.hal.science/tel-00002829>**

Submitted on 16 May 2003

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**ULP Université Louis Pasteur STRASBOURG I**

---

**THESE**

Présentée en vue de l'obtention du grade de

**Docteur de l'Université Louis Pasteur  
STRASBOURG I**

à

L'U.F.R. des Sciences de la Vie et de la Terre

Par

**Daniel MONTAGNON**

**Biologie Moléculaire et Phylogénèse des Lémuriformes  
(Primates de Madagascar)**

Directeur de Thèse : Pr. Yves RUMPLER

Soutenue le 6 Septembre 2001

Devant le jury d'examen composé de

**Pr. Y. RUMPLER**

**Pr. P. DARLU**

**Pr. R. ALBIGNAC**

**Dr. A. HANAUER**

A M. le Professeur **Yves RUMPLER**, je vous remercie pour votre aide et vos conseils durant toutes ces années passées à l'Institut d'Embryologie et plus particulièrement au cours de la préparation de cette thèse.

A M. le Professeur **Pierre DARLU**, trouvez ici toute ma reconnaissance pour avoir bien voulu juger ce travail.

A M. le Professeur **Roland ALBIGNAC** et à M. le Docteur **André HANAUER**, je vous remercie d'avoir bien voulu juger ce travail de thèse.

Je tiens à remercier tout particulièrement le Docteur **Iary B. RAVAOARIMANANA** pour son aide et son soutien ainsi que pour tous les travaux que nous avons pu réaliser en collaboration.

Je remercie aussi le Professeur **Berthe RAKOTOSAMIMANANA** grâce à qui j'ai pu obtenir les prélèvements de sub-fossile de Lémuriens.

Je remercie le Professeur **Pierre CHAMBON**, Directeur de l'IGBMC, ainsi que toutes les personnes de l'Institut qui m'ont apporté leur aide, et plus particulièrement **Serge VICAIRE** (service de séquençage) pour sa patience et sa gentillesse.

Je remercie également tout le personnel de l'Institut d'Embryologie pour son aide et ses conseils.

A mes parents.

A Nicolas et Thomas.

# SOMMAIRE

<b>SOMMAIRE</b> .....	<b>1</b>
<b>INTRODUCTION</b> .....	<b>2</b>
<b>TECHNIQUES D'ETUDE</b> .....	<b>3</b>
GENERALITES .....	3
RECONSTRUCTIONS PHYLOGENETIQUES .....	6
MODALITES D'UTILISATION .....	29
PROBLEMES PRINCIPAUX .....	30
METHODES COMPLEMENTAIRES .....	32
NIVEAU D'INFORMATION PHYLOGENETIQUE .....	33
DATATION DES SEQUENCES .....	34
RECONSTRUCTION DE SEQUENCES ANCESTRALES.....	35
EVOLUTION GENERALE DES SEQUENCES.....	36
<b>L'ADN MITOCHONDRIAL</b> .....	<b>37</b>
GENERALITES .....	37
CARACTERISTIQUES .....	38
GENES ETUDIES .....	38
<b>LES LEMURIFORMES ACTUELS ET SUB-FOSSILES</b> .....	<b>41</b>
<b>MATERIEL ET METHODES</b> .....	<b>52</b>
EXTRACTION DE L'ADN .....	52
AMPLIFICATION DES SEQUENCES MITOCHONDRIALES.....	52
ANALYSE DES PRODUITS D'AMPLIFICATION .....	53
SEQUENÇAGE DES PRODUITS D'AMPLIFICATION.....	53
PROGRAMMES INFORMATIQUES UTILISÉS .....	54
SEQUENCES UTILISEES.....	55
<b>RESULTATS</b> .....	<b>60</b>
CARACTERISTIQUES GENERALES DES SEQUENCES .....	60
TEST DE LA MONOPHYLIE DES LEMURIFORMES .....	74
DETERMINATION DE LA SEQUENCE OUTGROUPE OPTIMALE.....	78
TRAITEMENTS SUR LES SEQUENCES CYTOCHROME B .....	79
TRAITEMENTS SUR LES SEQUENCES 12S rRNA.....	124
TRAITEMENTS SUR LES DONNEES MIXTES .....	143
LES RELATIONS DMA – TUP .....	154
LES RELATIONS IND – MGA.....	157
<b>DISCUSSION</b> .....	<b>160</b>
<b>CONCLUSIONS</b> .....	<b>169</b>
<b>BIBLIOGRAPHIE</b> .....	<b>171</b>
<b>TABLE DES ILLUSTRATIONS</b> .....	<b>189</b>
<b>TABLE DES TABLEAUX</b> .....	<b>191</b>
<b>TABLE DES MATIERES</b> .....	<b>194</b>
<b>INDEX</b> .....	<b>197</b>

# INTRODUCTION

La faune de Primates de Madagascar semble être le produit d'une colonisation unique par un Primate ancestral venant d'Afrique. Cette colonisation aurait eu lieu à une époque à laquelle il n'existait plus aucune liaison terrestre entre le continent et l'île. Il y a en effet au moins 150 millions d'années une profonde faille océanique sépara Madagascar du continent Africain (267). Ceci impliquerait que cet ancêtre ait dû emprunter des radeaux de végétations (362). Mais on ne peut pas tout à fait exclure une autre hypothèse considérant que l'ancêtre était déjà présent sur l'île au moment où celle-ci s'est séparée du continent Africain.

Un certain nombre d'études ont fixé le début de l'évolution des Lemuriformes, soit à la fin du Paléocène (291), soit à la fin de l'Eocène (22); il existe une différence d'environ 20 millions d'années entre ces deux hypothèses. Les travaux de Yoder et coll. (387) ont permis de montrer que les Primates malgaches avaient divergés des Lorisiformes, tant en Afrique qu'en Asie, il y a à peu près 62 millions d'années. Mais il est toutefois difficile de déterminer cette date avec précision.

Au cours des 2 000 dernières années on peut considérer qu'environ 25 espèces de Vertébrés, de taille relativement grande, ont disparu de Madagascar, et la plupart de ceux-ci étaient des Primates (39).

La première description publiée d'un Lémurien malgache est due à Flacourt (100) qui, dans son 'Histoire de la grande isle Madagascar' note:

"Tretretrete ou Tratratratra, c'est un animal grand comme un veau de deux ans qui a le teste rond, & une face d'homme: les pieds de devant comme un singe, & les pieds de derrière aussi. Il a un poil frisé, la queue courte & les oreilles comme celles d'un homme. Il ressemble au Touachi décrit par Ambroise Paré. Il s'en est vu un proche l'estang de Lipomami, aux environs duquel est son repaire. C'est un animal fort solitaire, les gens du país en ont grand peur & s'enfuient de luy comme luy d'eux."

Cet animal existait, du temps de Flacourt, dans la région de Fort-Dauphin et, même si l'on peut considérer qu'il y ait une certaine exagération quant à la taille, il semble bien que ce soit là l'un des derniers spécimens de *Megaladapis madagascariensis* dont un crâne fut présenté, pour la première fois, à la Royal Society par Forsyth-Major (101) le 15 Juin 1893. Depuis cette date aucun autre spécimen vivant de cet animal n'a été trouvé.

Au cours des trente dernières années l'apparition et le développement des outils de biologie moléculaire, de même que l'utilisation d'outils mathématiques et statistiques de plus en plus performants, ont permis un important essor des phylogénèses moléculaires. Il a ainsi été possible de préciser, ou de définir, un grand nombre de relations phylogénétiques, entre autres, dans le cas des Primates malgaches.

# TECHNIQUES D'ETUDE

## *Généralités*

### **Etapes d'un traitement phylogénétique**

La réalisation d'un traitement phylogénétique complet peut être divisée en quatre étapes principales:

- Définition des données à traiter: il s'agit ici de sélectionner des séquences nucléotidiques correspondant à des gènes complets ou partiels, ou des séquences non codantes, ou encore des séquences protéiques.
- Alignement des séquences sélectionnées: cette étape est réalisée à l'aide de logiciels adaptés, ou dans certaines conditions elle peut être effectuée manuellement.
- Application d'une ou plusieurs méthodes de reconstruction: ces méthodes sont généralement choisies en fonction des caractéristiques des séquences traitées.
- Evaluation statistique de l'arbre, ou des arbres obtenus: des comparaisons peuvent également être effectuées avec les phylogénèses obtenues à partir de données autres que des séquences, par exemple des données morphologiques.

### **Données utilisables**

Deux hypothèses doivent être posées quant aux caractères qui vont être utilisés pour les reconstructions phylogénétiques:

- Les caractères doivent être indépendants.
- Les caractères doivent être homologues.

Dans la grande majorité des cas les caractères utilisés sont de type quantitatif, c'est à dire qu'ils peuvent présenter deux ou plusieurs états, encore appelés valeurs discrètes. Lorsque plus de deux états peuvent survenir ceux-ci peuvent être ordonnés ou non.

Dans le cas des séquences nucléotidiques les caractères utilisés sont représentés par les nucléotides. Généralement chacun d'eux peut présenter quatre états, non ordonnés, qui sont en relation par le biais des transitions et des transversions (**Figure 1**).

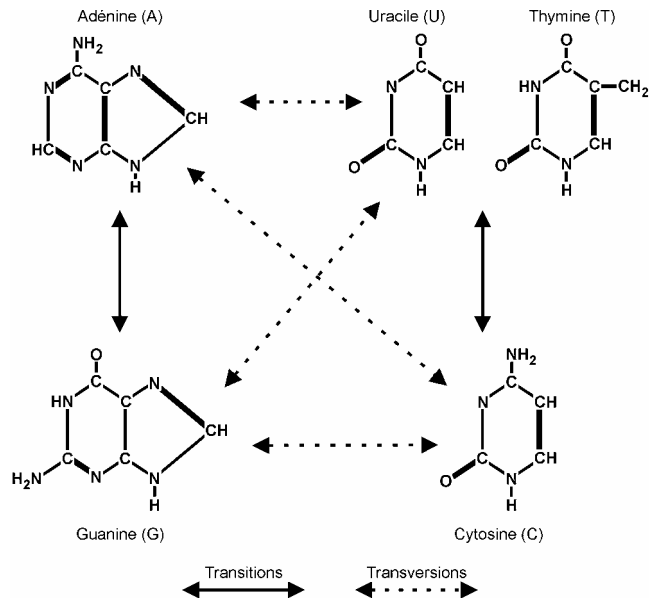


Figure 1 : Relations entre transitions et transversions

## Alignement des séquences

Pour un grand nombre de séquences nucléotidiques, l'alignement est un préalable essentiel pour les analyses phylogénétiques. L'algorithme généralement utilisé pour calculer cet alignement est une extension de la procédure décrite par Needleman et Wunsch (236) pour deux séquences protéiques. Mais lorsque le travail doit porter sur plus de deux séquences, qu'elles soient protéiques ou nucléotidiques, il existe trois sources principales d'ambiguïtés:

- La première est due à la nécessité d'attribuer des poids relatifs aux différents événements permettant le passage d'une séquence à l'autre et à la création de l'alignement. Ces événements se traduisent en termes d'insertions, de délétions, ou de changements de bases. En pratique un poids est attribué arbitrairement pour l'alignement généré à chaque type d'événement, ce qui permet de calculer un score. Plusieurs alignements sont testés et, en définitive, celui présentant le score optimal (généralement le plus faible) est retenu.
- Dans le cas des alignements multiples (comportant plus de deux séquences) le résultat final est obtenu soit par l'utilisation d'une méthode dite exacte, soit en utilisant une recherche de type heuristique. On peut montrer que le résultat final est dépendant de l'ordre dans lequel les séquences sont traitées. En effet l'algorithme commence par aligner deux séquences, puis la troisième est alignée en fonction des deux précédentes, et ainsi de suite. Il est alors possible, en utilisant le même ensemble de séquences, les mêmes paramètres d'alignement (poids des événements), mais des ordres d'introduction des séquences différents, d'obtenir plusieurs alignements ayant un score optimal. Une solution envisageable pour obtenir le meilleur alignement possible est alors de 'guider' le traitement à partir d'un arbre phylogénétique, donc en fixant les relations entre les différentes séquences.

- La dernière source d'ambiguïté réside dans le fait que l'algorithme utilisé peut suivre plusieurs voies d'alignement, ce qui, comme dans le cas précédent, peut conduire à l'obtention de plusieurs alignements présentant le score optimal. Une telle situation peut se présenter même avec l'utilisation d'un arbre phylogénétique comme guide.

Lorsque plusieurs alignements sont trouvés à l'issue de la procédure, le meilleur pourra être déterminé en considérant le fait que ce doit être celui fournissant le cladogramme le plus parcimonieux.

## **Horloge moléculaire et temps de divergence**

L'hypothèse de l'existence d'une horloge moléculaire pour l'évolution des séquences (nucléotidiques ou protéiques) considère que le taux d'évolution moléculaire est sensiblement constant pour toutes les lignées étudiées (393). Après avoir été fortement contestée, même récemment (83, 139, 162), l'existence de cette horloge est maintenant généralement admise. Elle a tout d'abord fourni un fort support pour la théorie de la neutralité des mutations, postulant que la majorité des changements moléculaires survenant au cours de l'évolution sont le fait de mutations neutres, ou pratiquement neutres (166, 168). En effet Miyata et Yasunaga (224) ont montré qu'il existait des taux quasi équivalents pour les mutations dites synonymes (n'induisant pas de modification de l'acide aminé codé), tant dans les différents gènes, que dans les différentes lignées. Toutefois des études plus récentes (33, 189, 372) ont montré que les taux de substitution étaient significativement plus élevés chez les Rongeurs que chez les Primates et les Artiodactyles.

L'utilisation de cette horloge moléculaire pour calibrer les divergences entre lignées a été suggérée par Margoliash (207) et Zuckerhandl et Pauling (393). Mais pour Hillis et coll. (138), si l'on considère une horloge parfaite, c'est à dire suivant un processus de Poisson, l'existence de variations aléatoires fait que l'horloge pouvant être définie devient très peu précise lorsqu'elle est utilisée pour l'estimation des divergences.

Chez les Mammifères il a toutefois été possible de définir une horloge globale (78), mais également des horloges dites locales (19, 248) rendant compte des caractéristiques d'évolution pour seulement une, ou quelques lignées.

Les travaux de Bonner et coll. (28) dans lesquels est utilisée une technique d'hybridation de l'ADN chez les Lemur, Loris, Tarsiers et Simiens, ont montré que ces groupes avaient divergé au début de l'évolution des Primates, il y a environ 60 millions d'années et que les taux de substitution nucléotidique étaient significativement plus lents chez les Lemurs que pour les autres groupes. Ces données ont été confirmées par Harris et coll. (127) dans son travail sur les gènes de la  $\beta$ -globine chez les Lemurs et les Primates supérieurs; de même que par Li et coll. (188). D'autre part Gu (126), en utilisant deux points de calibration internes, dont une date de divergence entre Mammifères et Oiseaux estimée à partir de fossiles, a pu montrer que les Métazoaires avaient divergé il y a environ 600 millions d'années.

Ce problème des temps de divergence a été abordé par de nombreux auteurs, même lorsqu'il n'était pas possible de prendre en compte l'existence d'une horloge moléculaire (146, 177, 268, 281, 346, 351). Par contre dans l'approche proposée par Takezaki et coll. (335), l'horloge moléculaire est testée pour l'ensemble des séquences et celles qui ne sont pas compatibles avec le modèle (avec un degré de significabilité supérieur à 1 %) sont éliminées de l'analyse. Les séquences restantes sont alors utilisées pour la reconstruction d'un arbre phylogénétique, en assumant un taux de substitution



constant. L'arbre obtenu (arbre linéarisé) peut alors être utilisé afin d'estimer les temps de divergence pour n'importe laquelle des paires de séquences. Les taux de substitution peuvent aussi être estimés à partir d'autres sources telles que des fossiles ou des dates géologiques.

Le calcul d'une date de divergence ne peut être effectué que si l'on connaît le contexte géologique et environnemental de cet événement. Généralement les données génétiques peuvent être utilisées dans ce but (18, 24, 25, 60, 132, 142, 257, 270, 355). Les estimations de divergences obtenues à partir de ces données génétiques donnent très souvent des valeurs plus anciennes que celles obtenues à partir des fossiles. Les contraintes sélectives, les mécanismes de réparation de l'ADN, ainsi que la taille effective des populations sont des paramètres affectant l'utilisation et le pouvoir d'un marqueur génétique, ce qui peut être une explication pour les différences observées entre datations effectuées à partir de séquences nucléotidiques, et celles effectuées à partir de séquences protéiques.

Le rôle de l'analyse statistique ne doit pas être négligé dans ce type d'étude. En effet l'utilisation de modèles trop simples peut provoquer l'apparition de biais dans l'estimation des dates de spéciation (379).

Il existe de nombreuses méthodes permettant de tester la constance des taux d'évolution. Parmi celles-ci on trouve les méthodes 'tree-species' ou 'species-group' (96, 185, 233, 329, 372), 'least squares' (90, 352), et 'Maximum Likelihood' (92).

## ***Reconstructions phylogénétiques***

Les reconstructions phylogénétiques sont utilisées pour la représentation des relations historiques existant entre différents 'objets biologiques' qui peuvent être, par exemple, des séquences (nucléotidiques ou protéiques) ou des espèces.

### **Modèles de substitution et distances génétiques**

L'estimation des distances génétiques est généralement basée sur le nombre de substitutions affectant les nucléotides, ou acides aminés, présents dans les deux séquences à analyser. Chacune de ces différences peut être la traduction d'une modification unique, soit d'une série plus ou moins importante de modifications. A l'heure actuelle il n'est pas possible de déterminer laquelle de ces deux hypothèses représente le cas réel. De plus il faut considérer les cas où, à la suite de plusieurs modifications affectant un même site, le caractère modifié retrouve l'état dans lequel il se trouvait dans la séquence initiale, ou ancestrale. Afin d'évaluer un tel phénomène il faudrait pouvoir disposer de tous les intermédiaires entre séquence ancestrale et actuelle, ce qui est très rarement le cas.

Au cours de l'évolution la substitution d'un nucléotide peut être divisée en deux étapes principales:

- Une mutation qui n'est pas corrigée par le système normal de réparation de l'ADN.
- L'intervention de forces de sélection et d'effets de dérive génétique aléatoire faisant que le nouvel allèle va se trouver fixé dans la population.

En fait le taux de substitution ou de mutation observé est le résultat d'un processus multifactoriel faisant intervenir la sensibilité aux mutagènes, la fiabilité de l'ADN polymérase et l'efficacité du système de réparation de l'ADN.

L'étude des taux de substitution est importante afin de comprendre la dynamique d'évolution des séquences (106, 168, 247). Les premières études effectuées dans ce domaine ont concerné les protéines (68), elles ont permis de montrer que les taux de substitution des acides aminés variaient considérablement en fonction de la protéine analysée (26, 187, 370). De même, pour les séquences nucléotidiques, les modalités de substitution à différents sites ne sont pas identiques, et de plus elles ne sont pas indépendantes.

Un autre estimateur fréquemment utilisé est la valeur du rapport 'Transitions/Transversions' (Ti/Tv). Ce calcul peut être effectué en utilisant les proportions de sites présentant des transversions (Q) ou des transitions (P), pour les deux séquences étudiées. Toutefois avec cette approche le principal problème est l'existence d'une corrélation entre la distance génétique et la valeur observée du rapport Ti/Tv. Cette corrélation fait qu'à de faibles taux de divergence le rapport apparaît élevé, alors qu'il apparaît bas pour de fortes divergences.

Dans un travail portant sur trois groupes d'Euthériens Yang et Yoder (382) montrent que la fréquence des transitions présente un pic pour des distances génétiques de l'ordre de 15 %; au delà les valeurs de transitions et transversions tendent à l'équilibre, et à partir d'une divergence de 20 % les deux types d'évènements présentent une fréquence égale. Un autre phénomène devant être pris en considération est la saturation des transitions à de forts degrés de divergence génétique (34, 228). De ce fait il semble que l'estimation des processus de substitution soit meilleure lorsqu'elle est effectuée sur des séquences très proches, que lorsqu'elle prend en compte des séquences plus divergentes (266). De plus, à de faibles niveaux de divergence, si le nombre de sites analysés est faible et si il existe un important biais pour les transitions, la variance du rapport Ti/Tv est généralement élevée (360).

On admet généralement que les substitutions nucléotidiques suivent un processus de Markov homogène, défini par une matrice de taux (appelée Q) et dont les éléments représentent les taux instantanés de substitution des quatre nucléotides. Ce modèle, encore appelé REV, a été introduit dans l'analyse des séquences nucléotidiques par Tavaré (345).

Les méthodes qui peuvent être utilisées pour l'estimation des distances génétiques sont fonction des modèles de substitution considérés (112, 277, 298). Comme des sites différents peuvent présenter des valeurs différentes du rapport Ti/Tv (366), les méthodes autorisant une correction pour les substitutions multiples fournissent généralement une meilleure estimation de ce rapport (156).

## **p-distances**

Cette méthode est la plus simple de toutes. Elle consiste à déterminer la proportion de sites pour lesquels il existe une différence entre les deux séquences. On obtient ainsi:

$$p = \frac{n_d}{n}$$

avec:  $n_d$  = nombre de différences observées et  $n$  = nombre total de sites étudiés

Lorsque la valeur de  $p$  est faible (inférieure à 0,1) on peut considérer qu'elle correspond sensiblement au nombre de substitutions par site. Dans ce cas son utilisation pour les reconstructions phylogénétiques fournit des résultats similaires à ceux qui pourraient être obtenus en utilisant des méthodes plus sophistiquées.

Si, pour toutes les lignées étudiées, les taux de substitution sont à peu près identiques, l'utilisation des p-distances permet une meilleure évaluation de la topologie correcte que d'autres méthodes, et ce du fait de la faible variance associée à ces p-distances (278, 286, 307, 331). Par contre cette méthode n'est pas valable lorsque l'on souhaite estimer la divergence entre deux séquences.

Sous ce modèle il est aussi possible d'estimer les proportions de sites auxquels se produisent des transitions (P) et ceux auxquels se produisent des transversions (Q). Ces estimations sont données par:

$$P = \frac{n_s}{n} \quad \text{et} \quad Q = \frac{n_v}{n}$$

avec:  $n_s + n_v = n_d$

On peut également calculer le rapport (R) de ces deux estimations:

$$R = \frac{P}{Q}$$

## **Modèle JC69**

Dans ce modèle défini par Jukes et Cantor (157) les taux de substitutions nucléotidiques sont supposés être identiques pour toutes les paires possibles de nucléotides. Ce modèle est en fait un cas particulier du modèle général de substitution défini pour l'ADN. Le nombre de substitutions ( $d$ ) entre deux séquences est alors estimé par une méthode de maximum de vraisemblance (Maximum Likelihood):

$$d = \frac{-3 \log_e \left(1 - \frac{4p}{3}\right)}{4}$$

Cette estimation n'est possible que si la p-distance présente une valeur inférieure à 0,75; dans ce cas elle représente alors une bonne estimation du nombre réel de substitutions si:

- ✓ Chacun des quatre nucléotides présente une fréquence voisine de 0,25.
- ✓ Il n'existe aucun biais dans le rapport Ti/Tv; celui-ci devant avoir une valeur voisine de 0,5.
- ✓ Les distances estimées (d) sont inférieures à 1,0.

	A	T	C	G
A	-	$\lambda$	$\lambda$	$\lambda$
T	$\lambda$	-	$\lambda$	$\lambda$
C	$\lambda$	$\lambda$	-	$\lambda$
G	$\lambda$	$\lambda$	$\lambda$	-

Lorsque le nombre de nucléotides analysés est faible (inférieur à 100) cette méthode tend à surestimer le nombre réel de substitutions (330).  
Ce modèle JC69 peut être représenté comme indiqué dans le **Tableau 1**.

$\lambda$  = taux de substitution

**Tableau 1 : Modèle JC69**

### Modèle TN84

Dans ce modèle décrit par Tajima et Nei (331) les fréquences de chacun des quatre nucléotides sont supposées être différentes de 0,25. Cette approche permet, en général, de mieux prendre en compte la composition réelle des séquences. On obtient ainsi une meilleure estimation des substitutions que dans le cas du modèle JC69. L'estimation qui est utilisée ici est basée sur le modèle décrit par Nei et Tajima (243):

$$d = -b \log_e \left(1 - \frac{p}{b}\right)$$

avec :  $b = \frac{1}{2} \left(1 - \sum_{i=1}^4 \pi_i^2 + \frac{p^2}{c}\right)$  et  $c = \sum_{i=1}^3 \sum_{j=i+1}^4 \frac{\pi_i \pi_j}{2\pi_i \pi_j}$

	A	T	C	G
A	-	$\beta$	$\gamma$	$\delta$
T	$\alpha$	-	$\gamma$	$\delta$
C	$\alpha$	$\beta$	-	$\delta$
G	$\alpha$	$\beta$	$\gamma$	-

Cette estimation est robuste et elle peut être appliquée dans un grand nombre de cas, du moins tant que le nombre de substitutions nucléotidiques par site est inférieure à 1,0.

Ce modèle peut être représenté comme indiqué dans le **Tableau 2**.

$\alpha, \beta, \gamma, \delta$  = taux de substitution pour chaque nucléotide

**Tableau 2 : Modèle TN84**

## Modèle K80

Ce modèle défini par Kimura (1967) est également appelé 'Kimura à deux paramètres'. Comme généralement, et en particulier dans l'ADN mitochondrial (mtDNA), les transitions surviennent avec une fréquence plus grande que les transversions (29), le modèle JC69 fournit une sous estimation du taux de substitution ( $\lambda$ ), spécialement lorsque la valeur de la distance génétique ( $d$ ) est inférieure à 0,1. De ce fait l'estimation, par le maximum de vraisemblance, de Kimura est utilisée:

$$d = \left( \frac{-\log_e(1-2P-Q)}{2} \right) - \left( \frac{\log_e(1-2Q)}{4} \right)$$

Sous ce modèle il est également possible d'estimer les substitutions par transitions ( $v$ ) et par transversions ( $s$ ):

$$s = \left( \frac{-\log_e(1-2P-Q)}{2} \right) + \left( \frac{\log_e(1-2Q)}{4} \right)$$

$$v = \frac{-\log_e(1-2Q)}{4}$$

ainsi que le rapport ( $R$ ) des transitions sur les transversions:

$$R = \left( \frac{\log_e(1-2P-Q)}{\log_e(1-2Q)} \right) - \frac{1}{2}$$

	A	T	C	G
A	-	$\beta$	$\beta$	$\alpha$
T	$\beta$	-	$\alpha$	$\beta$
C	$\beta$	$\alpha$	-	$\beta$
G	$\alpha$	$\beta$	$\beta$	-

Ce modèle peut être représenté par la matrice du **Tableau 3**.

$\alpha$  = taux de transitions

$\beta$  = taux de transversions

**Tableau 3 : Modèle K80**

## Modèle F81

Ce modèle a été décrit par Felsenstein (89).

	A	T	C	G
A	-	$\Pi_T\lambda$	$\Pi_C\lambda$	$\Pi_G\lambda$
T	$\Pi_A\lambda$	-	$\Pi_C\lambda$	$\Pi_G\lambda$
C	$\Pi_A\lambda$	$\Pi_T\lambda$	-	$\Pi_G\lambda$
G	$\Pi_A\lambda$	$\Pi_T\lambda$	$\Pi_C\lambda$	-

Il prend en compte des fréquences inégales à l'équilibre pour les quatre nucléotides. Il peut être représenté comme indiqué dans le **Tableau 4** où  $\Pi$  représente la fréquence d'une base

Tableau 4 : Modèle F81

## Modèle T92

Ce modèle a été défini par Tamura (336). Si dans le modèle K80 les fréquences des quatre nucléotides sont supposées être identiques à l'équilibre, l'analyse de nombreuses séquences montre qu'il existe des variations, parfois importantes dans le contenu en Guanine et Cytosine (G + C). Une telle situation se retrouve, en particulier, dans le mtDNA pour lequel le modèle T92 a été défini. L'estimation de (d) est ici effectuée par le maximum de vraisemblance:

$$d = -2\Theta(1-\Theta)\log_e\left(1 - \left(\frac{P}{2\Theta(1-\Theta)}\right) - Q\right) - (1-2\Theta(1-\Theta))\log_e\left(\frac{1-2Q}{2}\right)$$

avec  $\Theta$  = contenu en G + C

L'estimation des transitions et transversions est effectuée par :

$$s = 2\Theta(1-\Theta)\log_e\left(1 - \left(\frac{P}{2\Theta(1-\Theta)}\right) - Q\right) + \Theta(1-\Theta)\log_e(1-2Q)$$

$$v = \frac{\log_e(1-2Q)}{2}$$

et celle de (R) par :

$$R = \frac{4\Theta(1-\Theta)\log_e\left(1 - \left(\frac{P}{2\Theta(1-\Theta)}\right) - Q\right)}{\log_e(1-2Q) - 2\Theta(1-\Theta)}$$

	A	T	C	G
A	-	$(1-\Theta)\beta$	$\Theta\beta$	$\Theta\alpha$
T	$(1-\Theta)\beta$	-	$\Theta\alpha$	$\Theta\beta$
C	$(1-\Theta)\beta$	$(1-\Theta)\alpha$	-	$\Theta\beta$
G	$(1-\Theta)\alpha$	$(1-\Theta)\beta$	$\Theta\beta$	-

Dans toutes ces estimations le contenu en Guanine et Cytosine ( $\Theta$ ) peut être calculé, soit pour chaque paire de séquences, soit pour l'ensemble des séquences étudiées. La matrice représentant ce modèle est donnée au **Tableau 5**.

**Tableau 5 : Modèle T92**

### Modèle HKY85

Ce modèle a été décrit par Hasegawa et coll. (131). Il s'agit de l'un des plus utilisés au cours des études sur les séquences de mtDNA. Il n'existe pas ici de formule analytique permettant l'estimation de (d).

	A	T	C	G
A	-	$\Pi_T\beta$	$\Pi_C\beta$	$\Pi_G\alpha$
T	$\Pi_A\beta$	-	$\Pi_C\alpha$	$\Pi_G\beta$
C	$\Pi_A\beta$	$\Pi_T\alpha$	-	$\Pi_G\beta$
G	$\Pi_A\alpha$	$\Pi_T\beta$	$\Pi_C\beta$	-

La matrice définissant ce modèle est celle du **Tableau 6**. On peut remarquer qu'elle est très proche de celle du modèle F81.

**Tableau 6 : Modèle HKY85**

### Modèle TN93

Dans ce modèle décrit par Tamura et Nei (337) est intégré le fait que, dans les séquences réelles, les taux de transitions entre purines et entre pyrimidines ne sont pas identiques. Ce modèle apparaît donc plus réaliste que le modèle HKY85. L'estimation de (d) est donnée par:

$$d = -\frac{2\Pi_A\Pi_G}{\Pi_R} \log_e \left( 1 - \left( \frac{\Pi_R}{2\Pi_A\Pi_G} \right) P_1 - \left( \frac{1}{2\Pi_R} \right) Q \right) - \frac{2\Pi_T\Pi_C}{\Pi_Y} \log_e \left( 1 - \left( \frac{\Pi_Y}{2\Pi_T\Pi_C} \right) P_2 - \left( \frac{1}{2\Pi_Y} \right) Q \right) - 2 \left( \Pi_R\Pi_Y - \left( \frac{\Pi_A\Pi_G\Pi_Y}{\Pi_R} \right) - \left( \frac{\Pi_T\Pi_C\Pi_R}{\Pi_Y} \right) \right) \log_e \left( 1 - \left( \frac{1}{2\Pi_R\Pi_Y} \right) Q \right)$$

avec  $P_1$  = proportion de transitions entre A et G et  $P_2$  = proportion de transitions entre T et C

Il est également possible d'estimer les valeurs de (s) et (v) :

$$s = -\frac{2\Pi_A\Pi_G}{\Pi_R} \log_e \left( 1 - \left( \frac{\Pi_R}{2\Pi_A\Pi_G} \right) P_1 - \left( \frac{1}{2\Pi_R} \right) Q \right) - \frac{2\Pi_T\Pi_C}{\Pi_Y} \log_e \left( 1 - \left( \frac{\Pi_Y}{2\Pi_T\Pi_C} \right) P_2 - \left( \frac{1}{2\Pi_Y} \right) Q \right) + 2 \left( \left( \frac{\Pi_A\Pi_G\Pi_Y}{\Pi_R} \right) - \left( \frac{\Pi_T\Pi_C\Pi_R}{\Pi_Y} \right) \right) \log_e \left( 1 - \left( \frac{1}{2\Pi_Y\Pi_R} \right) Q \right)$$

$$v = -2\Pi_R\Pi_Y \log_e\left(1 - \left(\frac{1}{2\Pi_R\Pi_Y}\right)Q\right)$$

	A	T	C	G
A	-	$\Pi_T\beta$	$\Pi_C\beta$	$\Pi_G\alpha_1$
T	$\Pi_A\beta$	-	$\Pi_C\alpha_2$	$\Pi_G\beta$
C	$\Pi_A\beta$	$\Pi_T\alpha_2$	-	$\Pi_G\beta$
G	$\Pi_A\alpha_1$	$\Pi_T\beta$	$\Pi_C\beta$	-

La matrice correspondant à ce modèle est celle présentée dans le **Tableau 7**.

$\alpha_1$  = taux de transitions entre purines  
 $\alpha_2$  = taux de transitions entre pyrimidines  
 $\beta$  = taux de transversions

**Tableau 7 : Modèle TN93**

### Modèle F84

Ce modèle décrit par Felsenstein (90) est sensiblement identique au modèle HKY85. Les fréquences nucléotidiques à l'équilibre sont différentes pour les quatre types de nucléotides.

### Modèle REV

Ce modèle a été défini par Yang (375); il s'agit d'une modification du modèle REV général. Il requiert six paramètres: les taux de substitution de A en C, de A en G, de A en T, de C en G, de C en T et de G en T. Pour ces six paires de nucléotides les deux types de substitution possibles sont supposés survenir avec la même fréquence.

### Distances gamma

Dans les modèles précédents le taux des substitutions nucléotidiques est supposé être identique pour tous les sites étudiés. En fait ceci est rarement le cas. Des analyses statistiques ont montré que cette valeur variait en suivant approximativement une distribution de type gamma (173, 337, 353, 358). Il s'agit d'une densité de probabilité qui est souvent utilisée pour des études de probabilités et de statistiques (par exemple le test du  $X^2$  est un cas particulier de cette distribution gamma). La forme de la distribution change de manière très importante en fonction du paramètre de forme appelé (a) ou ( $\alpha$ ). Ce paramètre est défini comme étant l'inverse du taux de substitution ( $\lambda$ ), et de ce fait la moyenne de la distribution a une valeur de 1,0 et sa variance est de (1/a). Lorsque (a) tend vers l'infini ce modèle d'hétérogénéité des taux entre sites tend vers un modèle dans lequel les taux sont égaux.

La plupart des estimations empiriques du paramètre (a) donnent des valeurs comprises entre 0,1 et 0,5 (379), ce qui implique bien l'existence de variations non négligeables des taux entre sites. La distribution gamma peut être appliquée à trois des modèles précédents:



### Modèle JC69

$$d = \frac{3a(1 - (\frac{4P}{3}))^{-(1/a)} - 1}{4}$$

### Modèle K80

$$d = \frac{a}{2}(1 - 2P - Q)^{-(1/a)} + (\frac{1}{2})(1 - 2Q)^{-(1/a)} - (\frac{3}{2})$$

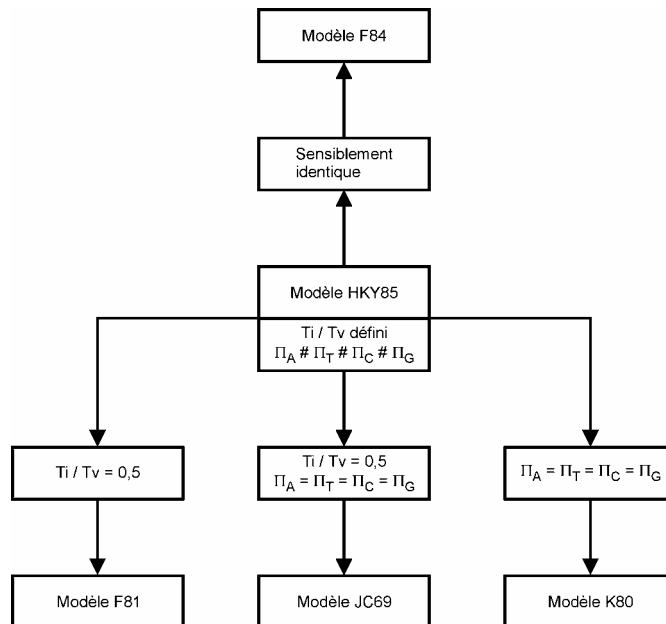
$$s = \frac{a}{2}(1 - 2P - Q)^{-(1/a)} - (\frac{1}{2})(1 - 2Q)^{-(1/a)} - (\frac{1}{2})$$

$$v = \frac{a}{2}((1 - 2Q)^{-(1/a)} - 1)$$

### Modèle TN93

$$d = 2a \left( \frac{\Pi_A \Pi_G}{\Pi_R} \left( 1 - \left( \frac{\Pi_R}{\Pi_A \Pi_G} \right) P_1 - \left( \frac{1}{2\Pi_R} \right) Q \right)^{-(1/a)} + \frac{\Pi_T \Pi_C}{\Pi_Y} \left( 1 - \left( \frac{\Pi_Y}{\Pi_T \Pi_C} \right) P_2 - \left( \frac{1}{2\Pi_Y} \right) Q \right)^{-(1/a)} \right. \\ \left. + (\Pi_R \Pi_Y - \left( \frac{\Pi_A \Pi_G \Pi_Y}{\Pi_R} \right) - \left( \frac{\Pi_T \Pi_C \Pi_R}{\Pi_Y} \right)) \left( 1 - \left( \frac{1}{2\Pi_R \Pi_Y} \right) Q \right)^{-(1/a)} - \Pi_A \Pi_G - \Pi_T \Pi_C - \Pi_R \Pi_Y \right)$$

Les relations existant entre les principaux modèles de substitutions nucléotidiques sont résumées dans la **Figure 2**.



**Figure 2 : Relations entre les principaux modèles de substitutions**

## Séquences protéiques

Dans le cas des séquences protéiques il est possible d'utiliser les calculs de distances suivants:

### p-distances

Il s'agit ici de l'estimation de la proportion d'acides aminés différents entre les deux séquences traitées:

$$p = \frac{n_d}{n}$$

### Distances avec correction de Poisson

Le calcul prend en compte le fait que le nombre de substitutions par site suit une loi de Poisson:

$$d = \log_e(1 - p)$$

### Distances gamma

Dans ce modèle, comme pour les séquences nucléotidiques, le taux de substitutions des acides aminés varie de site à site en suivant une distribution de type gamma. Les calculs pour ce type de distances ont été établis par Nei et coll. (241):

$$d = a((1 - p)^{-(1/a)} - 1)$$

Les modèles de substitutions qui viennent d'être définis peuvent également être modifiés afin de tenir compte d'évènements tels que les insertions et les délétions (27, 349, 350), ou la structure secondaire des protéines formées dans le cas de séquences codantes (230, 348), ou encore de la corrélation existant entre taux de substitution et sites (48, 93, 376).

## Substitutions synonymes et non synonymes

Dans le cas des séquences codantes, l'interprétation des modifications des taux de substitution doit tenir compte de la manière dont ces changements vont influencer sur le produit final (la protéine). On est ainsi amené à définir deux types de changements (296):

- Ceux n'altérant pas la protéine. Ces changements sont dits silencieux ou synonymes.
- Ceux induisant des changements d'acides aminés, ils sont alors appelés changements non silencieux ou non synonymes.

Lorsque l'on constate une augmentation dans le nombre des changements synonymes, cela est généralement le signe d'une évolution adaptative (213). D'autre part une prédiction importante de la théorie de la neutralité est que les taux de substitutions synonymes et non synonymes doivent être identiques tant entre les espèces qu'à l'intérieur des espèces (197). Dans le cas de l'ADN

mitochondrial cette prédiction n'est pas fondée puisque l'on constate que le taux des substitutions non synonymes est de 5 à 10 fois supérieur à l'intérieur des espèces qu'entre les espèces.

Les mécanismes de sélection font que la plupart des substitutions non synonymes se trouvent éliminées, ce qui fait que, généralement, le taux des substitutions synonymes apparaît plus élevé que celui des substitutions non synonymes (168, 224). Toutefois, dans certains cas, à la suite d'une sélection Darwinienne positive, on peut constater une accélération de l'apparition des substitutions non synonymes (148, 183). Dans la sélection naturelle les substitutions synonymes sont invisibles (7), alors que les mutations non synonymes sont soumises à de fortes pressions de sélection. Le taux de fixation de ces deux types de substitutions fournit alors un outil performant pour l'étude des mécanismes d'évolution des séquences nucléotidiques. Par exemple un taux variable du rapport des deux types de modifications peut indiquer l'existence d'une évolution adaptative (213), ou de contraintes sélectives (55), et ainsi permettre la détection de sites présentant une sélection positive (245).

L'approche la plus basique pour l'estimation de ces deux types de substitutions est la suivante:

- ✓ Dénombrer les sites synonymes et non synonymes dans les séquences à analyser.
- ✓ Comparer chaque paire de codons homologues, site à site, et en déduire le nombre de différences synonymes et non synonymes en utilisant la voie d'évolution la plus parcimonieuse entre les deux codons.
- ✓ Ajuster les valeurs obtenues afin de tenir compte des substitutions multiples, en utilisant un modèle standard d'évolution, tel que le modèle JC69, ou le modèle K80.

Cette estimation peut également être effectuée en utilisant différentes autres méthodes (190, 224, 242) dont certaines utilisent le principe du maximum de vraisemblance (115, 231, 232). Le modèle décrit par Nei et Gojobori fournit de bonnes estimations dans les cas où il n'existe pas de biais important du rapport  $T_i / T_v$ . Par contre dans le modèle de Goldman et Yang le biais du rapport  $T_i / T_v$  est pris en compte, de même que les propriétés physico-chimiques des acides aminés en utilisant les distances de Grantham (120): plus cette distance est élevée, plus les deux acides aminés en cause sont différents. Toutefois cette approche ne tient pas compte du fait que le nombre et la localisation des deux types de sites peuvent être modifiés dans le temps. Les approches basées sur l'étude des codons ont permis, entre autre, d'expliquer l'hétérogénéité des taux trouvés dans le génome des chloroplastes, en les reliant à des différences survenant dans les taux de substitutions non synonymes (232). Pour Kimura (168) les substitutions non synonymes faisant intervenir des acides aminés similaires surviennent plus fréquemment que celles faisant intervenir des acides aminés très différents; cette interprétation est consistante avec la théorie neutre de l'évolution moléculaire. Par contre pour Gillespie (106), les substitutions les plus fréquentes ne sont pas celles impliquant des acides aminés identiques, mais ceux pour lesquels les distances de Miyata (mesure de la dissimilarité des acides aminés positivement corrélée avec la distance de Grantham) sont égale à 1,0.

La comparaison et le classement de ces deux types de sites sont délicats car leur nombre et leur position sont susceptibles de changer dans le temps.

Chez les Mammifères, le taux des substitutions synonymes varie significativement en fonction des gènes étudiés (26, 229, 370). Ces variations peuvent être le résultat tant d'une

variabilité dans les taux de mutation que de différences de pressions sélectives portant sur les sites synonymes. Pour de nombreux auteurs ces sites synonymes sont neutres car les taux de substitution les affectant sont très proches de ceux que l'on trouve au niveau des pseudo gènes, ou encore au niveau du génome total (187, 370). Un autre facteur qui doit être pris en compte est l'existence, comme cela a été montré chez la souris, d'une sélection dans l'usage des codons (64). De plus la comparaison des deux types de substitutions suggère que les positions synonymes et non synonymes peuvent avoir la même étendue sous l'influence de pressions sélectives (9, 229, 247). De plus l'intensité de la sélection portant sur les codons est positivement corrélée avec le niveau d'expression du gène considéré (72, 297). Certains travaux tendent également à montrer que les taux de mutations pourraient varier en fonction du modèle d'expression du gène considéré (72, 297, 321). Enfin si l'on compare les fréquences estimées des substitutions synonymes et non synonymes survenant sans aucune sélection et que ces valeurs sont soustraites de celles effectivement observées, la valeur résiduelle est celle que l'on peut attribuer à un effet de sélection.

Un certain nombre de travaux montrent qu'il existe des différences significatives dans les taux de substitutions trouvés chez les Mammifères et les Vertébrés, en général, lorsque des séquences de mtDNA sont utilisées. En particulier ce taux, calculé à partir des sites du type 'four fold degenerate' est de 7 à 8 fois plus lent chez les Requins que chez les Primates ou les Ungulates (210). Dans le cas des poissons téléostéens l'évolution est de 4 à 5 fois plus faible que pour les Mammifères lorsque l'on considère les substitutions non synonymes (40). Mais pour les Tortues les résultats obtenus sont discordants et dépendent des séquences analysées: en utilisant le cytochrome b, ou les sites de restriction, le taux obtenu est plus faible que celui trouvé chez les autres Vertébrés (17, 30); alors qu'en utilisant le 12S rRNA cette hypothèse ne se vérifie plus (293). Une telle hétérogénéité des substitutions, pour les séquences mitochondriales, a également été reportée lors de l'étude des séquences protéiques (1): dans ce cas on constate que les Mammifères évoluent au moins six fois plus rapidement que les Poissons. Ces études montrent également que les taux augmentent des Amphibiens aux Oiseaux et aux Mammifères.

L'étude des 13 protéines, ou des gènes correspondants, disponibles chez les Mammifères (correspondant à 6 espèces appartenant à 5 ordres) montre que pour la majorité des gènes mitochondriaux les substitutions non synonymes suivent une horloge moléculaire (152). Des données plus récentes, obtenues à partir d'un plus grand nombre de gènes, montrent pour leur part que les Primates évoluent plus rapidement que tous les autres ordres (6, 12, 14, 141) et parmi ceux-ci ce sont les babouins et les orangs-outangs qui présentent les taux les plus élevés (3, 13, 143, 374).

L'évaluation du taux absolu d'évolution du mtDNA a été effectuée par Penny et Hendy (259). Ces auteurs ont comparé des espèces proches pour lesquelles les dates de divergence étaient connues et ils ont ainsi montré l'existence d'une certaine hétérogénéité des taux au niveau des sites non synonymes, ce qui permet de calculer que les Primates évoluent 1,8 fois plus rapidement que les Perissodactyles et 1,5 fois plus vite que les Carnivores.

Tous les modèles de substitutions implémentés par le maximum de vraisemblance, de même que les méthodes d'estimation de distances génétiques, supposent que les substitutions suivent un processus de Poisson homogène dans le temps. Afin de tester la validité de ces modèles il est possible d'effectuer un test statistique dans lequel l'hypothèse nulle est celle dans laquelle transitions et transversions surviennent avec des fréquences identiques. Le modèle F81 est particulièrement bien adapté pour de tels calculs. Les valeurs ainsi obtenues pourront alors être comparées à celles obtenues par l'utilisation du modèle HKY85 dans lequel il est possible de différencier les taux de transitions et de transversions.

## Méthodes de reconstruction

La reconstruction d'arbres phylogénétiques est avant tout un problème statistique (45). En ce sens l'arbre reconstruit doit être considéré comme une estimation de l'arbre réel, cette estimation étant caractérisée par une topologie et des longueurs de branches données. Mais il est également important de tenir compte des données biologiques disponibles pour accepter, ou rejeter, une topologie particulière.

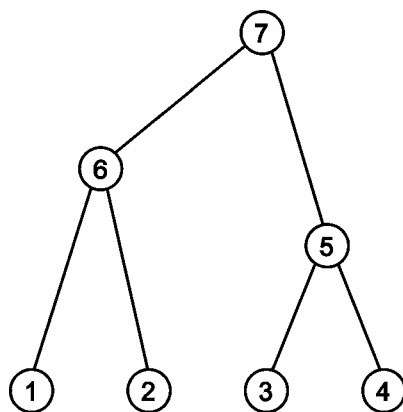
Lorsqu'un arbre est obtenu par l'étude d'un gène, cet arbre est appelé un 'gene tree'; alors que lorsqu'il est obtenu en prenant en compte l'enchaînement d'évènements tels que séparations et divergences entre espèces, il est appelé un 'species tree'. Un 'gene tree' est sensible aux erreurs d'échantillonnage due à la présence d'allèles polymorphiques dans la population ancestrale (237, 244, 254, 328). D'autre part si les séquences utilisées pour la reconstruction proviennent d'espèces différentes le 'gene tree' peut ne pas présenter la même topologie que l'arbre reliant les espèces.

De nombreuses méthodes de reconstruction peuvent être utilisées, tant avec des séquences nucléotidiques qu'avec des séquences protéiques (92, 221). Ces méthodes utilisent:

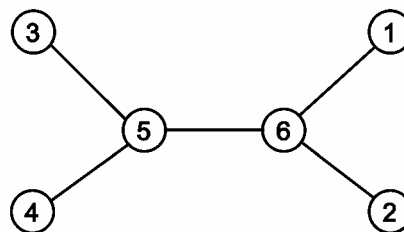
- Soit des distances estimées entre chaque paire possible de séquences. L'arbre est alors calculé en utilisant certains principes et algorithmes.
- Soit des caractères discrets, et dans ce cas l'arbre est reconstruit en considérant les relations d'évolution des différentes séquences, et ce pour chaque caractère ou position nucléotidique.

La plupart des méthodes de reconstruction ne tiennent compte que des homologies entre séquences, c'est à dire des similitudes héritées d'un ancêtre commun. Les homoplasies, ou similitudes non héritées d'un ancêtre commun, ne sont pas prises en compte. D'autre part ces méthodes sont généralement basées sur des principes d'optimisation (240, 326), dans ce cas la topologie finale est sélectionnée en attribuant à chaque arbre possible un score et en choisissant, selon la méthode, l'arbre ayant le score le plus faible ou le plus élevé.

Il est possible de reconstruire des arbres enracinés et non enracinés:



Arbre enraciné



Arbre non enraciné

Le nombre de topologies possibles est fonction du type de reconstruction et il augmente fortement avec le nombre de séquences, ou taxa, étudiés. Ainsi pour un arbre non enraciné comprenant 'n' taxa le nombre de topologies possibles est donné par la formule:

$$\frac{(2n-5)!}{(n-2)!2^{n-3}}$$

alors que dans le cas d'un arbre enraciné ce nombre est donné par la formule:

$$\frac{(2n-3)!}{(n-2)!2^{n-2}}$$

Ainsi pour des reconstructions comprenant de 4 à 10 taxa les nombres de topologies possibles seront:

<b>n</b>	<b>Non enraciné</b>	<b>Enraciné</b>
4	3	15
5	15	105
6	105	945
7	945	10 395
8	10 395	135 135
9	135 135	2 027 025
10	2 027 025	34 459 425

Dans les alignements de séquences utilisés pour les reconstructions phylogénétiques il existe des positions dites constantes. Celles-ci peuvent apparaître soit sous l'effet du hasard, soit parce que la, ou les substitutions ayant eu lieu à ces positions ont laissé les séquences identiques. Une telle constatation suggère que certaines positions sont si vitales que la sélection naturelle empêche qu'elles soient modifiées (98, 99, 131, 253). De tels sites sont dits invariables. De ce fait un site peut être constant, soit par chance, soit parce qu'il est invariable, et il est difficile de savoir si un site apparemment constant, pour un ensemble donné de séquences, deviendra variable si une nouvelle séquence est ajoutée au groupe.

## **Méthode UPGMA**

La méthode UPGMA (Unweighted Pair Group Method Analysis) a été développée en premier lieu dans le cadre d'études taxonomiques.

Si les taux de substitution (que ce soit pour les nucléotides ou pour les acides aminés) sont identiques dans toutes les lignées analysées, cette méthode peut être appliquée tant aux séquences nucléotidiques que protéiques. Des simulations sur ordinateur ont montré que l'utilisation d'une horloge moléculaire, combinée avec des distances évolutives importantes entre lignées, font que la probabilité d'obtenir l'arbre correct est faible (313, 339). La méthode UPGMA est donc à réserver pour l'étude des lignées proches ou très proches. Si le traitement porte sur plusieurs gènes est si le nombre de caractères étudiés est important, l'arbre obtenu sera fiable. Cet arbre est 'non enraciné', les taux d'évolution étant ici considérés comme constants. Toutefois il est nécessaire que les estimations de distances, sur lesquelles repose l'arbre, soient linéaires par rapport au temps d'évolution.

## Méthode ME

Dans la méthode ME (Minimum Evolution) décrite par Saitou et Nei (279) les estimations de distances sont corrigées afin de tenir compte des modifications multiples pouvant survenir à un même site.

Au cours de la recherche l'arbre qui sera considéré comme étant le plus 'correct' sera celui pour lequel la somme des longueurs des branches sera la plus faible. Si (n) séquences sont traitées et si il existe uniquement des bipartitions dans l'arbre, la méthode ME nécessite l'évaluation, pour chaque arbre possible, de  $(2n - 3)$  longueurs de branches. Lorsque les estimations de distances ne présentent aucun biais la méthode ME présente un solide fond théorique (276).

Toutefois la méthode ME présente un inconvénient: le temps de calcul est particulièrement long car il est toujours nécessaire d'évaluer un grand nombre de longueurs de branches.

## Méthode NJ

La méthode NJ (Neighbor-Joining) est une version simplifiée de la méthode ME dans laquelle les longueurs des branches ne sont pas calculées pour toutes les topologies possibles. L'algorithme initial a été modifié par Studier et Keppler (320), il permet d'obtenir un arbre non enraciné. Cet arbre pourra être, ultérieurement, enraciné par l'intermédiaire d'un outgroupe. Mais sans la présence de cet outgroupe la racine peut parfois être donnée par le point séparant les deux séquences les plus éloignées, en admettant que le taux d'évolution est constant dans toutes les branches. Si le nombre de séquences analysées est élevé et si les divergences entre séquences sont faibles, les méthodes NJ et ME fournissent des topologies très voisines. D'autre part des simulations ont permis de montrer que l'existence de petites erreurs dans les estimations de distances n'empêchent pas d'obtenir l'arbre correct (16).

## Méthode MP

Initialement la méthode MP (maximum Parsimony) a été utilisée pour le traitement de caractères morphologiques (204, 308, 325). Par la suite de très nombreuses approches ont été basées sur ce modèle et la plus utilisée est celle décrite par Eck et Dayhoff (79). Dans ce cas les changements évolutifs sont supposés survenir entre n'importe laquelle des paires possibles de nucléotides. Dans certaines conditions il est également possible d'attribuer un poids différent aux transitions et aux transversions afin de mieux prendre en compte certaines contraintes d'évolution (324, 365).

Les travaux de Felsenstein (92) ont permis de montrer que si le taux d'évolution varie largement avec l'évolution des séquences, la méthode MP représente un estimateur non consistant de l'arbre correct. Les conditions dans lesquelles la méthode MP peut être utilisée sont les suivantes (239):

- La divergence entre les séquences doit être faible, les valeurs estimées de (d) devant être inférieures à 0,1.
- Le taux des substitutions nucléotidiques doit être plus ou moins constant.

- Le biais du rapport Transitions / Transversions doit être faible.
- Le biais du contenu en Guanine et Cytosine doit être faible.
- Le nombre de nucléotides analysés doit être élevé (plusieurs centaines).

De nombreux travaux (278, 340) tendent à montrer que la méthode MP a une efficacité moindre que les méthodes NJ ou ME.

La méthode MP suppose l'existence d'un certain nombre d'hypothèses:

- Chacun des sites étudiés évolue de manière indépendante.
- Les différentes lignées utilisées dans l'analyse évoluent indépendamment.
- La probabilité pour que survienne la substitution d'une base à un site donné est faible en regard de la longueur de la branche de l'arbre (celle-ci étant une approximation du temps de divergence).
- Le nombre de changements dans les différentes branches de la topologie ne doit pas varier en dehors des limites suivantes: un changement dans les branches présentant un faible taux de variation et deux changements dans les branches avec un taux de variation élevé.
- La quantité, supposée ou estimée, de changements ne doit pas présenter de trop grandes variations entre les différents sites analysés.

Au cours de l'analyse l'ensemble des sites présents sur les séquences n'est pas étudié; seuls sont pris en considération ceux pour lesquels il existe au moins deux nucléotides différents, chacun d'eux étant présent au moins deux fois. De tels sites sont appelés 'parsimony informative sites'. Pour chaque site retenu, et pour une topologie donnée, les nucléotides des séquences ancestrales sont estimées, de même que le nombre minimum de substitutions nécessaires pour expliquer les différences observées. La somme de toutes ces valeurs, pour une topologie, constitue ce que l'on appelle le nombre de pas, ou longueur d'arbre. La topologie pour laquelle la longueur d'arbre est la plus faible représente l'arbre le plus parcimonieux. En pratique on trouve souvent deux ou plusieurs arbres présentant cette valeur minimale. De tels arbres sont dits également parcimonieux. Dans ce dernier cas la résolution finale sera effectuée en faisant appel à une méthode de consensus.

Les arbres obtenus par la méthode MP ne sont normalement pas enracinés. D'autre part, dans certaines conditions, les longueurs des branches individuelles peuvent être déterminées (95, 204, 324).

Lorsque l'analyse porte sur moins de 10 séquences, il est possible d'examiner l'ensemble des topologies pouvant être reconstruites. Par contre si le nombre de séquences est plus élevé, le temps de calcul devient très long, et afin de le diminuer on a généralement recours à une méthode dite 'branch-and-bound' (133) dans laquelle tous les arbres ayant une longueur supérieure à celle de l'arbre précédemment examiné sont ignorés dans la suite du traitement. Si le nombre de séquences est supérieur à 20 la méthode MP devient très lente et l'on utilise alors une recherche de type



heuristique. Dans celle-ci seulement un petit nombre d'arbres possibles est examiné, ce qui fait que l'on ne peut être absolument certain de trouver le ou les arbres les plus parcimonieux.

L'algorithme utilisé pour la méthode 'branch-and-bound' débute par un arbre constitué de trois séquences, pour lequel il n'existe qu'une seule possibilité de topologie non enracinée. Cet arbre peut être choisi, comme c'est le cas dans le logiciel MEGA, avec une longueur relativement importante. Afin de trouver cet arbre initial les différences nucléotidiques sont examinées pour toutes les paires possibles de séquences. Les deux séquences les plus divergentes sont alors retenues et l'on recherche, dans les séquences restantes, celle qui, ajoutée aux deux précédentes, permet d'obtenir un arbre dont la longueur, calculée par le principe du maximum de parcimonie, est maximale. Les autres séquences sont alors ajoutées une à une et la longueur de l'arbre est calculée à chaque étape. Si cette longueur est supérieure à une valeur prédéterminée, la topologie correspondante, ainsi que toutes celles qui en découlent, est ignorée.

L'approche par le maximum de parcimonie pose un certain nombre de problèmes:

- Elle ne présente aucune structure en fonction du temps.
- La probabilité pour qu'une substitution survienne dans un laps de temps court est supposée identique à celle pour qu'elle survienne dans un laps de temps long (114, 347).
- Dans la topologie obtenue les longueurs individuelles des branches ne sont pas prises en compte, bien qu'il s'agisse là d'un paramètre important pour la détermination des changements sur ces branches.
- La matrice de substitutions utilisée ne présente pas de bases claires.
- La réversion des substitutions n'est pas prise en compte, alors que ce facteur semble avoir une certaine importance lorsque les séquences analysées sont très voisines.
- L'existence d'une horloge moléculaire, combinée avec une faible valeur d'évolution, ne garantit pas la fiabilité statistique de la parcimonie, en particulier lorsque le nombre de séquences traitées est élevé.

La démarche d'une reconstruction par parcimonie pour 5 séquences (S1 à S5) de 5 caractères peut être résumée de la manière suivante:

Site	1	2	3	4	5
<b>S1</b>	G	C	A	T	T
<b>S2</b>	G	C	T	T	T
<b>S3</b>	A	A	C	C	T
<b>S4</b>	A	A	T	G	T
<b>S5</b>	A	A	T	T	A

L'une des topologies, non enracinée, possible est définie par: (((S3, S4), S5), (S1, S2)). Si l'on considère, par exemple, le site 3 il est possible de calculer le nombre minimal de changements nécessaires pour que ce site rende compte de la topologie choisie. Pour le groupe (S3, S4) ce site présente les nucléotides C et T, donc le nœud basal (nœud 6) pour ce groupe contiendra, au site 3, l'un de ces deux nucléotides. La séquence S5 est reliée par l'intermédiaire du nœud 7 à (S3, S4) et le

nucléotide présent au site 3 de S5 est T. La voie la plus parcimonieuse est alors d'assigner au site 3 du nœud 7 le nucléotide T. De même pour le groupe (S1, S2) le nœud basal (nœud 8) contiendra, au site 3, soit le nucléotide A, soit le nucléotide T. Comme pour le nœud 7 ce site contient T, la voie la plus parcimonieuse est d'attribuer au site 3 du nœud 8 le nucléotide T. De ce fait pour l'ensemble de l'arbre le site 3 présentera deux changements de nucléotides (A et C). Cette opération étant répétée pour chaque site il est alors possible d'estimer le nombre total de changements pour la topologie choisie. Enfin cette analyse sera répétée pour chacune des topologies possibles et celle, ou celles, présentant le nombre minimal de changements constituera l'arbre retenu comme étant le plus parcimonieux.

## Méthode ML

Dans la méthode ML (Maximum Likelihood), si l'on considère une topologie comprenant (b) branches définies à partir de (c) caractères, chacun de ceux-ci pouvant présenter (s) états, la vraisemblance pour l'ensemble des caractères de l'arbre est donnée par la formule suivante:

$$\prod_c (\sum_s (\prod_b p(b)))$$

Dans la pratique c'est le logarithme de cette valeur qui est calculé (ou log likelihood).

	A	T	C	G
A	- 0,75μ(κ+2)	0,25μ	0,25μ	0,25μ
T	0,25μ	- 0,75μ(κ+2)	0,25μ	0,25μκ
C	0,25μκ	0,25μ	- 0,75μ(κ+2)	0,25μ
G	0,25μ	0,25μκ	0,25μ	- 0,75μ(κ+2)

Tableau 8 : Matrice ML pour le modèle K80

Dans le cas des séquences nucléotidiques si l'on considère que tous les changements sont équiprobables, et que l'estimation du nombre de changements par site est (μ), il est alors possible d'utiliser le modèle JC69 (dans lequel chaque cellule de a matrice à une valeur de 0,25μ) comme matrice de substitution. Par contre si l'on veut tenir compte des la valeur (κ) du rapport Transitions / Transversions, il faut utiliser le modèle K80 et la matrice prend la forme présentée au **Tableau 8**.

Dans la méthode ML les caractères, qu'il s'agisse de nucléotides ou d'acides aminés, sont considérés séparément pour toutes les séquences, c'est à dire qu'ils sont vus comme indépendants. L'utilisation conjointe d'un modèle de substitution du type 'Markov chain' et de la méthode ML est actuellement bien établie, mais de nombreux mécanismes ne sont pas encore suffisamment connus pour permettre une modélisation aussi fiable que possible (113).

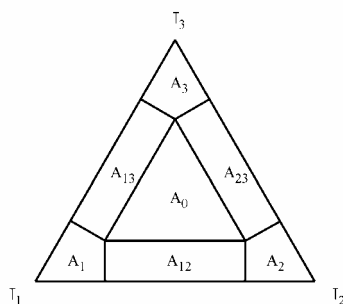
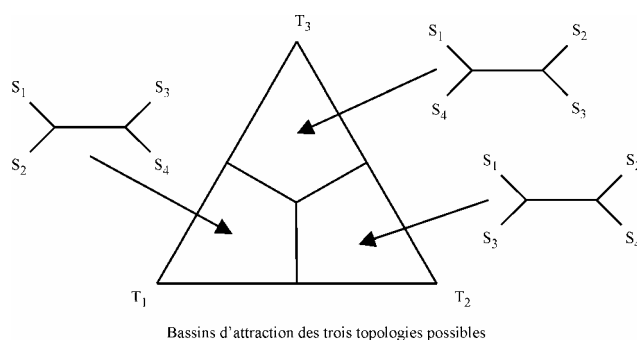
La lenteur de l'algorithme initial de la méthode ML peut être optimisé par l'utilisation de calculs comportant un certain nombre d'approximations (2, 250). D'autre part l'enracinement de l'arbre final peut être obtenu en utilisant la méthode de comparaison des outgroupes (84).

## Quartet Puzzeling

L'utilisation de la méthode ML pour la reconstruction d'arbres à partir de séquences, soit nucléotidiques, soit protéiques, ne pose aucun problème dans le cas du traitement de quatre séquences. De nombreuses méthodes existent pour la reconstruction d'une topologie en utilisant tous les assemblages possibles de quatre séquences (21, 71, 97, 284). Ces méthodes présentent des performances similaires à celles de la méthode NJ (286).

La méthode de 'Quartet Puzzeling' décrite par Strimmer et Von Haeseler (319) reconstruit, en utilisant la méthode ML, tous les quartets possibles à partir de l'ensemble des séquences à analyser. Ces quartets sont ensuite combinés afin d'obtenir l'arbre final. Dans cette méthode les séquences sont ajoutées, de manière aléatoire, à un sous arbre existant. L'arbre final est obtenu en utilisant un procédé de consensus du type 'majority rule' (208) à partir de tous les arbres reconstruits au cours de multiples traitements, ou itérations. Cet arbre peut être binaire ou multifurqué et une valeur indique la 'crédibilité' de chacune des branches internes.

Le pourcentage de quartets non résolus, c'est à dire ceux pour lesquels la branche interne a une valeur nulle, parmi l'ensemble des quartets possibles est un bon indicateur de la possibilité d'utiliser les séquences pour des reconstructions phylogénétiques. Ces données peuvent être représentées, de manière graphique, dans un diagramme triangulaire (**Figure 3**). Dans cette représentation les quartets non résolus se trouvent localisés dans la région centrale du triangle, alors que ceux totalement résolus se trouvent dans les angles du triangle.



Bassins d'attraction pour les topologies résolues et non résolues  
 A<sub>0</sub> = zone de 'tree star-like'  
 A<sub>1</sub>, A<sub>2</sub>, A<sub>3</sub> = zones des topologies totalement résolues  
 A<sub>12</sub>, A<sub>13</sub>, A<sub>23</sub> = zones des topologies partiellement résolues

**Figure 3 : Analyse 'likelihood mapping'**

## **Parsimony Ratchet**

La méthode de 'Parsimony Ratchet' (246) a été mise au point afin d'apporter une solution aux problèmes posés par les méthodes de réarrangements qui constituent l'étape finale de nombreuses méthodes de reconstruction. Elle comporte les étapes suivantes:

- ✓ Génération d'un arbre initial, par exemple un arbre du type 'Wagner', sur lequel sont effectuées une ou plusieurs étapes de 'branch swaping'.
- ✓ Sélection aléatoire d'une partie des caractères traités. Les poids de ces caractères sont alors modifiés: on peut, par exemple, ajouter 1,0 au poids initial.
- ✓ Réalisation d'un 'branch swaping' par la méthode TBR ou 'branch breaking' sur l'arbre en utilisant la nouvelle matrice de poids de caractères. A l'issue de cette opération un ou plusieurs arbres sont conservés.
- ✓ Les poids des caractères, précédemment sélectionnés, sont remis à leurs valeurs initiales.
- ✓ Réalisation d'un 'branch swaping' sur l'arbre ou les arbres sélectionnés à l'étape 3. A l'issue du traitement un ou quelques arbres sont conservés.
- ✓ Retour à l'étape 2.

En pratique 50 à 100 itérations sont effectuées en utilisant la sélection aléatoire de 5 à 25 % des caractères.

A titre d'exemple l'analyse d'un ensemble de 500 taxa, en utilisant l'option TBR du programme PAUP, réalisée sur un ordinateur Mac-Quadra, nécessite un mois afin d'obtenir un arbre ayant une longueur minimale de 16 225. Le même ensemble de données a été réanalysé par Rice et coll. (272) en ne conservant à chaque étape qu'un nombre minimal d'arbres et en effectuant l'analyse par la méthode TBR; dans ce cas, après 11,6 mois, un arbre de longueur 16 220 a été trouvé. Toujours avec le même ensemble de données Nixon et Davis, en utilisant le 'Parsimony Ratchet' et un ordinateur Pentium trouvent en 78 heures des arbres de longueur 16 220 (lorsque deux arbres sont conservés à chaque itération), et de longueur 16 218 en 150 heures.

Il semble que la méthode 'Parsimony Ratchet' soit en mesure de fragmenter les îlots d'arbres sub optimaux, et de permettre ainsi une meilleure analyse des données.

## **Mesures de la validité des arbres**

La méthode de test décrite par Wilson (368) utilise la technique des quartets et elle peut être décomposée en deux étapes:

- Décider, pour chaque quartet, lequel de ceux totalement résolus décrit le mieux les relations entre les quatre lignées.
- Combiner tous les quartets retenus en un 'super arbre'.

La première étape peut être réalisée en utilisant le principe du maximum de parcimonie: les longueurs des trois arbres possibles sont calculées et celui ayant la plus faible longueur est sélectionné. Il est également possible d'utiliser le principe du maximum de vraisemblance: la topologie est celle présentant la valeur maximale de vraisemblance.

La seconde étape peut être réalisée par addition des quartets en ordre aléatoire, le processus étant répété un certain nombre de fois et une méthode de consensus est utilisée pour l'obtention de l'arbre final. Mais il est aussi possible d'effectuer l'addition dans un ordre tel que la 'force du signal' soit maximale à chaque étape. Une dernière solution consiste à rechercher l'arbre le moins inconsistant.

## Méthodes de rééchantillonnage

Si l'on considère la topologie d'un arbre phylogénétique, il est possible de la subdiviser en un ou plusieurs 'sous arbres'. Les méthodes de rééchantillonnage peuvent alors être utilisées afin d'évaluer les limites de confiance de ces sous arbres (91). Deux méthodes sont principalement utilisées pour ce faire (82):

- Le bootstrap: les séquences obtenues par cette méthode ont des tailles identiques à celles des séquences d'origine. Le tirage aléatoire d'un site est effectué avec remise (c'est à dire que le même site peut être sélectionné plusieurs fois pour la constitution d'une séquence). Les nucléotides présents dans chacune des séquences à traiter, à la position du site tiré sont alors ajoutés aux séquences en cours de 'construction' et l'opération est répétée jusqu'à ce que les nouvelles séquences aient une taille identique à celle des séquences d'origine. Si le rééchantillonnage est effectué (N) fois, la probabilité pour qu'un site ne soit jamais rééchantillonné est de  $(1/N)^N$ . Ce rééchantillonnage suit une loi de Poisson, de moyenne 1,0; de ce fait on peut considérer que 37 % des sites ne sont jamais rééchantillonnés.
- Le jackknife: dans ce procédé il y a purge, de manière aléatoire, d'une partie des sites des séquences. Si la moitié des sites sont ainsi éliminés on effectue un 'half jackknife'. Les séquences obtenues sont donc toujours de taille inférieure aux séquences d'origine.

Ces deux méthodes fournissent généralement des résultats très peu divergents et sont utilisées afin de tester la robustesse des différents nœuds d'une topologie.

## Tests de vraisemblance

L'un des points essentiels au cours des analyses phylogénétiques est l'estimation de la manière dont les modèles de substitution utilisés, et les hypothèses formulées, jouent un rôle dans le résultat final (85, 219).

Dans une analyse de parcimonie sont inclus l'indépendance des caractères ainsi qu'un modèle spécifique pour la transformation de ces caractères. L'estimation de l'arbre 'correct' peut être effectuée avec une probabilité élevée, même si plusieurs hypothèses ne sont pas correctes, du moins partiellement. En effet les méthodes MP, ML et celles permettant l'estimation des distances génétiques sont robustes quant à la violation des hypothèses, en particulier celles concernant les

processus de substitution, la variation des taux entre sites et l'indépendance entre sites (144, 147, 287).

L'utilisation, dans l'analyse, de modèles explicites permet non seulement une estimation de la phylogénèse, mais aussi une approche des processus d'évolution en prenant en compte les hypothèses testées.

La mesure de la validité relative de deux hypothèses est obtenue par l'évaluation du rapport des vraisemblances de chaque hypothèse. Ce test est le LTR (Likelihood Ratio Test) décrit par Edwards; Kendall et Stuart; Milinkovitch et coll. (80, 163, 217):

$$\Delta = \frac{\text{Max}(L_0(\text{Null\_model}(\text{Data})))}{\text{Max}(L_1(\text{Alternative\_model}(\text{Data})))}$$

Si  $\Delta$  est inférieur à 1,0 l'hypothèse nulle ( $L_0$ ) est rejetée; dans le cas contraire c'est l'hypothèse alternative ( $L_1$ ) qui est rejetée. Cette dernière situation ne peut se produire que si les deux modèles ne sont pas liés (en d'autres termes si l'hypothèse nulle ne représente pas un cas particulier de l'hypothèse alternative. Lorsque  $\Delta$  est inférieur à 1,0 la valeur ( $- 2 \log \Delta$ ) suit sensiblement une distribution du  $X^2$  sous l'hypothèse nulle et avec un nombre de degrés de liberté égal à la différence entre le nombre de paramètres libres pour les deux hypothèses. De plus ce test présente certaines propriétés statistiques faisant que sa performance est optimale dans le cas d'hypothèses simples et qu'il se montre sous performant lorsque les hypothèses sont plus complexes (271).

Au cours de ces études phylogénétiques Felsenstein (92) a introduit un autre type de test: le 'parametric bootstrap' (52, 53) qui permet de générer une distribution nulle de ( $- 2 \log \Delta$ ) en utilisant une simulation de type Monte Carlo. Ce test est en mesure d'évaluer la robustesse d'une phylogénèse.

D'autres facteurs peuvent également être pris en compte pour tester la validité, ou la vraisemblance, d'une topologie. Par exemple Lanave et coll. (180) suggèrent l'analyse de la fréquence des bases. En effet tous les modèle de type 'Markov chain', habituellement utilisés, prédisent une égalité des fréquences des bases pour toutes les séquences étudiées. Il serait également possible d'utiliser le rapport des variances des différences entre les différentes paries de séquences (105, 168).

## Méthodes de consensus

Ces méthodes sont utilisées afin de déterminer l'arbre 'moyen' d'une série d'arbres possibles ou probables, ce qui est par exemple le cas d'une série d'arbres également parcimonieux. Deux types de consensus sont principalement utilisés:

- Le consensus strict dans lequel ne sont pris en compte que les groupes, ou 'sous arbres' présents dans 100 % des topologies analysées.
- Le consensus de type 'majority rule' dans lequel sont pris en compte les groupes présents avec une fréquence définie, comprise entre 50 % et 100 %.

## Méthodes de réarrangements

Dans le cas de la méthode MP, la recherche de l'arbre le plus parcimonieux est un processus complexe utilisant des techniques dites de réarrangements ou 'branch swaping'. Ces méthodes peuvent être classées en deux groupes:

- Les réarrangements locaux: ils impliquent un segment interne de l'arbre qui est échangé avec un autre segment de cet arbre. Si le nouvel arrangement fournit un arbre présentant de meilleures caractéristiques que l'arbre initial, il est retenu pour la suite du traitement. Un tel processus est très sensible à l'ordre d'introduction des séquences et il ne peut être mis en œuvre qu'à partir de l'addition de la cinquième séquence. Si le nombre (n) de séquences analysées est important le temps de traitement est long car il y a  $(n - 1)(2n - 3)$  topologies différentes à prendre en considération.
- Les réarrangements globaux: dans ce processus chaque 'sous arbre' de la topologie est supprimé, puis ajouté à toutes les positions possibles. Ce processus est appelé SPR (Subtree Pruning and Regrafting) (326). Il est aussi possible d'utiliser des procédés quelque peu différents qui sont le NNI (Nearest Neighbor Interchange) et le TBR (Tree Bisection and Reconnection).

A l'heure actuelle les méthodes les plus utilisées sont celles de 'branch breaking' (87) et de TBR. Le principal problème rencontré lors de l'utilisation de ces méthodes est le temps de traitement qui augmente logarithmiquement avec le nombre de taxa traités, et ceci est particulièrement sensible avec la méthode TBR. De plus lorsque les données à traiter sont nombreuses il existe toujours un grand nombre d'arbres 'sub optimaux' qui se répartissent dans des formations appelées 'islands' (202). Plusieurs stratégies ont été développées afin de minimiser ces effets, par exemple en diminuant le nombre d'arbres initiaux, ou en réduisant le nombre d'arbres conservés à chaque réplication, ou encore en collectant les résultats de nombreuses réplifications et en utilisant un sous ensemble de ces résultats pour des analyses plus complètes.

## Indices

Quatre types principaux de mesures peuvent être effectuées afin de juger de la validité d'un arbre phylogénétique, de sa crédibilité:

- Mesures de la réalité des branches: elles peuvent être effectuées en utilisant le 'Bremer indice', ou le SI (Support Index) (31), ou encore le bootstrap (91).
- Mesures de la validité de l'ensemble des caractères: elles peuvent utiliser
  - **Le CI** (Consistency Index) (171). Cet indice de cohérence est défini comme étant le rapport entre le nombre minimal de transformations nécessaires pour expliquer les états de tous les caractères, et le nombre effectif de transformations dans l'arbre considéré. Cet indice est une mesure du taux d'homoplasies présentes dans les données étudiées. Lorsque l'indice est voisin de 1,0 l'arbre est dépourvu d'homoplasies et l'on peut considérer que pour des valeurs inférieures ou égales à 0,3 l'arbre n'a plus de cohérence car il est trop riche en homoplasies (61). Le nombre de

caractères étudiés, de même que le nombre de séquences analysées, affectent de manière importante la valeur du CI (11, 282).

- **Le RI** (Retention Index) (36, 86, 116). Cet indice est défini comme étant le rapport entre le nombre d'homoplasies observables et le nombre d'homoplasies observées. Cet indice semble être la meilleure estimation du nombre d'homoplasies (61), de plus il est insensible au nombre de taxa inclus dans l'analyse (11).
  - **Le TSI** (Total Support Index) (32). Cet indice est une approche de la mesure de la crédibilité d'une phylogénèse. Il est influencé par l'incongruence des caractères, ainsi que par les caractères non informatifs (autapomorphies, caractères invariants). Cet indice est dérivé du SI en ce sens qu'il représente la somme des valeurs des SI déterminés pour tous les clades d'une topologie (161).
  - **Le PSI** (Proportional Support Index) (182). Il s'agit d'une version modifiée du TSI. Cet indice n'est pas affecté par la présence de caractères non informatifs et il représente ainsi une manière plus fiable de rendre compte de la force du signal phylogénétique présent dans les données.
- Mesures de la validité des changements de caractères pour l'arbre entier: les trois méthodes principales pouvant être utilisées sont le 'Dynamic weighting' (365), les 'Exact Indices' (295), la 'Self weighting optimization' (117).
- Mesures de la validité des optimisations effectuées pour les nœuds internes: de très nombreuses méthodes ont été décrites dans ce domaine, entre autres par Maddison; Pagel; Schuler; Schuler et coll.; Yang et Roberts (202, 252, 288, 289, 381).

Il est par ailleurs possible d'effectuer une approche cladistique de la robustesse des nœuds d'un arbre en utilisant le DC (Decay Index) (31, 54, 70). Ces résultats peuvent être comparés avec ceux obtenus par l'utilisation d'une méthode de rééchantillonnage.

## ***Modalités d'utilisation***

Pour les méthodes ME et NJ l'estimation des p-distances semble être plus efficace que les estimations effectuées en utilisant des modèles plus complexes (279, 313, 331). Ceci serait dû à la faible variance estimée sur les p-distances par rapport aux autres types d'estimations de distances génétiques.

Dans le cas de la méthode ML le modèle de substitution HKY85 est moins efficace que le modèle JC69 (35, 378). Ceci serait dû au fait que les méthodes complexes nécessitent un nombre important de paramètres; mais aussi à la nature complexe du traitement ML (238, 378, 380).

Des études empiriques récentes, utilisant des séquences réelles (10, 140, 209, 212, 338) suggèrent que les méthodes MP, NJ et ML fournissent généralement des phylogénèses identiques lorsque le test du bootstrap leur est appliqué.

Lorsque certains caractères dits 'pertinents' doivent être privilégiés par rapport aux homoplasies, il est possible d'utiliser des méthodes de pondérations successives; un tel traitement peut, par exemple, être appliqué après une analyse de parcimonie. La pondération décrite par Farris (86) permet d'attribuer à chaque caractère un indice, appelé RCI (Rescaled Consistency Index) qui



est en réalité le produit entre CI et RI. Une telle méthode est généralement en mesure de réduire le nombre d'arbres retenus.

## **Analyses combinées**

L'analyse combinée de différents gènes, ou de différentes zones de séquences, semble être en mesure, à priori, de fournir des informations plus consistantes, pouvant être utilisées dans le cadre des reconstructions phylogénétiques. L'utilisation de données provenant de gènes multiples peut permettre l'exécution de tests concernant les similitudes ou les différences pouvant exister dans les dynamiques d'évolution des différents gènes considérés. Il est par exemple possible d'étudier les fréquences nucléotidiques et les biais du rapport Transitions / Transversions dans ces gènes.

Toutefois il existe une controverse quant à l'utilisation, séparée ou conjointe, de différents gènes pour les analyses phylogénétiques (38, 323).

## **Effets d'échantillonnage**

Dans une étude sur les Eulemur Yoder et Irwin (388) montrent que les résultats de parcimonie effectués à partir de différents gènes (mitochondriaux et nucléaires) donnent des topologies différentes, que les modèles de substitutions utilisés soient plus ou moins complexes.

L'augmentation du nombre de caractères analysés a été étudiée par Sober (308) sur des simulations et par Bull et coll. (38), ou Huelsenbeck et coll. (145) sur des données réelles et il semble bien que la faible résolution obtenue avec un petit nombre de caractères soit à relier à ce faible nombre. D'autre part des données empiriques montrent que l'on gagne en résolution en utilisant de nombreux ensembles de données, relativement petits, et provenant de plusieurs gènes, plutôt qu'en augmentant le nombre de caractères pour un seul gène (42, 59).

Le nombre de taxa introduits dans l'analyse est un facteur dont l'augmentation permet très souvent de résoudre certains problèmes se posant au niveau des nœuds de la topologie (122, 135, 136, 164, 165, 263, 311). D'autre part le fait d'attribuer des poids différents aux caractères peut permettre l'augmentation du seuil de résolution d'une phylogenèse (47, 137, 220, 234, 390). Mais il existe aussi des arguments en faveur de l'attribution de poids égaux à tous les caractères (8, 49).

## ***Problèmes principaux***

### **Longueur des branches**

La longueur des branches d'une topologie joue un rôle important dans les méthodes ML et ME. Pour de nombreuses raisons les estimations de ces longueurs de branches ne sont pas statistiquement indépendantes sous les deux types de reconstructions (195). La fréquence et l'amplitude des corrélations (dont la plupart sont négatives) entre les estimations des longueurs des branches sont des facteurs augmentant l'hétérogénéité de ces longueurs. De nombreuses méthodes ont été proposées afin de surmonter, du moins en partie, les problèmes liés aux longueurs des branches rencontrés au cours des reconstructions. Par exemple Felsenstein (88) suggère que l'utilisation d'algorithmes incorporant un modèle pour les changements d'état des caractères peut être en mesure de résoudre le problème des branches internes courtes. Pour d'autres auteurs

l'augmentation du nombre de caractères, ou l'augmentation du nombre de taxa, pourrait contribuer à une meilleure définition de la topologie.

Dans le cas des gènes mitochondriaux les phylogénèses réalisées à partir d'un gène unique (de taille inférieure à 2 kb) semblent manquer de réalité et de résolution (51). Différents gènes peuvent générer des topologies différentes (41, 42, 47) car des séquences trop courtes, ou hautement conservées, peuvent diminuer l'information phylogénétique nécessaire à la détection des branches courtes. Les séquences uniques peuvent également être sujettes à des contraintes dont le résultat pourra être le biais du signal phylogénétique. Tous ces phénomènes entraînent des distorsions de l'arbre (36, 37).

## **Le phénomène LBA**

Ce phénomène (Long Branch Attraction) a une grande importance car il peut entraîner la formation d'artefacts dans les reconstructions phylogénétiques. Pour Felsenstein (88) il doit être considéré comme un biais de topologie et Brown et coll. (34) ont identifié ce biais au supplément de fréquence avec lequel ces branches longues sont associées.

Afin de juger de l'effet LBA un arbre est construit par la méthode ML en utilisant un modèle de substitutions supportant un taux égal entre tous les sites. Cet arbre est alors comparé à celui obtenu en utilisant d'autres méthodes dans lesquelles, soit une partie des sites est considérée comme invariable, soit en admettant que les taux pour tous les sites suivent une distribution de type gamma, soit encore en considérant une partie des sites comme invariants et l'autre partie suivant une distribution gamma. Les indices de vraisemblance de ces arbres sont comparés avec les résultats d'arbres obtenus par la méthode NJ en utilisant les p-distances. Généralement ces tests montrent que la vraisemblance des arbres ML augmente si l'hétérogénéité des taux entre sites est prise en compte.

Lorsque les taux de substitutions sont très différents entre les lignés étudiées il est important d'utiliser un modèle représentant les probabilités actuelles des substitutions afin d'obtenir une topologie correcte (191, 326, 334, 377). Par exemple dans les phylogénèses utilisant les rRNA la première violation du modèle concerne le contenu en Guanine et Cytosine (129). Toutefois si toutes les lignées étudiées présentent le même biais de ce contenu en Guanine et Cytosine la topologie reste constante (184) et dans le cas contraire il est nécessaire d'utiliser un modèle sensible au biais de ce facteur, par exemple la 'Transversion parcimony' (369), ou la 'Det Log distance' (191). Un autre facteur qu'il est nécessaire de prendre en compte est la variation du taux d'évolution entre sites (RAS ou Rate Among Sites) (249, 377).

Certaines méthodes, telle la méthode HOP (High Order Parsimony), introduisent des corrections dans les formules MP et sont en mesure de réduire le phénomène LBA. Dans le 'taxon sampling' (269) les taxa sont sélectionnés de manière telle que leurs positions 'casse' les branches longues. Il est également possible d'attribuer des poids différents à certains types de substitutions: dans la 'Transversion parsimony' (326) les transversions ont un poids différent de celui des transitions. Enfin Sankoff (283) a proposé un système beaucoup plus général pour attribuer des poids aux caractères. Ce système serait en mesure d'apporter une solution au phénomène LBA.

On peut enfin noter qu'il existe aussi un phénomène appelé LBD (Long Branch Distracts) (102, 264).

## Méthodes complémentaires

### Interior branch test

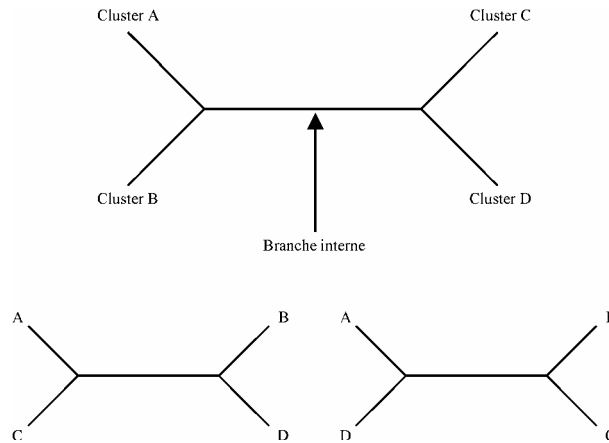


Figure 4 : Interior branch test

Le test 'least squares' permet l'estimation de la longueur et de la variance de la branche interne d'une topologie construite à partir de quatre séquences ou quatre clusters de séquences. Le test 'Interior branch' est alors construit afin d'examiner si la longueur de la branche interne est statistiquement différente de zéro (271) en utilisant le calcul 'Interior branch' de Sitnikova et coll. (304), tel qu'il est implémenté dans le logiciel PHYLTEST. Le complément de la probabilité ( $CP = 1 - \alpha$ ) d'une erreur de type 1, mesurée au moyen d'un test 'two tailed normal deviate' est calculé et plus la valeur de CP est élevée, plus la crédibilité de la branche interne est forte.

### Four cluster analysis

Cette méthode a été mise au point par Rzhetsky et coll. (275) et elle fournit un moyen efficace de comparer les trois topologies possibles pouvant être obtenues à partir de quatre groupes monophylétiques de séquences. Dans ces trois topologies la somme des branches entre chaque cluster est égale et leurs contributions se trouvent annulées lorsque les différences sont calculées. Il est ainsi possible, en utilisant le test 'Interior branch' d'obtenir trois valeurs de probabilités, et donc trois valeurs de CP dont la plus élevée indique la meilleure topologie.

## Relative rate test

Ce test est le test 'deux clusters' de Takezaki et coll. (335) dans lequel la constance de l'horloge moléculaire est évaluée entre deux lignées, en utilisant une troisième lignée comme outgroupe. Si  $L_A$  et  $L_B$  représentent les moyennes des nombres de substitutions par site (soit la longueur des branches) calculées à partir de l'ancêtre commun des lignées A et B, l'hypothèse nulle du test est que  $L_A - L_B = 0$ .

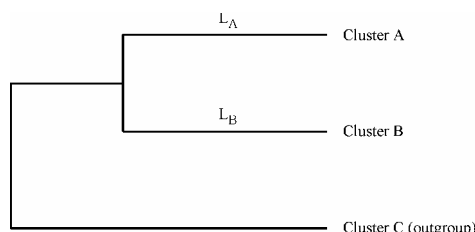


Figure 5 : Relative rate test

Cette hypothèse est posée en admettant la constance de l'horloge moléculaire. On peut alors calculer  $\delta = L_A - L_B = 0$ . Comme il est possible d'estimer la variance de  $\delta$ , on peut tester la déviation de ce facteur par rapport à la valeur nulle et ainsi évaluer la constance du taux d'évolution entre les lignées A et B en utilisant un test 'two tailed normal deviate'. Lorsque l'outgroupe est choisi proche des deux lignées à tester, cela rend le test statistique plus performant dans le rejet de l'hypothèse nulle.

## Arbres linéarisés

La méthode des arbres linéarisés a été décrite par Takezaki et coll. (335). Dans un premier temps une série de tests sont conduits afin d'éliminer les séquences hétérogènes, c'est à dire celles pour lesquelles le taux d'évolution est significativement différent de la moyenne des taux d'évolution. Puis, en utilisant une topologie donnée (généralement obtenue par la méthode NJ) les longueurs des branches sont estimées sous l'hypothèse d'un taux d'évolution constant. Certaines branches peuvent alors avoir une longueur négative (ceci étant du aux méthodes de calcul) et dans ces cas leur longueur est considérée comme nulle, ce qui conduit à une multifurcation dans la topologie. Un tel arbre linéarisé peut alors être utilisé afin d'estimer les dates de divergence entre séquences ou entre clusters de séquences. Cette méthode devient difficilement applicable lorsque le nombre de séquences traitées est important.

## Niveau d'information phylogénétique

Les distances génétiques sont parfois utilisées afin de juger de la présence, ou de l'absence, d'un signal phylogénétique dans une séquence donnée. Dans le cas du cytochrome b, et de celui d'autres gènes mitochondriaux, cette approche peut poser un certain nombre de problèmes. En effet des distances génétiques élevées peuvent indiquer l'existence d'une saturation des gènes et conduire à la conclusion de leur non utilisation pour des analyses phylogénétiques. Certains auteurs, dont Meyer (214), ont suggéré que l'existence de distances génétiques élevées (supérieures à 15 %) disqualifiaient un gène chez la grenouille. Par contre Graybeal (121) avait trouvé que pour des distances génétiques comprises entre 15 % et 20 % les séquences protéiques déduites des séquences nucléotidiques étaient pratiquement invariables. Pour le cytochrome b, et en utilisant des distances

non corrigées, les valeurs calculées entre n'importe lequel des *Lemur* et n'importe lequel des *Loris* sont au maximum de 21 % et de 25 % au moins pour les comparaisons effectuées entre Strepsirhini et Anthropoïdes (390). Pour les Strepsirhini la plupart des distances inter-taxa sont comprises entre 15 % et 20 %, ce qui indique sans doute une saturation de la troisième position des codons (214). Par contre les distances calculées à partir des séquences protéiques sont en moyenne 1,7 fois plus faibles, mais elles ne sont pas dramatiquement basses, comme cela est le cas chez les grenouilles (121).

## **Datation des séquences**

La datation d'une divergence dépend de l'attribution de dates fiables à au moins l'un des nœuds de l'arbre phylogénétique. Chez les Mammifères, le peu de fossiles disponibles fait que les datations ne peuvent pas toujours être très précises. Si dans un premier temps la calibration de l'horloge moléculaire pour les branches impliquées dans la divergence était nécessaire pour effectuer la datation, des travaux récents ont suggéré la possibilité de ne pas prendre en compte cette horloge moléculaire (146, 268, 281, 351).

Dans le cas des Primates, on peut tenir compte d'une date d'origine comprise entre 80 millions d'années (211) et de 63 à 70 millions d'années (108). La branche séparant les Strepsirhini du nœud correspondant au Primate ancestral représente seulement un faible pourcentage de la longueur cumulée des branches menant aux différents Strepsirhini. Une telle constatation peut indiquer que les Strepsirhini, ou du moins la grande majorité de ceux-ci, ont évolué dès leur divergence d'avec les Anthropoïdes.

Après avoir évalué les taux relatifs de substitutions dans différentes lignées Arnason et coll. (14) estiment que Hominoïdes et Chimpanzés ont divergé il y a 6,1 millions d'années (de 5,2 à 6,9 pour un intervalle de confiance de 95 %). Pour *Pan troglodytes* et *Pan paniscus* la divergence daterait de 3 millions d'années (de 2,4 à 3,6). Ces résultats suggèrent que la séparation *Gorille / Pan - Homo* serait intervenue il y a 8,4 millions d'années (de 7,3 à 9,4). Toutes ces estimations sont indépendantes de l'utilisation de fossiles hominoïdes et d'interprétations morphologiques, elles utilisent le '*Phoca* standard' en tant que référence externe pour la calibration.

Chez les Métazoaires les séquences de 12S rRNA ont été largement utilisées pour rechercher la date d'origine de ce groupe (94, 178, 357), ainsi que pour estimer la date de la première évolution des Eucaryotes (309, 310). L'utilisation de ces séquences implique les hypothèses formulées dans les différents modèles de reconstruction qui, si elles sont valables pour des espèces proches, ne le sont plus pour des espèces distantes. Dans le cas de séquences distantes il est alors possible d'utiliser des gènes, dits fondamentaux, dans lesquels les taux de substitutions nucléotidiques à différents sites sont fortement variables du fait de l'existence de domaines structurels et fonctionnels (régions variables et conservées). Si l'on ignore les variations de taux ceci affecte dramatiquement l'analyse phylogénétique avec une sous estimation de la longueur des branches et du rapport Transitions / Transversions (104, 333, 359, 375), dans ce cas la topologie est sans doute également affectée. Afin de contourner ce problème les distances utilisées peuvent être estimées en utilisant un modèle de type gamma (154, 186, 337).

Une autre hypothèse couramment utilisée considère que les taux de substitutions sont homogènes dans le temps et qu'ils sont stationnaires en suivant le même modèle (aléatoire) dans les différentes lignées considérées. Toutefois cette hypothèse n'est pas très réaliste lorsque l'on s'adresse à des séquences très distantes. Pour les reconstructions effectuées à partir de séquences de

12S rRNA les topologies sont perturbées si les fréquences des bases présentent des valeurs très différentes entre les lignées étudiées (129, 193). Par contre si des génomes ont acquis, indépendamment, des compositions en bases identiques, les méthodes de parcimonie tendent à suggérer l'existence d'un arbre faux dans lequel sont regroupées les séquences ayant des composition en bases identiques, plus que celles ayant une histoire génétique commune (192, 317, 363). Ce problème des fréquences inégales est confondu avec le phénomène LBA et les topologies ne peuvent être utilisées pour des datations. En choisissant les taxa de manière à 'casser' les branches longues le problème est généralement surmonté (381).

Pour Hasegawa et coll. (130) les arbres reconstruits à partir de séquences protéiques seraient plus fiables, les fréquences des acides aminés étant plus constantes que celles des nucléotides (1, 193). Ils pourraient fournir un support fiable pour des datations.

## ***Reconstruction de séquences ancestrales***

Les séquences protéiques actuelles peuvent être utilisées pour reconstruire des séquences ancestrales (79, 258), et ce en utilisant une approche par parcimonie (4, 153, 206). La reconstruction des séquences ancestrales est également possible à partir de séquences nucléotidiques.

Il est ainsi possible d'estimer les voies de substitution des nucléotides, ou des acides aminés, à chacun des sites des séquences. Une telle approche peut être utilisée afin d'identifier des nucléotides, ou des acides aminés, spécifiques susceptibles de provoquer des changements fonctionnels du gène étudié, ainsi que pour détecter des évolutions convergentes ou des sélections Darwiniennes positives (318, 322).

Un certain nombre d'algorithmes de reconstruction ont été développés par Fitch (95) pour les arbres enracinés et bifurqués, et par Hartigan (128) pour des topologies générales. La validité de la reconstruction est généralement inconnue, sauf dans le cas où les séquences analysées sont très proches.

Dans les modèles qui sont utilisés avec la méthode ML les états des caractères dans les séquences ancestrales sont vus comme des variables aléatoires (89, 114). L'approche par vraisemblance de Yang et coll. (380) utilise les longueurs des branches de la topologie des séquences actuelles et les modèles de substitutions définis pour la reconstruction des séquences ancestrales. Ces reconstructions peuvent alors être effectuées par l'une des deux méthodes suivantes:

- **Reconstruction marginale**: un nucléotide, ou un acide aminé, est associé à chaque nœud interne de la topologie et à chaque site, en tenant compte de ses 'posterior probabilities'.
- **'Joint reconstruction'**: dans cette méthode on considère l'attribution d'un jeu de caractères à tous les nœuds interne de la topologie et à chaque site. Dans le logiciel PAML, si il y a moins de 15 nœuds internes pour un arbre obtenu à partir de séquences nucléotidiques, ou moins de 7 pour un arbre obtenu à partir de séquences protéiques, toutes les reconstructions possibles sont évaluées. Dans le cas contraire un algorithme de parcimonie est utilisée, générant à chaque nœud le 'equivocal character state' qui est ensuite utilisé pour la reconstruction des séquences ancestrales.

Ces deux méthodes sont sensées fournir des résultats identiques, mais la reconstruction marginale semble être la plus fiable.

La connaissance de la, ou des séquences ancestrales est d'une grande importance en biologie. Par exemple dans le cas du lysozyme (318) il a été possible d'identifier des sites pour lesquels le même acide aminé intervenait indépendamment dans toutes les lignées menant du bœuf au langur.

### ***Evolution générale des séquences***

Komiyama et coll. (174) ont émis l'hypothèse que d'importantes modifications, ou le réajustement de la fonction d'une protéine, peuvent être le résultat de l'évolution de seulement un petit nombre de sites. Un tel processus a pu être décrit dans le cas de l'hémoglobine des crocodiles, des lysozymes de l'estomac (175, 318), de certains pigments visuels (15, 332, 338).

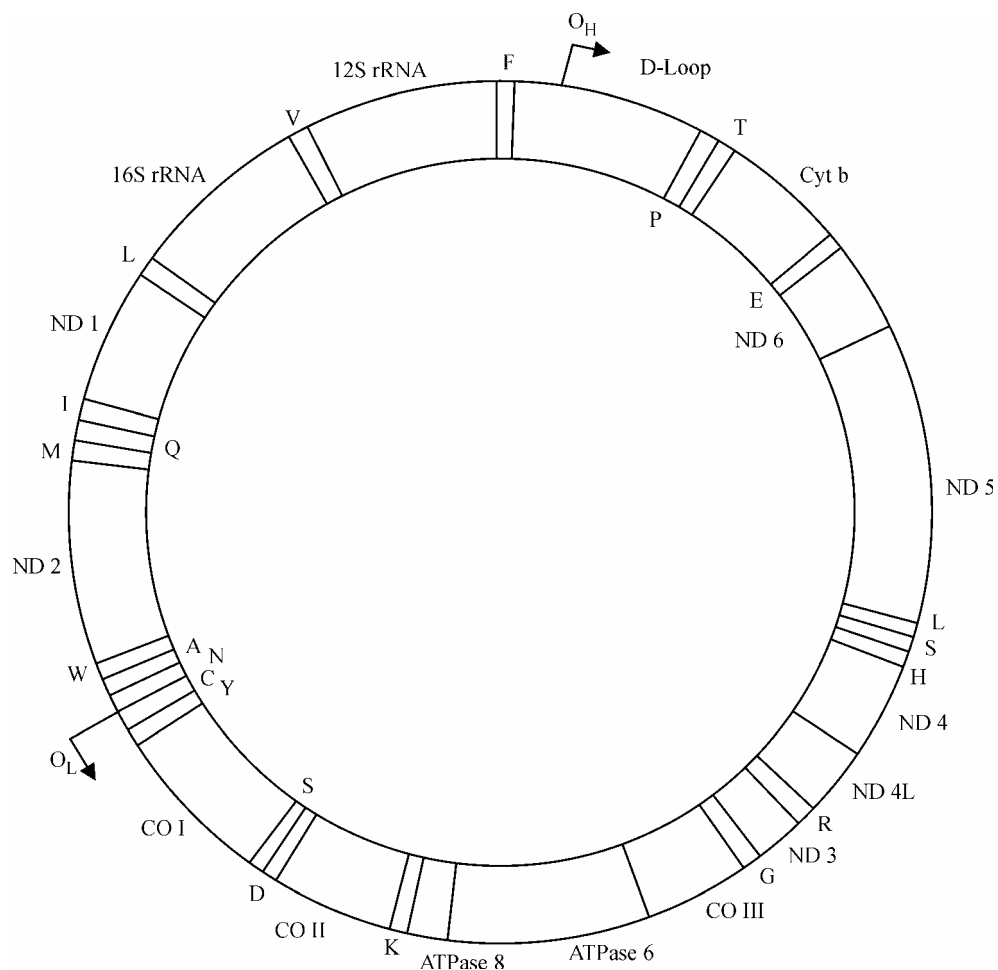
On admet généralement que la genèse d'une nouvelle séquence est le résultat du clivage en deux d'une lignée ancestrale, les deux nouvelles lignées divergeant ensuite graduellement. D'autre part les relations d'évolution ainsi que les facteurs modelant les séquences nucléotidiques sont des événements qui ne peuvent être facilement dissociés du fait de leur interaction.

# L'ADN MITOCHONDRIAL

## Généralités

Les gènes présents dans l'ADN mitochondrial des animaux supérieurs sont fortement conservés. A quelques exceptions près ce génome circulaire dont la taille est comprise entre 16 000 et 17 000 paires de bases code pour 13 protéines, 2 ARN ribosomiaux et 22 ARN de transfert. Généralement il existe une région non codante présentant des fonctions de régulation (CR ou Control Region, ou D-Loop), ainsi qu'une faible quantité d'ADN non codant.

Chez l'homme la cartographie du mtDNA a pu être établie (**Figure 6**):



**Figure 6 : Cartographie de l'ADN mitochondrial humain**

Chez la plupart des Vertébrés pour lesquels la totalité du génome mitochondrial a pu être séquencé, on constate une forte conservation de l'ordre des gènes (67) et les réarrangements possibles constituent des événements évolutifs uniques pouvant être utilisés pour des reconstructions phylogénétiques, en particulier lorsque les séquences analysées appartiennent à des taxa distants (29, 306).



Le génome mitochondrial est transmis à la manière d'un locus non recombinant. Tous les gènes mitochondriaux portent une seule histoire d'évolution et, théoriquement, devraient fournir des arbres phylogénétiques identiques lorsque les séquences analysées proviennent d'un même taxon.

Les forces agissant sur le génome mitochondrial peuvent avoir pour résultat l'établissement de taux de substitutions différents entre les taxa. Dans ce contexte interviennent les taux de mutation ainsi que les taux de fixation des mutations qui, tous deux, se trouvent sous l'influence de facteurs multiples (169, 219). Il semble également que le génome nucléaire puisse influencer l'évolution du génome mitochondrial en modifiant les contraintes sélectives globales de ce dernier.

D'une manière générale l'évolution de l'ADN mitochondrial est plus rapide que celle de l'ADN nucléaire (367), ce qui fait que ce génome a été largement utilisé dans des études génétiques de différenciation des populations (77).

## **Caractéristiques**

Les travaux de Gissi et coll. (109) ont permis de montrer que le taux d'évolution du génome mitochondrial était variable, aussi bien entre les différents ordres, qu'à l'intérieur même de ces ordres, et dans certains cas entre espèces proches ou très proches.

Les gènes mitochondriaux codant pour des protéines montrent une saturation de la troisième position des codons pour des distances génétiques (estimées suivant le modèle JC69) comprises entre 0,15 et 0,20 (5, 121, 214).

Chez les Mammifères les substitutions synonymes présentent un taux plus élevé dans le mtDNA que dans l'ADN nucléaire. De plus les taux de substitutions ne sont pas identiques pour tous les gènes mitochondriaux: par exemple le taux des mutations synonymes dans le gène NADH 4 (Nicotinamine Adenine Dehydrogenase sous unité 4) est seulement de 60 % du taux qui peut être calculé à partir des données des neuf gènes complets actuellement publiés.

## **Gènes étudiés**

### **Cytochrome b**

Le cytochrome b est l'une des sous unités du complexe du cytochrome dont une représentation est donnée à la **Figure 7**. La région de 357 bp étudiée dans notre travail est figurée en rouge dans ce schéma, elle correspond aux acides aminés 24 à 142 et comprend trois zones hélicales (notées I, II et III). Les acides aminés essentiels au fonctionnement de la molécule (ceux participant au site actif) sont localisés au niveau de l'hélice II, ou à proximité de celle-ci, aux positions 80, 83, 97 et 100 (soit les acides aminés 57, 60, 74 et 77 de la séquence analysée).

Traditionnellement le cytochrome b est considéré comme un marqueur génétique valable et utilisable à différents niveaux d'analyse phylogénétique (81, 216, 228, 391). Ce gène présente un taux d'évolution accéléré dans les lignées menant à l'Homme (149, 196). Il est constitué d'une mosaïque de zones conservées et de zones à plasticité variable. Comme pour d'autres protéines il semble qu'il existe ici une corrélation entre la localisation, le type fonctionnel de la protéine et le degré de conservation des acides aminés. Ainsi les régions transmembranaires sont relativement peu contraintes pour les fonctions spécifiques de la protéine, alors que les régions de la membrane

externe, ou interne, sont le siège des processus biologiques essentiels. Il en résulte que les régions transmembranaires du cytochrome b sont plus variables que les autres régions. En particulier dans les régions IV et IX au moins 43 % des sites peuvent changer une ou plusieurs fois, et la région III est la plus conservée de toute la protéine. Dans ce gène la troisième position des codons évolue plus rapidement que la seconde position, ceci implique l'existence d'une relation inverse entre la pression de sélection sur les acides aminés et le taux d'évolution. Toutefois on peut remarquer que la première position évolue aussi plus rapidement que la seconde, bien que pour ces deux sites la pression sur l'acide aminé correspondant soit sensiblement identique.

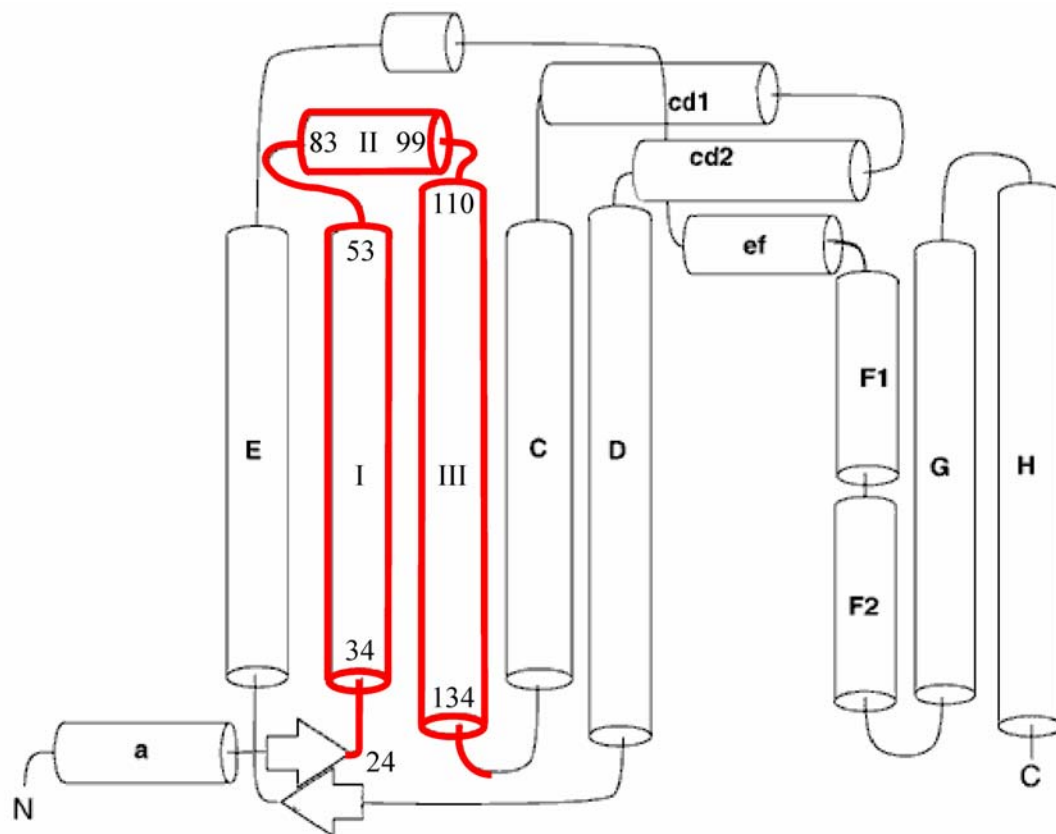


Figure 7 : Représentation schématique du cytochrome b

Le rapport Transitions / Transversions à une valeur voisine de 1,0 pour la première et la troisième position. D'autre part pour les premières et secondes positions les taux de substitutions sont très variables entre sites.

Comme dans le cas des autres Mammifères (149) les Primates présentent, à la première position des codons une distribution à peu près identique des quatre nucléotides, alors qu'à la seconde position il existe une préférence pour la Thymine et à la troisième position une forte sous représentation de la Guanine. En outre le cytochrome b des Mammifères montre un fort biais des Guanines dans la chaîne L (ce qui est typique du mtDNA) et ceci peut poser un problème lors des reconstructions phylogénétiques car la surabondance de l'état d'un caractère augmente la tendance des sites présentant ce caractère à se saturer prématurément (149, 214). D'autre part le biais des mutations peut conduire à une violation des hypothèses de parcimonie (260). En effet si il existe une mutation maintenant la Guanine à une fréquence plus faible que celle de l'Adénine, la vraisemblance d'une transformation de A en G sera plus faible que celle d'une transformation de G en A.

Chez tous les Primates la seconde position des codons est celle présentant les variations les plus faibles, alors que la troisième position montre les variations les plus importantes. Toutes ces variations peuvent être corellées avec le degré de contrainte sélective agissant sur ces sites.

Le taux d'évolution, de même que les transitions et les transversions, peuvent varier en fonction de la position sur les codons. D'autre part des caractères dont l'évolution est rapide (transitions, troisième position), du fait de leur fort taux de substitutions, peuvent être particulièrement sujets à l'homoplasie résultant de substitutions multiples.

Les caractères pour lesquels la sélection est intense (les secondes positions par exemple) sont tellement contraints qu'ils ne peuvent être informatifs pour les phylogénèses, ou bien les substitutions observées à ces sites reflètent plus l'existence de contraintes de sélection similaires que l'existence d'une histoire commune (235).

Chez les Primates sur les 453 sites phylogénétiquement informatifs du cytochrome b, 103 sont localisés au niveau des premières positions, 48 au niveau des secondes positions et 302 au niveau des troisièmes positions. Toutes les substitutions affectant les secondes positions conduisent à un changement d'acide aminé alors que cela est le cas pour seulement 30 % des substitutions affectant les troisièmes positions.

Les phénomènes de transversions à la troisième position, ainsi que tous les changements sur les secondes positions, doivent être considérés comme 'conservatifs', dans le sens où ils sont relativement libres des effets de saturation; et le degré avec lequel ils reflètent la phylogénèse et la sélection est différent (390).

## **12S et 16S rRNA**

Les séquences nucléotidiques correspondant aux 12S et 16S rRNA ont été utilisées avec succès pour la résolution de nombreux problèmes phylogénétiques (20, 50, 151, 215, 217, 218, 222).

Comme ces gènes codent pour des ARN ribosomiaux, de plus nombreuses variations par rapport à un gène codant pour une protéine sont possibles. Toutefois ces variations sont soumises à des pressions sélectives destinées à maintenir une structure secondaire correcte de la molécule nécessaire au bon fonctionnement des ARN. Il existe également dans ces gènes des blocs très conservés pouvant être utilisés pour la construction de primers (172).

Les gènes rRNA de l'ADN mitochondrial évoluent de deux à trois fois plus rapidement que les gènes codant pour des protéines, dans la même molécule d'ADN mitochondrial.

## LES LEMURIFORMES ACTUELS ET SUB-FOSSILES

Les Lemuriformes forment avec les Lorisiformes le groupe des Strepsirhini. A la suite des travaux de Grandidier et Milne-Edwards, à la fin du XIX<sup>ème</sup> siècle et au début du XX<sup>ème</sup> siècle, les bases essentielles de la taxonomie des Primates malgaches furent établies par (290). Puis l'utilisation de techniques cytogénétiques a permis à Bender et Chu (23), et à Rumpler et Albignac (273) d'étudier de manière plus précise la systématique ainsi que l'évolution des Lémuriens. Quelques années plus tard Dutrillaux (73) réalise une reconstruction phylogénétique de l'ensemble des Primates en utilisant les informations fournies par les réarrangements chromosomiques pouvant être détectés au cours de l'évolution des caryotypes.

Les Lémuriens malgaches comprennent 36 espèces classées en 14 genres et 5 familles. Ces animaux montrent une très large palette de morphologies, de styles de vie, ainsi que de comportements. En outre parmi les sub-fossiles qui ont pu être collectés on a dénombré au moins 16 espèces de Primates appartenant à au moins 7 genres (111, 160, 300, 301). Plusieurs de ces espèces étaient encore présentes à Madagascar il y a 500 ans (110), ce qui tend à faire considérer ces sub-fossiles, au sens de l'évolution, comme contemporains des Lémuriens actuels.

Les relations phylogénétiques qui ont pu être proposées pour ce groupe de Primates endémiques ne sont pas toujours concordantes, tant pour celles basées sur des données de biologie moléculaire, que pour celles utilisant des caractères morphologiques ou caryologiques.

Au cours des années 70, toute une série d'articles furent publiés concernant l'unité phylogénétique des Primates malgaches (43, 46, 292, 327, 343). Ces données ont pu être confirmées à la suite d'études récentes de biologie moléculaire (5, 58, 158, 265, 384, 390). Après une vingtaine d'années de controverses ces phylogénies apparaissent maintenant comme robustes. L'un des points problématiques que l'on pouvait alors relever concernait la position phylogénétique du *Daubentonia* (Aye-Aye) qui était apparu pour la première fois dans les classifications en 1870 (312) et était alors classé avec les Tarsiers. Puis en 1795, St. Hilaire en fait un Rongeur et c'est en 1800 que Cuvier crée le genre *Cheiromys*. Depuis qu'il est relié au groupe des Primates certains auteurs l'ont placé dans une sous famille des Indriidae (291, 314, 343), ou encore comme un infra- ou un sub-ordre séparé (103, 123, 150, 155, 262, 354), ou encore une super-famille séparée (77). A l'heure actuelle il est classé dans la famille des Daubentoniidae qui est un infra-ordre des Lemuriformes (134, 261, 341). De nombreuses données cytologiques (74, 274) ainsi que des travaux de phylogénèses moléculaires (65, 66, 265) indiquent que le *Daubentonia* est le taxon le plus basal parmi les Lemuriformes. Toutefois il faut noter qu'une étude de (5), utilisant des séquences de CO II (Cytochrome Oxidase sous unité II,) est en contradiction avec ce résultat et place le *Daubentonia* en tant que groupe frère de tous les autres Strepsirhini, avec toutefois un faible support statistique de cette topologie. Dans le travail de Yoder et coll. (390) l'utilisation de séquences du cytochrome b place le *Daubentonia* à la base de tous les Primates malgaches et à la base des Strepsirhini lorsque des séquences de CO II sont utilisées. La combinaison des deux types de séquences, en utilisant un rapport de Transitions / Transversions de 1,0 permet de montrer que le clade des Primates malgaches, dans lequel est inclus le *Daubentonia*, présente un support de bootstrap de 53 % si toutes les positions des codons sont utilisées, et de 93 % si seules les troisièmes positions sont prises en compte. Ce résultat tendrait à prouver l'existence de contraintes fonctionnelles au niveau des acides aminés.

L'un des points apparaissant comme le plus constant dans les phylogénèses moléculaires est le classement des Primates malgaches dans un groupe monophylétique (28, 65, 66, 176, 265).

Toutefois certains travaux, tels ceux de Stanger-Hall et Cunningham (316) dans lesquels sont utilisées des séquences de 16S rRNA, ainsi que les positions 1 et 2 des codons du cytochrome b et du CO II, fournissent des résultats contradictoires: le groupe est monophylétique dans les reconstructions effectuées à partir des séquences de 16S, et diphylétique avec les positions 1 et 2 du cytochrome b et du CO II. Certains auteurs avaient d'ailleurs montré, auparavant, que pour le cytochrome b et le CO II il pouvait exister d'importantes différences dans les phylogénèses obtenues à partir de l'une ou l'autre des trois positions des codons (5, 371). Si les phylogénèses des Lémurs ne fournissent pas une organisation monophylétique il faut admettre qu'il a existé entre le continent Africain et Madagascar au moins deux migrations ayant concouru au peuplement de l'île (43). Mais un ensemble de données, tant paléogéographiques (256, 267) que paléontologiques (107, 108, 211) tendent à montrer que l'éventualité des migrations multiples est peu probable. Lorsque les caractères morphologiques sont utilisés les résultats sont quelque peu différents, en effet trois des quatre études actuellement disponibles indiquent une paraphylie, ou une diphylie, des Strepsirhini. Par contre la plupart des données génétiques vont dans le sens d'une monophylie des Lemuriformes avec inclusion des Cheirogaleidae dans ce clade (5, 28, 57, 66, 176, 265, 274, 384, 390).

Sur la base des données morphologiques les Cheirogaleidae, Galagidae et Lorisidae présentent des relations spéciales (43). Dans le cas des Cheirogaleidae et Lorisidae les individus sont de petite taille, insectivores ou omnivores, nocturnes et ils présentent généralement un bas taux métabolique. Ces deux groupes pourraient alors constituer un clade séparé de celui des autres Primates malgaches. Mais des travaux portant sur certaines régions variables du génome semblent fournir des résultats en contradiction avec cette hypothèse et il faudrait admettre que les similitudes observées ne seraient que des convergences ou des rémanences d'un Strepsirhine ancestral (46). Cette dernière hypothèse est clairement favorisée par le critère de parcimonie. Toutefois des travaux plus récents (5, 65, 66, 74, 176, 265, 274, 385) tendent à montrer que les Cheirogaleidae doivent être inclus dans les Lémurs malgaches.

La majorité des études réalisées à ce jour indiquent l'existence de problèmes concernant la résolution de certains nœuds de l'arbre (316, 388). Si différents schémas, attribuant des poids différents aux caractères étudiés, sont utilisés, chacun d'eux favorisera une topologie différente pour les nœuds posant des problèmes. Des analyses utilisant le principe du maximum de parcimonie se heurtent aux mêmes problèmes, mais les arbres sont généralement mieux résolus que dans le cas de ceux utilisant les troisièmes positions des codons. A une exception près tous les nœuds entièrement résolus sont en accord avec les phylogénèses obtenues antérieurement pour les Mammifères (44, 149, 265). L'exception concerne le *Daubentonia*. La position de *Varecia* est également quelque peu problématique: il a été placé en dehors des Lemuridae (198, 315); à la base de tous les Lemurids (5, 58, 384, 390); en tant que groupe frère des *Eulemur* (124, 125, 342); ou encore en position non résolue par rapport aux *Lemur*, *Hapalemur* et *Eulemur* (58, 316). D'autre part pour Yoder et coll. (390) *Propithecus* apparaît étant le clade frère des Cheirogaleidae, alors que *Varecia* est le clade frère de l'ensemble (Cheirogaleidae, *Propithecus*, autres Lémurs). Mais si l'analyse est effectuée en attribuant comme poids aux différentes positions l'inverse du nombre de caractères génétiquement informatifs, et en standardisant les valeurs pour les troisièmes positions (soit un schéma 3 : 6 : 1) *Varecia* et *Propithecus* se trouvent placés comme précédemment, mais les Primates anthropoïdes forment un groupe frère du *Daubentonia* seul, cette association étant supportée par une valeur élevée de bootstrap (86 %).

Dans le cas des Strepsirhini les fossiles disponibles sont peu nombreux et la plupart de ceux que l'on connaît actuellement sont postérieurs à la date de séparation qui peut être fixée entre Galagidae et Lorisidae (199). L'utilisation du Primate ancestral pour les calibrations dans ce groupe pose également des problèmes car sa date d'apparition est assez incertaine (entre 80 et 63 millions

d'années). Un autre événement qui pourrait être pris en compte pour les datations est le taux de changement des nucléotides que Irwin et coll. (149) ont estimé, chez les Artiodactyles, à 0,5 % par million d'années et Smith et Patton (306) à 1,7 % chez les Rongeurs murins. La voie la plus sûre est alors d'accepter l'estimation la plus conservative, c'est à dire celle donnant la date de divergence la plus récente pour l'origine des Primates. Cette date peut alors être utilisée pour calibrer l'horloge moléculaire et ce en considérant que les analyses de distances génétiques fournissent des valeurs qui sont relatives aux temps de divergences. Ainsi en fixant l'origine des Primates à 63 millions d'années, l'origine des Strepsirhini peut être évaluée à 61 millions d'années lorsque l'on utilise les informations contenues dans les trois positions des codons; ou à 62 millions d'années si l'on ne prend en compte que les troisièmes positions de ces codons. L'une ou l'autre de ces estimations place cette origine au milieu du Paléocène.

Les travaux de Yoder (383) et Yoder (384) ont permis d'établir un modèle parcimonieux de l'origine des Lémuriens dans lequel le premier événement rencontré est la séparation du *Daubentonia* des autres Lemuriformes, puis intervient la séparation entre les Cheirogaleidae et les autres Lémurs. D'autre part en prenant en compte les travaux les plus récents il est possible de définir, ou de préciser, les points suivants:

- ✓ *Varecia* se trouve en position basale de la famille des Lemuridae et il existe un fort support statistique de cette organisation.
- ✓ *Hapalemur* et *Lemur catta* forment un clade qui est le clade frère de celui regroupant les individus du genre *Eulemur*, et cette disposition présente un fort support statistique.
- ✓ Les sous espèces de *Eulemur fulvus* forment un groupe monophylétique regroupant trois lignées distinctes.
- ✓ Dans toutes les études pour lesquelles un poids égal est attribué à tous les caractères *Hapalemur coronatus* apparaît comme étant le taxon frère de *Eulemur macaco*.
- ✓ Les positions de *Eulemur rubriventer* et *Eulemur mongoz* ne peuvent jamais être définies avec un fort support statistique. Toutefois il semble que *E. mongoz* soit le plus basal dans le genre *Eulemur*.

Les trois classifications suivantes montrent les différentes modifications apportées, au fil de l'évolution des techniques d'étude. La première est celle de Osman-Hill (1953) reprise d'après celle établie par Simpson (302):

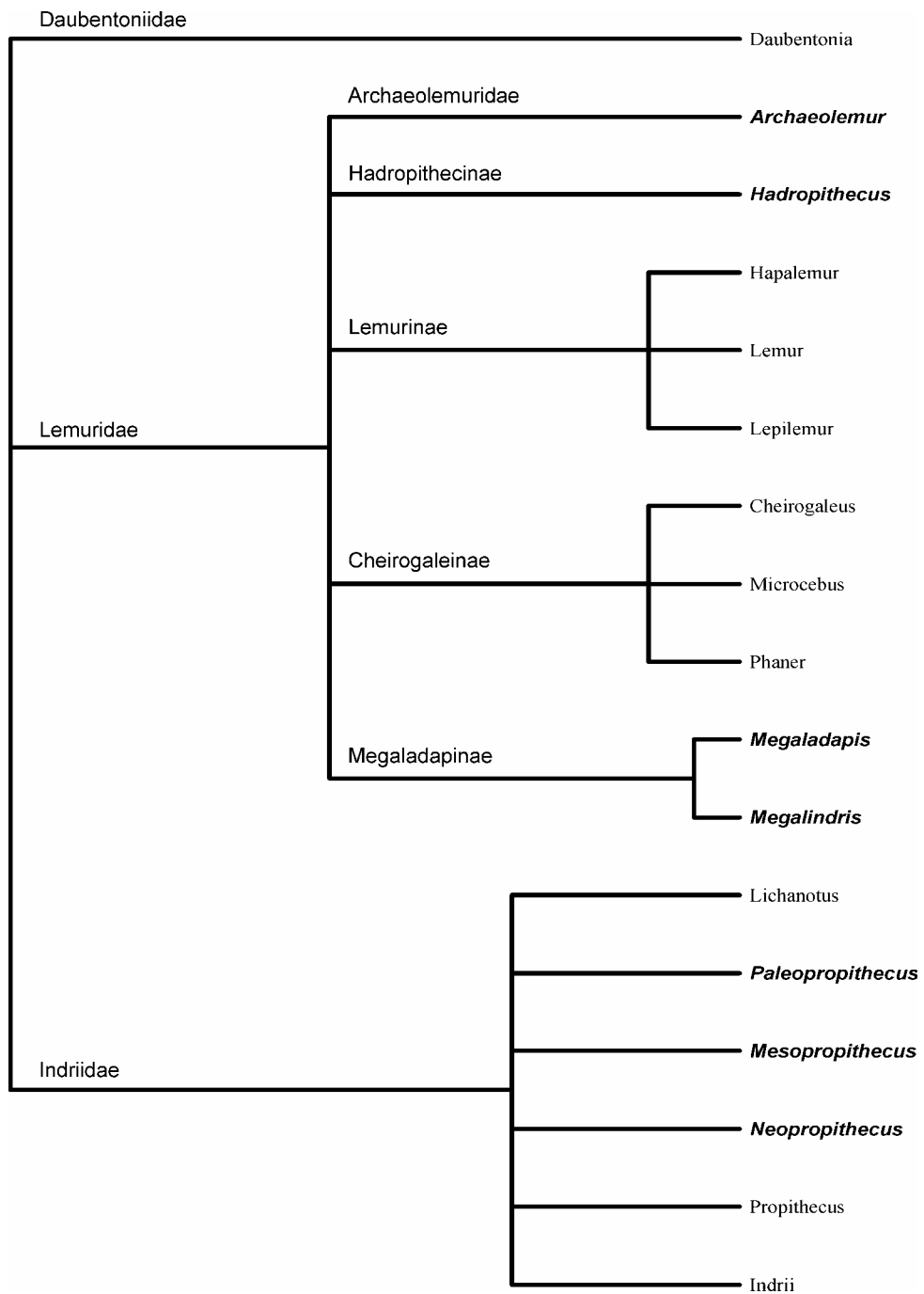


Figure 8 : Classification des Lemuriformes de Osman-Hill (1953)

La seconde est tirée des travaux de Petter et coll. (261):

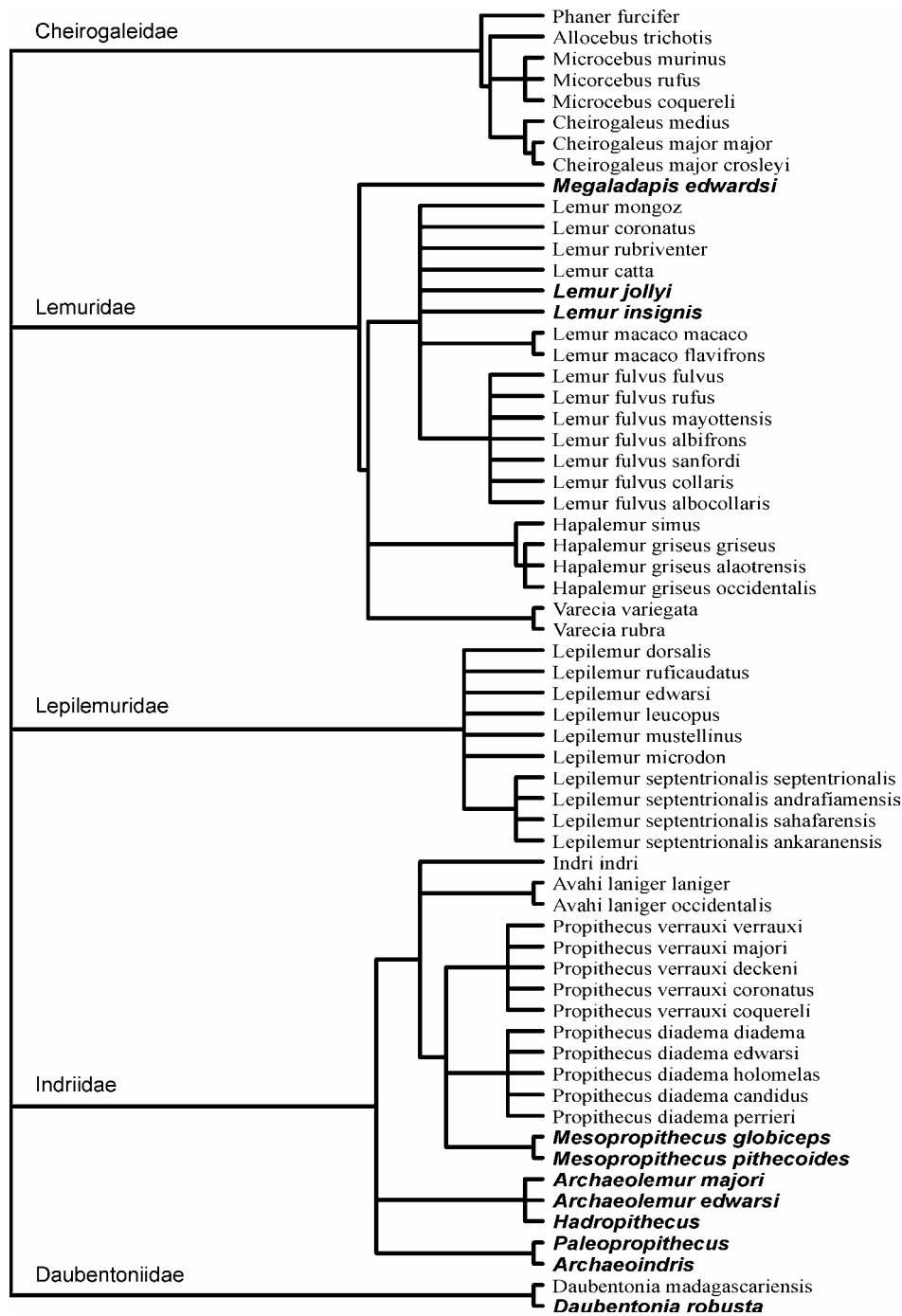


Figure 9 : Classification des Lemuriformes de Petter et al. (1977)



La troisième représente une compilation des données les plus récentes:

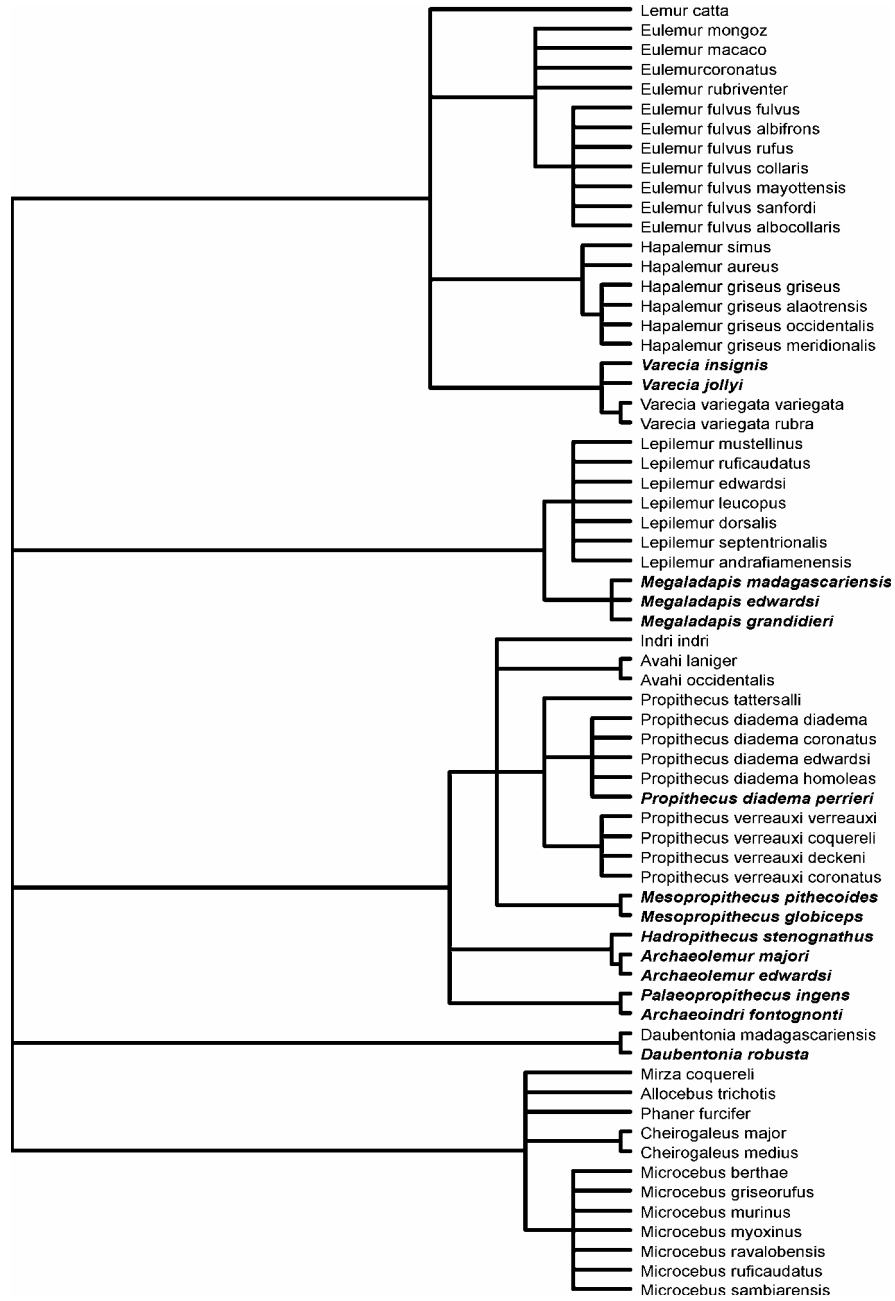


Figure 10 : Classification générale des Lemuriformes

Dans les trois classifications précédentes les formes sub-fossiles sont indiquées en caractères gras-italique.

Les références de première description des familles, sous-familles, genres et taxa pour le groupe des Lemuriformes sont les suivantes:

**Famille** Lemuridae (Gray, 1821)

**Sous-famille** Lemurinae (Gray, 1821)

**Genre** Lemur (Linnaeus, 1758)

*L. catta* (Linnaeus, 1758)

**Genre** Eulemur (Simons)

*E. mongoz* (Linnaeus, 1766)

*E. macaco* (Linnaeus, 1766)

*E. fulvus* (Geoffroy, 1796)

*E. f. fulvus* (Geoffroy, 1796)

*E. f. albifrons* (Geoffroy, 1796)

*E. f. rufus* (Audebert, 1800)

*E. f. collaris* (Geoffroy, 1812)

*E. f. mayottensis* (Schlegel, 1886)

*E. f. sanfordi* (Archbold, 1932)

*E. f. albocollaris* (Rumpler, 1975)

*E. coronatus* (Gray, 1942)

*E. rubriventer* (Geoffroy, 1850)

**Genre** Hapalemur (Geoffroy, 1851)

*H. griseus* (Link, 1793)

*H. g. griseus* (Link, 1793)

*H. g. alaotrensis* (Rumpler, 1975)

*H. g. occidentalis* (Rumpler, 1975)

*H. g. meridionalis* (Warter)

*H. simus* (Gray, 1870)

*H. aureus* (Meier)

**Genre** Varecia (Gray, 1863)

*V. variegata* (Kerr, 1792)

*V. v. variegata* (Kerr, 1792)

*V. v. rubra* (Geoffroy, 1812)

*V. insignis* † (Filhold, 1895)

*V. jollyi* † (Standing, 1908)

**Famille** Lepilemuridae (Rumpler, 1972)

**Sous-famille** Lepilemurinae (Rumpler, 1972)

**Genre** Lepilemur (Geoffroy, 1851)

*L. mustellinus* (Geoffroy, 1851)

*L. ruficaudatus* (Grandidier, 1867)

*L. dorsalis* (Forsyth-Major, 1894)

*L. leucopus* (Forsyth-Major, 1894)

*L. edwardsi* (Forsyth-Major, 1894)

*L. septentrionalis* (Rumpler, 1975)

*L. andrafiamenensis* (Rumpler, 1975)

**Sous-famille** Megalapinae † (Forsyth-Major, 1894)

**Genre** Megaladapis † (Forsyth-Major, 1894)

*M. madagascariensis* † (Forsyth-Major, 1894)

*M. edwardsi* † (Grandidier, 1899)

*M. grandidieri* † (Standing, 1903)

**Famille** Indriidae (Bennett, 1828)

**Sous-famille** Indriinae (Bennett, 1828)

**Genre** Indri (Geoffroy, 1796)

*I. indri* (Gmelin, 1788)

**Genre** Avahi (Jourdan, 1834)

*A. laniger* (Gmelin, 1788)

*A. occidentalis* (Lorenz, 1898)

**Genre** Propithecus (Bennett, 1832)

*P. diadema* (Bennett, 1832)

*P. d. diadema* (Bennett, 1832)

*P. d. candidus* (Grandidier, 1871)

*P. d. edwardsi* (Grandidier, 1871)

*P. d. homoleas* (Gunther, 1875)

*P. d. perrieri* † (Levauden, 1931)

*P. verreauxi* (Grandidier, 1867)

*P. v. verreauxi* (Grandidier, 1867)

*P. v. coquereli* (Milne-Edwards, 1867)

*P. v. deckeni* (Peters, 1870)

*P. v. coronatus* (Milne-Edwards, 1871)

*P. tattersalli* (Simons)

**Genre** Mesopropithecus † (Standing, 1905)

*M. pithecoides* † (Standing, 1905)

*M. globiceps* † (Lamberton, 1936)

**Sous-famille** Archaeolemurinae † (Standing, 1908)

**Genre** Archaeolemur † (Filhold, 1895)

*A. majori* † (Filhold, 1895)

*A. edwardsi* † (Filhold, 1895)

**Genre** Hadropithecus † (Lorenz, 1899)

*H. stenognathus* † (Lorenz, 1899)

**Sous-famille** Palaeopropithecinae † (Tattersall, 1973)

**Genre** Palaeopropithecus † (Grandidier, 1899)

*P. ingens* † (Grandidier, 1899)

**Genre** Archaeoindris † (Standing, 1908)

*A. fontognonti* † (Standing, 1908)

**Famille** Daubentoniidae (Gray, 1870)

**Genre** Daubentonia (Geoffroy, 1795)

*D. madagascariensis* (Gmelin, 1788)

*D. robusta* † (Lamberton, 1934)

**Famille** Cheirogaleidae (Gregory, 1915)

**Sous-famille** Cheirogaleinae (Geoffroy, 1812)

**Genre** Cheirogaleus (Geoffroy, 1812)

*C. major* (Geoffroy, 1812)

*C. medius* (Geoffroy, 1812)

**Genre** Microcebus (Geoffroy, 1828)

*M. berthae*

*M. griseorufus*

*M. murinus* (Miller, 1777)

*M. myoxinus* (Peters, 1852)

*M. ravalobensis* (Zimmermann, 1998)

*M. ruficaudatus* (Lesson, 1840)  
*M. sambiarensis*  
**Genre** *Mirza* (Lesson, 1840)  
*M. coquereli* (Grandidier, 1867)  
**Genre** *Allocebus* (Peter-Rousseau, 1967)  
*A. trichotis* (Gunther, 1875)  
**Genre** *Phaner* (Gray, 1870)  
*P. furcifer* (Blainville, 1833)

On peut considérer que les études phylogénétiques deviennent de plus en plus précises et que leur fondement théorique s'accroît (62, 63, 181, 285) de même que leurs conditions d'utilisation (41, 373). Ainsi les classifications précédentes seront elles certainement modifiées ou précisées, au cours des prochaines années.

Les travaux de Yoder et coll. (390) fournissent une comparaison des valeurs relatives des branches inter-nodales (ce qui correspond à un temps de divergence relatif) chez les Strepsirhini. Lorsque sont prises en compte les trois positions des codons, ou lorsque l'analyse porte uniquement sur les troisièmes positions, la longueur de la branche séparant l'ancêtre des Strepsirhini de l'ancêtre des Primates représente seulement une faible fraction de la distance séparant le taxon des Strepsirhini de l'ancêtre des Primates (entre 1,7 % et 2,6 %). Deux explications peuvent être apportées à cette observation:

- Du fait de l'évolution rapide du mtDNA, le cytochrome b a pu accumuler un nombre suffisant de mutations pour que le gène entier se trouve saturé. Dans ce cas les algorithmes utilisés pour les estimations de distances ne sont plus en mesure de corriger correctement les données. La branche courte ne serait alors qu'un artefact.
- La branche courte peut être le reflet de l'existence d'un laps de temps relativement court séparant les ancêtres des Primates et des Strepsirhini.

Un certain nombre d'arguments permettent de privilégier la seconde hypothèse. En effet la topologie de l'arbre obtenu au moyen des estimations de distances est tout à fait concordante avec celle obtenue par la méthode MP, ainsi que celle obtenue par la méthode ML. D'autre part une analyse détaillée des changements relatifs des caractères par rapport aux distances génétiques montre que le cytochrome b est une molécule informative au niveau de la divergence Strepsirhini / Anthropoïdes. Mais il faut toutefois tenir compte du fait que ces lignées divergent etradient si rapidement que cela peut conduire à un manque de précision des reconstructions effectuées (303).

En ce qui concerne les formes sub-fossiles des études morpho-fonctionnelles ont permis de montrer qu'elles présentaient une plus grande diversité que les Lémurs actuels. Par exemple la forme *Megaladapis* apparaît comme beaucoup plus proche des koalas que de tous les autres Primates (159, 361). La forme *Palaeopropithecus* peut être reliée à un 'paresseux arboricole' (200) et la forme *Archaeoindri* à un 'paresseux terrestre' (159, 356). D'autre part sur la base d'études cranio-dentaires les genres *Archaeolemur* et *Hadropithecus* sont plus proches des singes que des Lémurs (205, 314). En ce qui concerne les *Daubentonia* des espèces plus grandes que celle vivant actuellement ont du exister dans le Sud-Ouest de Madagascar il y a environ 1 000 ans (118, 119, 179, 201), l'un des représentants de ces sub-fossiles est le *Daubentonia robusta*. Lorsque l'on cherche à étudier les relations phylogénétiques existant entre ces sub-fossiles et les Lémuriens actuels le problème de l'homoplasie, typique des primates malgaches endémiques (76), doit être pris en compte.

En fonction des données disponibles à l'heure actuelle il est possible de postuler un certain nombre d'hypothèses concernant la phylogénèse de ces sub-fossiles:

- ✓ *Palaeopropithecus*, *Mesopropithecus*, *Babakotia* et *Archaeoindri* constituent une famille monophylétique (*Palaeopropithecidae*) apparaissant comme étant le clade frère des Indriidae actuels.
- ✓ *Palaeopropithecidae* et Indriidae forment le groupe frère des Archaeolemuridae (160, 301).
- ✓ *Megaladapis* est proposé comme étant le groupe frère des Lepilemurs actuels (291).
- ✓ *Daubentonia robusta* et le *Daubentonia* actuel forment deux clades frères (299).
- ✓ *Pachylemur* est plus proche des *Varecia* que des Lemurs (56, 294).

Il est ainsi possible de représenter ces relations par l'arbre phylogénétique de la **Figure 11**, tiré du travail de Yoder et Irwin (388) et dans lequel les formes sub-fossiles sont indiquées en caractères gras-italique.

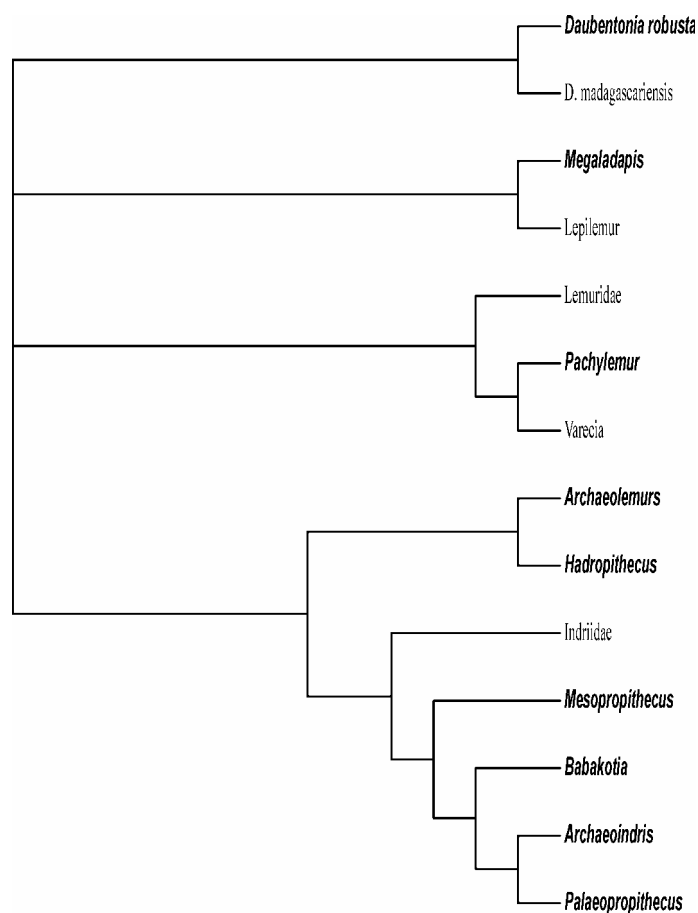


Figure 11 : Relations entre formes sub-fossiles et actuelles (d'après Yoder et Irwin – 1999)

Dans ce travail les auteurs ont étudié un fragment de cytochrome b provenant des genres *Palaeopropithecus*, *Megaladapis* et *Archaeolemur*. L'analyse effectuée par le maximum de parcimonie montre que si tous les taxa sont pris en compte *Palaeopropithecus* et *Megaladapis* forment un clade qui est basal par rapport à tous les autres Lemuriformes, à l'exception du *Daubentonia*. Mais un tel résultat n'est pas retrouvé lorsque sont prises en compte les données morphologiques. Cette association existe, bien que les séquences présentent une divergence de 5,5 % (de l'ordre de celle qui existe entre *Propithecus* et *Varecia*: 4,6 % par exemple). Ceci semblerait indiquer que ces deux taxa, tout comme les lignées de Lemuriformes actuels, ont débuté leur évolution indépendante il y a plusieurs millions d'années (386).

Comme nous l'avons vu précédemment il existe un certain nombre d'études systématiques réalisées sur différents groupes de Primates malgaches. Les reconstructions ont été effectuées en utilisant soit des données morphologiques, soit des données cytogénétiques, soit encore des données moléculaires (portant sur l'étude de l'ADN hautement répété et l'amplification de gènes nucléaires ou mitochondriaux). Dans la grande majorité des cas, et en particulier pour le traitement des données de biologie moléculaire, ces études portent sur un groupe particulier de Lémurs, mais il en existe très peu dans lesquelles sont pris en compte l'ensemble des Lémurs. Dans ce dernier cas chaque taxon n'est généralement représenté que par un seul individu, ou une séquence consensus.

Il nous a ainsi semblé intéressant d'effectuer des reconstructions, sur la base de données moléculaires, en utilisant toutes les espèces de Lémuriens pour lesquelles de telles données sont disponibles.

Nous avons utilisé, dans ce travail, deux segments de deux gènes mitochondriaux (un fragment de 357 bp issu du cytochrome b, et un fragment de 393 bp issu du 12S rRNA) qui ont pu être amplifiés pour les espèces pour lesquelles les données n'étaient pas disponibles sur la banque de données EMBL. Comme ces deux gènes peuvent avoir des évolutions différentes, les différentes reconstructions seront effectuées en prenant en compte l'un ou l'autre de ces gènes, ainsi que la combinaison des deux. Nous avons également pu amplifier les mêmes segments à partir d'une forme sub-fossile (*Megaladapis edwardsi*) et ces résultats sont utilisés dans notre analyse.

Les principaux objectifs de ce travail vont être de tenter de préciser un certain nombre de points sujets à controverse, en particulier concernant la position des *Tupaia*, du *Daubentonia*, des *Varecia*, ainsi que des *Lepilemurs*. L'utilisation d'un nombre important de taxa devrait être en mesure de produire des reconstructions 'fiables'. D'autre part l'étude de l'évolution des différentes séquences permettra d'effectuer des datations des principales divergences entre groupes.

## MATERIEL ET METHODES

### *Extraction de l'ADN*

L'ADN est extrait à partir de petits fragments d'oreilles prélevés sur le terrain, sous anesthésie générale au Ketalar. Ces fragments sont stockés à sec dans de l'azote liquide jusqu'à leur utilisation.

Après décongélation les fragments sont incubés pendant au moins 24 heures à 37 °C dans 300 µl de tampon d'extraction :

Tris-HCl 50 mM pH 8,0  
EDTA 10 mM pH 8,0  
SDS 0,5 %

auquel sont ajoutés 10 µl de protéinase K (solution à 10 mg/ml). L'ADN est ensuite extrait en utilisant le protocole phénol / chloroforme décrit par Sambrook et coll. (280), puis précipité par l'éthanol absolu en présence d'acétate d'ammonium 6 M (1/10 volume à volume). Le culot final d'ADN est lavé deux fois par l'éthanol 70°, puis il est repris par de l'eau bidistillée. La solution finale est ajustée à une concentration de 12 ng/µl.

### *Amplification des séquences mitochondriales*

#### **Cytochrome b**

Les primers utilisés pour l'amplification sont ceux décrits par Kocher et coll. (172); leurs séquences sont les suivantes:

P181 (L14841): 5' - **AAA AAG CTT CCA TCC ACC ATC TCA GCA TGA TGA AA** -  
3'

P182 (H15149): 5' - **AAA CTG CAG CCC CTC AGA ATG ATA TTT GTC CTC A** -  
3'

(Les positions sur les chaînes L et H sont données en référence à la séquence d'ADN mitochondrial humain)

L'amplification est réalisée dans un volume de 100 µl. Le milieu de réaction contient:

Tris-HCl 20 mM pH 8,4  
KCl 50 mM  
MgCl<sub>2</sub> 2 mM  
200 mM de chaque dNTP  
15 pM de chaque primer  
Taq DNA polymerase 5 U  
Albumine sérique bovine 500 µg  
ADN 100 ng

Après une phase de prédénaturation de 10 minutes à 94 ° C, 40 cycles d'amplification sont réalisés:

Dénaturation: 1 minute à 94 ° C

Association: 1 minute à 48 ° C

Elongation: 2 minutes à 72 ° C

L'amplification se termine par une phase d'élongation de 10 minutes à 72 ° C.

## **12S rRNA**

Les primers utilisés sont également dérivés de ceux décrits par Kocher et coll. (172). Ce sont:

RL211 (L1104): 5' - CTG GGA TTA GAT ACC CCA CTA A - 3'

RL212 (H1486): 5' - GAG GTG GAC GGG CGG TGT GT - 3'

Le milieu d'amplification est identique à celui utilisé pour le cytochrome b.

Après une phase de prédénaturation de 10 minutes à 94 ° C, 35 cycles d'amplification sont réalisés:

Dénaturation: 30 secondes à 94 ° C

Association: 45 secondes à 52 ° C

Elongation: 1 minute 30 secondes à 72 ° C

L'amplification se termine par une élongation de 10 minutes à 72 ° C.

## **Analyse des produits d'amplification**

Les produits d'amplification sont analysés sur des gels d'agarose 1,2 % (TBE 1X) contenant du bromure d'éthidium (0,5 µg/µl). Après une migration de 2 à 4 heures sous 150 volts, et examen sous lumière UV, les bandes amplifiées sont découpées et l'ADN est extrait de celles-ci par la méthode de freeze-squeeze décrite par Tautz et Renz (344).

## **Séquençage des produits d'amplification**

Les fragments amplifiés de cytochrome b et de 12S rRNA sont séquençés directement en utilisant les mêmes primers que ceux définis pour l'amplification.

Les séquences sont déterminées à l'aide d'un séquenceur automatique (ABI – Modèle 377) en utilisant le kit 'Taq Deoxy Terminal Polymerase Sequencing' avec de l'Ampli Taq Polymerase et quatre dideoxynucléotides marqués par des fluorochromes.



## **Programmes informatiques utilisés**

- ✓ **ANCESTOR**: Zhang J., Nei M. (1997) Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood and distance methods. *J. Mol. Evol.* 44 (Suppl. 1):S139-S146
- ✓ **ANC-GENE**: Zhang J., Nei M. (1997) Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood and distance methods. *J. Mol. Evol.* 44 (Suppl. 1):S139-S146
- ✓ **BIOEDIT**: Hall T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/97/NT. *Nucl. Acids Symp. Ser.* 41:95-98. (*Version 5.0.6.*)
- ✓ **BN-BS**: Zhang J., Rosenberg H.F., Nei M. (1998) Positive Darwinian selection after gene duplication in Primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA.* 95:3708-3713
- ✓ **CLUSTAL X**: Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., Higgins D.G. (1997). The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res.* 25(24):4876-4882
- ✓ **DAMBE**: Xia X. (2000) DAMBE: Data Analysis and Molecular Biology and Evolution. Department of Ecology and Biodiversity. University of Hong Kong
- ✓ **DNASP**: Rozas J., Rozas R. (1999) DnaSP *Version 3*: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15:174-175
- ✓ **GZ-GAMMA**: Gu X., Zhang J. (1997) A simple method for estimating the parameter of substitution rate variation among sites. *Mol. Biol. Evol.* 14:1100-1113
- ✓ **LINETREE**: Takezaki N., Rzhetsky A., Nei M. (1995) Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* 12:823-833
- ✓ **MEGA**: Kumar S., Tamura K., Jakobsen I.B., Nei M. (2001) Mega 2: Molecular Evolutionary Genetics Analysis software. *Bioinformatics*
- ✓ **MFOLD**: Zuker M., Mathews D.H., Turner D.H. (1999) Algorithms and Thermodynamics for RNA secondary structure prediction: a practical guide. In "RNA biochemistry and Biotechnology" Barciszewski J., Clark B.F.C. (eds.) NATO ASI Series. Kluwer Academic Press, New York
- ✓ **PAML**: Yang Z. (2000) Phylogenetic Analysis by Maximum Likelihood (PAML). *Version 3.0* University College London. London, England
- ✓ **PHYLIP**: Felsenstein J. (2000) *Version 3.6 (alpha)* Department of Genetics. University of Washington. Seattle WA, USA
- ✓ **PHYLTEST**: Kumar S. *Version 2.0* A program for testing phylogenetic hypothesis. Institute of Molecular Evolutionary and Department of Biology. The Pennsylvania State University. University Park. Pennsylvania, USA
- ✓ **TREE-PUZZLE**: Schmidt H.A., Strimmer K., Vingron M., von Haeseler A. (1999-2000) *Version 5.0*
- ✓ **TREEVIEW**: Page R.D.M. (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. *Comp. Appl. Biosci.* 12:357-358. *Version 1.6.5.* (2001)
- ✓ **WINCLADA**: Nixon H.C. (1999) WINCLADA (beta). *Version 0.9.9.* Published by the author. Ithaca, NY, USA

## Séquences utilisées

### Origine des séquences et groupes d'études

Le **Tableau 9** regroupe l'ensemble des séquences de cytochrome b et de 12S rRNA utilisées dans notre étude. Certaines ont été extraites de la banque de données EMBL (elles sont référencées par leurs numéros d'accès). Celles qui ont été déterminées au Laboratoire sont indiquées par '\*\*' (non encore déposées à EMBL), ou par un numéro d'accès suivi de '\*'. Lorsque, pour un même taxon, plusieurs séquences sont disponibles le logiciel BIOEDIT est utilisé pour construire une séquence consensus (consensus majoritaire).

Pour certaines analyses, les séquences de cytochrome b et de 12S rRNA ont été extraites des génomes mitochondriaux complets, déposés sur la banque EMBL. Les références de ces séquences sont indiquées dans le Tableau 10. D'autre part des séquences de Lorisiidae supplémentaires ont été utilisées pour les études de monophylie (**Tableau 11**).

Les séquences de Lemuriformes analysées dans cette étude sont réparties dans les groupes taxonomiques suivants: Eulemurs, Hapalemurs, Varecia, Indriidae, Lepilemuridae, Cheirogaleidae et Daubentoniidae. Deux séquences du groupe des Tupaia sont également utilisées. La composition de ces différents groupes est présentée dans le **Tableau 12**. De plus les séquences d'un sub-fossile (*Megaladapis edwardsi*), récemment déterminées au laboratoire (225) sont aussi incluses dans l'étude. Enfin au cours des traitements des associations de ces différents groupes ont été réalisées, elles sont indiquées dans le **Tableau 13**.

Espèce	Cytochrome b	12S rRNA
ALA	**	**
AOC	**	**
CMA	U53570	**
DMA	Af175840-Z34952-U53569	AF175781
ECO	AF175831-AF175832-Z34981	AF175772-AF175773
EFA	AF081048-AF175856	AF175797
EFC	AF175833-AF175834-U53576	AF175774-AF175774
EFF	AF175841-AF175842-Z38983	AF175782-AF175783
EFR	AF175854-AF175855-U53577	AF175795-AF175796
EMF	AF175835-AF175836-AF081050	AF175776-AF175777
EMM	AF175837-AF175849-Z34982-AF081049	AF175778-AF175790
EMO	AF175843-AF175844-Z35096-AF081051	AF175784-AF175785
ERU	AF175847-AF175848-Z35097-AF081052	AF175788-AF175789
HAU	**	**
HGA	**	**
HGG	**	**
HGM	**	**
HGO	**	**
HGR	Z34986-U53574-**	**
HSI	AF175859-AF175860-**	AF175800-AF175801-**
IIN	Z35098-AJ406044*	**
LCA	AF175838-AF175839-Z34980-AF175953 à AF175960-U53575-U38271	AF175779-AF175780
LDO	AJ406001 à AJ406014*-AJ243867 à AJ243973*- AJ007619*	AJ270669 à AJ270684*-AJ270611 à AJ270621*
LED	AJ405992 à AJ406000*	**
LLE	AJ406015 à AJ406021*-AJ243869	**

LMU	AJ007782*	**
LRU	AJ406022 à AJ406030*-AJ007781*	**
LSE	AJ304652 à AJ304778*-AJ406031 à AJ406039*-AJ243868*-AJ007620*	AJ270622 à AJ270668*
LTA	AF212963-U53581	AF212943
MBE	AF285540 à AF285543	-
MCO	U53571	**
MED	AJ278142*	**
MGR	AF285567 à AF285568	-
MMU	AF285557 à AF285566-U53572-**	**
MMY	AF285535 à AF285539	-
MRA	AF285529 à AF285532	-
MRU	AF285544 à AF285553	-
MSA	AF285554 à AF285556	-
MTA	AF285533 à AF285534	-
PDI	**	**
PTA	AF175853-AJ406044*-U53573	AF175793-AF175794-**
PVC	AF285528-**	**
TGL	Z34984	AF153003
TUB	AF217811	AF217811
VVR	AF175850-U55578	AF175791
VVV	AF175851-Z34986-AF081047	AF175792

Tableau 9 : Origine des séquences utilisées

Groupe	Référence	N° d'accès	Nom
Carnivora	HGY	X72004	<i>Halichoetus gryphus</i>
	PHI	X63726	<i>Phoca vitulina</i>
	CFA	U90639	<i>Canis familiaris</i>
	FCA	U20753	<i>Felis catus</i>
Primates	HLA	X99256	<i>Hylobates lar</i>
	PPY	D38115	<i>Pongo pygmaeus</i>
	GGO	D38114	<i>Gorilla gorilla</i>
	PTR	D38113	<i>Pan troglodytes</i>
Cetacea	BPH	X61145	<i>Balaneoptera physallus</i>
	BMU	X72204	<i>Balaenoptera musculus</i>
Rodentia	RNO	X14848	<i>Rattus norvegicus</i>
	MUM	AB042524	<i>Mus musculus</i>
Aves	FPE	AF090338	<i>Falco peregrinus</i>
	GGA	X62392	<i>Gallus gallus</i>
	CFR	Y18522	<i>Corvus frugileus</i>
	RAM	AF090399	<i>Rhea americana</i>
Hymenoptera	AME	L06178	<i>Apis mellifera</i>

Tableau 10 : Génomes mitochondriaux complets utilisés

Espèce	Nom	Cytochrome b	12S rRNA
GAD	<i>Galagoides demidoff</i>	AF271411	AF212945
GAL	<i>Galago alleni</i>	AF212966-AF235095	AF212946
GCR	<i>Galago crassicaudatus</i>	U53579	AF179889
GDE	<i>Galago demidovii</i>	AF271409	AF212947
GMA	<i>Galago matschei</i>	AF271409	-
GMO	<i>Galago moholi</i>	AF271410	AF212951
NCO	<i>Nycticebus coucang</i>	U53580	AF212942
PPO	<i>Perodicticus potto</i>	AF271413	-

Tableau 11 : Origine des séquences de Lorisiidae

Groupe	Identification	Séquences	Noms
<b>Eulemurs</b>	EUL	ECO	<i>E. coronatus</i>
		EFA	<i>E. fulvus albocollaris</i>
		EFC	<i>E. fulvus collaris</i>
		EFF	<i>E. fulvus fulvus</i>
		EFR	<i>E. fulvus rufus</i>
		EMF	<i>E. macaco flavifrons</i>
		EMM	<i>E. macaco macaco</i>
		EMO	<i>E. mongoz</i>
		ERU	<i>E. rubriventer</i>
<b>Hapalemurs</b>	HAP	HAU	<i>H. aureus</i>
		HGA	<i>H. griseus alaotrensis</i>
		HGG	<i>H. griseus griseus</i>
		HGM	<i>H. griseus meridionalis</i>
		HGO	<i>H. griseus occidentalis</i>
		HGR	<i>H. griseus rufus</i>
		HSI	<i>H. simus</i>
<b>Indriidae</b>	IND	ALA	<i>Avahi laniger</i>
		AOC	<i>Avahi occidentalis</i>
		IIN	<i>Indri indri</i>
		PDI	<i>Propithecus diadema</i>
		PTA	<i>Propithecus tattersalli</i>
		PVC	<i>Propithecus verreauxi coquerelli</i>
<b>Megaladapidae</b>		MED	<i>Megaladapis edwardsi</i>
<b>Lepilemuridae</b>	LEP	LDO	<i>L. dorsalis</i>
		LED	<i>L. edwardsi</i>
		LLE	<i>L. leucopus</i>
		LMU	<i>L. mustellinus</i>
		LRU	<i>L. ruficaudatus</i>
		LSE	<i>L. septentrionalis</i>
<b>Cheirogaleidae</b>	CHE	CMA	<i>Cheirogaleus major</i>
		MBE	<i>Microcebus berthae</i>
		MCO	<i>Microcebus coquerelli</i>
		MGR	<i>Microcebus griseorufus</i>
		MMU	<i>Microcebus murinus</i>
		MMY	<i>Microcebus myoxinus</i>
		MRA	<i>Microcebus ravalobensis</i>
		MRU	<i>Microcebus ruficaudatus</i>
		MSA	<i>Microcebus sambiarensis</i>
<b>Daubentoniidae</b>		DMA	<i>Daubentonia madagascariensis</i>
<b>Tupaia</b>	TUP	TGL	<i>T. glys</i>
		TUB	<i>T. belangeri</i>
<b>Varecia</b>	VAR	VVR	<i>V. variegata rubra</i>
		VVV	<i>V. variegata variegata</i>

Tableau 12 : Définition des groupes taxonomiques

Groupe	Nombre de séquences	Composition
EHP	17	Eulemurs + Hapalemurs
MGA	7	Megaladapis + Lepilemuridae
EHV	19	Eulemurs + Hapalemurs + Varecia
IMG	13	Indriidae + Megaladapis + Lepilemuridae
GRA	32	Eulemurs + Hapalemurs + Varecia + Indriidae + Megaladapis + Lepilemuridae
GRB	42	Eulemurs + Hapalemurs + Varecia + Indriidae + Megaladapis + Lepilemuridae + Cheirogaleidae
GRC	45	Eulemurs + Hapalemurs + Varecia + Indriidae + Megaladapis + Lepilemuridae + Cheirogaleidae + Daubentonia + Tupaia
DTU	3	Daubentonia + Tupaia

Tableau 13 : Composition des groupes d'étude

## Prétraitement des séquences

### Cytochrome b

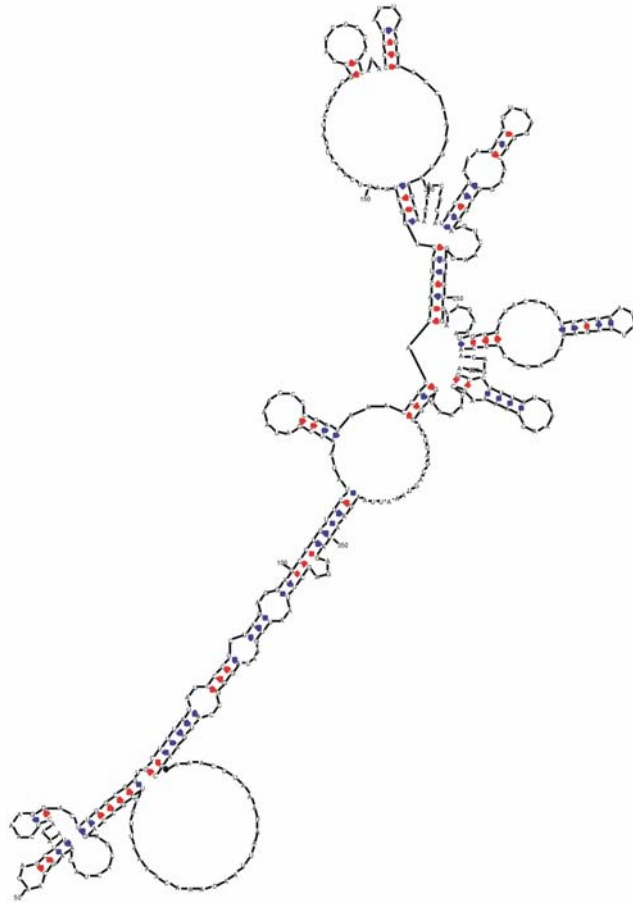
Le fragment de cytochrome b analysé a une taille de 357 bp, ce qui correspond à 119 codons entiers. Du fait de l'absence d'insertions ou de délétions ces séquences peuvent facilement être alignées sans l'aide d'un logiciel. Le programme MEGA est utilisé pour extraire les positions 1 et 3 des codons, ainsi que pour les traductions protéiques de ces séquences en utilisant le code génétique défini pour le génome mitochondrial des Mammifères. Les logiciels BIOEDIT et MEGA sont utilisés pour produire les fichiers d'alignements de ces séquences.

### 12S rRNA

Dans un premier temps ces séquences sont alignées en utilisant le logiciel CLUSTAL X. Ce traitement permet d'obtenir des séquences de 393 bp après introduction des 'gaps' nécessaires.

Les séquences alignées sont ensuite traitées dans le logiciel MFOLD qui permet la détermination de la structure secondaire de la molécule. Cette étude, dont le résultat est présenté à la **Figure 12**, permet de définir des zones pour lesquelles il n'existe pas de complémentarité (zones 'Loop'), et d'autres présentant cette complémentarité (zones 'Steem'). Le logiciel BIOEDIT est alors utilisé afin de produire, à partir des séquences totales, les deux types de séquences partielles: 213 bp pour les zones 'Loop' et 180 bp pour les zones 'Steem'.

pt22ps by D. Stewart and M. Zuker  
2001 Washington University



dG = -61.67 [initially -75.0] LCA2\_12S

**Figure 12 : Structure secondaire du fragment de 12S rRNA étudié**

# RESULTATS

## Caractéristiques générales des séquences

### Cytochrome b

#### Séquences nucléotidiques

##### Composition

Les caractéristiques de composition des différents groupes étudiés, ainsi que de l'ensemble des séquences sont indiquées aux **Tableaux 14** (séquences totales – 357 bp) et **15** (séquences partielles – comprenant uniquement les positions 1 et 3 des codons). Pour les deux types de séquences on note un faible pourcentage de Guanine: de 14,8 % à 17,5 % pour les séquences totales; et de 12,4 % à 16,0 % pour les séquences partielles. Ce biais de composition est une caractéristique générale des séquences mitochondriales. Si dans le cas des séquences totales les pourcentages d'Adénine, Thymine et Cytosine sont sensiblement identiques, dans le cas des séquences partielles le pourcentage d'Adénine est toujours supérieur à celui de la Thymine ou de la Cytosine. Ces différences se traduisent par un biais général de composition relativement faible (de 0,1007 à 0,1360) pour les séquences totales et plus élevé (de 0,1227 à 0,1713) pour les séquences partielles. Pour ce qui est des différents groupes il n'existe que très peu de variations dans les compositions nucléotidiques, sauf en ce qui concerne VAR et DMA pour lesquels le biais de composition est significativement plus élevé que la moyenne, tant pour les séquences totales que partielles.

Groupe	T	C	A	G	G + C	Biais
EUL	29,2	27,0	27,7	16,1	43,1	0,1126
HAP	32,3	24,9	26,7	16,2	41,1	0,1193
VAR	30,0	27,5	27,6	15,0	42,5	0,1340
IND	27,6	27,8	27,4	17,2	45,0	0,1040
LEP	27,8	28,2	26,6	17,5	45,7	0,1007
MED	28,3	28,0	26,3	17,4	45,4	0,1013
CHE	32,2	23,9	28,0	16,0	39,9	0,1153
DMA	29,7	28,0	27,5	14,8	42,8	0,1360
TUP	28,9	25,1	28,9	17,2	42,3	0,1073
Total	30,0	26,2	27,4	16,4	42,6	0,1147

Tableau 14 : Compositions nucléotidiques du cytochrome b (séquences de 357 bp)

Le biais est calculé en utilisant la formule:

$$\frac{2}{3} (|0,25 - \Pi_A| + |0,25 - \Pi_T| + |0,25 - \Pi_C| + |0,25 - \Pi_G|)$$

Groupe	T	C	A	G	G + C	Biais
EUL	25,3	27,5	33,1	14,0	41,5	0,1460
HAP	29,7	24,5	31,6	14,2	38,7	0,1507
VAR	26,1	28,6	33,0	12,4	41,0	0,1687
IND	22,8	28,9	32,6	15,7	44,6	0,1533
LEP	23,2	29,4	31,4	16,0	45,4	0,1573
MED	26,4	29,0	30,7	16,0	45,0	0,1340
CHE	29,7	22,9	33,4	17,9	40,8	0,1487
DMA	26,1	29,4	32,4	12,2	41,6	0,1713
TUP	25,2	25,0	34,0	15,8	40,8	0,1227
Total	26,5	26,4	32,6	14,5	40,3	0,1400

Tableau 15 : Compositions nucléotidiques du cytochrome b (positions 1 et 3)

Les pourcentages de sites variables, parcimonieux informatifs et singletons sont reportés dans les **Tableaux 16** (séquences totales) et **17** (séquences partielles). L'utilisation des seules positions 1 et 3 des codons fait passer le pourcentage de sites variables, pour l'ensemble des séquences, de 49,58 % (séquences totales) à 64,70 % (séquences partielles). Ce résultat montre que l'élimination des sites correspondant aux positions 2 des codons ne retire que peu d'information. La même constatation peut être faite en ce qui concerne les sites parcimonieux informatifs qui sont des sites pour lesquels il existe au moins deux changements nucléotidiques, chaque nucléotide étant représenté au moins deux fois. Par contre pour les sites singletons, qui sont des sites présentant un seul changement à un même site pour toutes les séquences analysées, le pourcentage est légèrement plus faible pour les séquences partielles que pour les séquences totales. Lorsque ces mêmes valeurs sont étudiées en fonction des différents groupes on peut remarquer d'importantes variations. Les groupes VAR, EUL et TUP sont ceux présentant les plus faibles pourcentages de sites variables, et le groupe CHE le pourcentage le plus élevé.

Groupe	Pourcentage de sites variables	Pourcentage de sites parcimonieux informatifs	Pourcentage de sites singletons
EUL	17,37	10,36	7,00
HAP	22,69	14,56	8,12
VAR	1,40	-	-
IND	24,93	15,69	9,24
LEP	7,56	2,24	5,32
CHE	28,85	19,61	9,24
TUP	5,6	-	-
Total	49,58	42,58	7,00

Tableau 16 : Caractéristiques des séquences cytochrome b (357 bp)

Groupe	Pourcentage de sites variables	Pourcentage de sites parcimonieux informatifs	Pourcentage de sites singletons
EUL	24,79	15,55	9,24
HAP	33,61	21,43	12,18
VAR	2,10	-	-
IND	35,29	22,69	12,60
LEP	8,82	2,94	5,88
CHE	42,44	28,99	13,44
TUP	8,40	-	-
Total	64,70	58,40	6,30

Tableau 17 : Caractéristiques des séquences cytochrome b (Positions 1 et 3)



## *Transitions et transversions*

Le **Tableau 18** montre qu'il existe des variations importantes dans la valeur du rapport des Transitions (Ti) sur les Transversions (Tv), aussi bien pour les séquences totales que pour les séquences partielles, lorsque ce rapport est calculé pour chaque groupe indépendamment. Toutefois si l'on prend en compte l'ensemble des séquences les valeurs ne sont que très peu différentes: 2,0 pour les séquences totales et 1,9 pour les séquences partielles. Les valeurs les plus faibles de ce rapport sont trouvées pour les groupes LEP et TUP (1,4 et 1,9 respectivement pour les séquences totales) et c'est le groupe EUL qui montre la valeur la plus élevée (10,0).

Groupe	Séquences totales (357 bp)			Séquences partielles (positions 1 et 3)		
	% <i>Ti</i>	% <i>Tv</i>	<i>R</i>	% <i>Ti</i>	% <i>Tv</i>	<i>R</i>
<b>EUL</b>	6,16	0,56	10,0	8,82	0,84	10,9
<b>HAP</b>	7,84	1,40	5,5	11,34	2,10	5,4
<b>VAR</b>	1,40	-	-	2,10	-	-
<b>IND</b>	9,80	2,24	4,4	14,28	2,94	4,8
<b>LEP</b>	1,68	1,40	1,4	2,10	1,26	1,6
<b>CHE</b>	9,52	2,24	4,4	13,86	2,94	4,6
<b>TUP</b>	3,64	1,96	1,9	5,46	2,94	1,9
<b>Total</b>	11,20	5,60	2,0	15,97	7,98	1,9

**Tableau 18 : Transitions et transversions pour les séquences cytochrome b**

A partir des comparaisons des séquences deux à deux pour chaque groupe comme pour l'ensemble des séquences, et en prenant en compte les compositions nucléotidiques de ces séquences il est possible, dans le logiciel DAMBE, de déterminer les types de transitions et de transversions, d'évaluer leur nombre réel en fonction des nucléotides impliqués et d'estimer quelle devrait être la répartition théorique de ces évènements. Les résultats concernant les transitions sont présentés aux **Tableaux 19** (séquences totales) et **20** (séquences partielles), et ceux concernant les transversions aux **Tableaux 21** (séquences totales) et **22** (séquences partielles).

Groupe	AG		CT	
	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>
<b>EUL</b>	170	104	611	183
<b>HAP</b>	152	108	624	200
<b>VAR</b>	3	1	2	1
<b>IND</b>	151	83	381	139
<b>LEP</b>	29	20	65	34
<b>CHE</b>	272	229	1236	387
<b>TUP</b>	7	3	6	4
<b>Total</b>	8584	7196	31128	12571

**Tableau 19 : Transitions observées et calculées pour le cytochrome b (357 bp)**

Groupe	AG		CT	
	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>
EUL	162	105	603	159
HAP	152	111	609	180
VAR	3	1	2	1
IND	151	86	362	112
LEP	29	17	49	24
CHE	272	234	1227	342
TUP	1	3	6	3
<b>Total</b>	8540	7296	29246	10786

Tableau 20 : Transitions observées et calculées pour le cytochrome b (positions 1 et 3)

Groupe	AC		AT		CG		GT	
	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>
EUL	52	174	4	188	21	101	1	110
HAP	55	165	45	215	21	100	21	130
VAR	0	1	0	1	0	1	0	1
IND	37	134	31	133	20	84	13	83
LEP	18	33	28	32	21	22	2	21
CHE	176	336	146	453	13	192	10	259
TUP	2	4	3	4	1	2	1	3
<b>Total</b>	9234	11471	6872	13154	1882	6873	1454	7886

Tableau 21 : Transversions observées et calculées pour le cytochrome b (357 bp)

Groupe	AC		AT		CG		GT	
	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>
EUL	52	209	4	192	13	88	1	81
HAP	55	191	45	232	21	86	21	104
VAR	0	1	0	1	0	0	0	0
IND	57	159	17	526	20	77	13	61
LEP	13	32	18	25	18	16	0	13
CHE	160	385	146	499	13	160	10	207
TUP	2	5	3	5	1	2	1	2
<b>Total</b>	2124	13278	6344	13301	1796	5916	1452	5926

Tableau 22 : Transversions observées et calculées pour le cytochrome b (positions 1 et 3)

On peut tout d'abord constater que les deux types de transitions ( $A \Leftrightarrow G$  et  $C \Leftrightarrow T$ ) surviennent plus fréquemment que prévu en fonction de la composition nucléotidique. Cette augmentation est très marquée pour les transitions Cytosine  $\Leftrightarrow$  Thymine et beaucoup plus faible pour les transitions Adénine  $\Leftrightarrow$  Cytosine, ce qui est à relier au fait que les Guanines sont les nucléotides présentant la plus faible fréquence. Cette constatation est valable tant pour les séquences totales que pour les séquences partielles.

Dans le cas des transversions il existe une hétérogénéité marquée: les transversions Adénine  $\Leftrightarrow$  Cytosine sont plus fréquentes que les trois autres types pour tous les groupes, sauf LEP et TUP chez lesquels ce sont les transversions Adénine  $\Leftrightarrow$  Thymine qui sont les plus fréquentes. D'autre part les transversions impliquant l'Adénine sont toujours plus fréquentes que celles impliquant la Cytosine, que ce soit pour les séquences totales ou pour les séquences partielles.

En fonction des caractéristiques de composition qui viennent d'être définies, le modèle de substitution TN93 sera utilisé pour les traitements portant sur les séquences de cytochrome b 357 bp, et le modèle K80 pour celles regroupant les positions 1 et 3 des codons.

### *Distances*

Le **Tableau 23** présente les distances de Tamura-Nei, prenant en compte les transitions et les transversions, calculées à partir des séquences consensus déterminées pour chaque groupe (partie inférieure de la matrice), à partir de l'ensemble des séquences de chaque groupe (partie supérieure de la matrice); ainsi que les moyennes intra groupe (diagonale de la matrice).

En utilisant les séquences consensus ces distances sont comprises entre 0,124 (entre EUL et VAR) et 0,317 (entre MED et TUP). En ce qui concerne la séquence MED (sub-fossile) la distance minimale (0,047) est trouvée pour la comparaison avec le groupe LEP. Les groupes DMA et TUP sont ceux présentant les distances les plus élevées par rapport aux autres groupes. Des résultats allant dans le même sens sont trouvés si l'on utilise cette fois l'ensemble des séquences disponibles pour chaque groupe.

Dans le cas des distances intra groupe la valeur la plus faible est trouvée pour VAR (0,014) et la plus élevée pour IND (0,142). On peut remarquer que ces valeurs sont nettement supérieures à 0,100 pour deux groupes: IND et CHE, ce qui peut être le signe d'une certaine hétérogénéité de ces groupes.

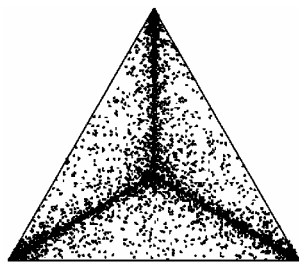
	<b>EUL</b>	<b>HAP</b>	<b>VAR</b>	<b>IND</b>	<b>LEP</b>	<b>MED</b>	<b>CHE</b>	<b>DMA</b>	<b>TUP</b>
<b>EUL</b>	<b>0,073</b>	0,167	0,152	0,193	0,207	0,194	0,212	0,244	0,277
<b>HAP</b>	0,168	<b>0,104</b>	0,180	0,197	0,203	0,201	0,221	0,257	0,315
<b>VAR</b>	0,124	0,173	<b>0,014</b>	0,188	0,185	0,178	0,215	0,260	0,274
<b>IND</b>	0,149	0,169	0,165	<b>0,142</b>	0,214	0,203	0,232	0,254	0,267
<b>LEP</b>	0,182	0,181	0,178	0,62	<b>0,031</b>	0,052	0,255	0,285	0,328
<b>MED</b>	0,172	0,188	0,176	0,164	0,047	-	0,246	0,279	0,339
<b>CHE</b>	0,182	0,201	0,178	0,196	0,220	0,215	<b>0,135</b>	0,269	0,299
<b>DMA</b>	0,236	0,250	0,250	0,243	0,270	0,279	0,256	-	0,275
<b>TUP</b>	0,252	0,302	0,274	0,231	0,292	0,317	0,302	0,261	<b>0,059</b>

**Tableau 23 : Distances génétiques pour les séquences de cytochrome b (357 bp)**

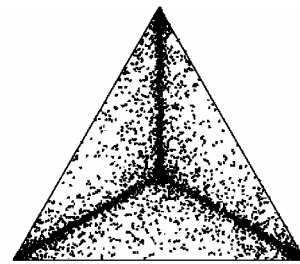
Une étude similaire a été réalisée en utilisant les séquences partielles (**Tableau 24**) et les distances de Kimura-2-paramètres prenant en compte les transitions et les transversions. Les valeurs de distances obtenues sont légèrement supérieures à celles calculées avec les séquences totales, mais les variations sont identiques pour les deux types de séquences.

	EUL	HAP	VAR	IND	LEP	MED	CHE	DMA	TUP
EUL	<b>0,109</b>	0,252	0,230	0,301	0,308	0,293	0,336	0,365	0,391
HAP	0,247	<b>0,159</b>	0,269	0,302	0,299	0,303	0,347	0,387	0,445
VAR	0,188	0,251	<b>0,021</b>	0,286	0,279	0,273	0,337	0,403	0,395
IND	0,230	0,250	0,248	<b>0,214</b>	0,318	0,308	0,365	0,384	0,373
LEP	0,278	0,283	0,275	0,240	<b>0,037</b>	0,070	0,389	0,432	0,472
MED	0,259	0,277	0,269	0,249	0,067	-	0,381	0,430	0,499
CHE	0,287	0,301	0,275	0,299	0,337	0,329	<b>0,205</b>	0,413	0,425
DMA	0,353	0,366	0,383	0,365	0,415	0,430	0,390	-	0,405
TUP	0,349	0,415	0,393	0,303	0,419	0,458	0,416	0,378	<b>0,090</b>

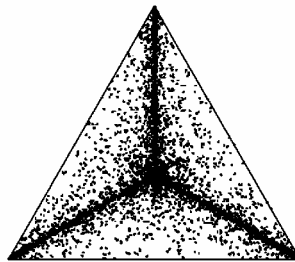
Tableau 24 : Distances génétiques pour les séquences de cytochrome b (positions 1 et 3)



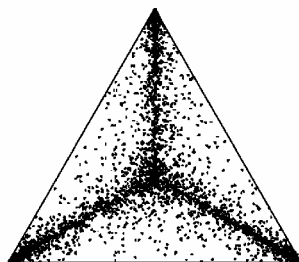
Cytochrome b 357 bp



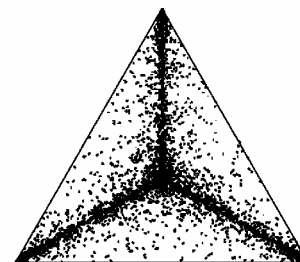
Cytochrome b Positions 1 et 3



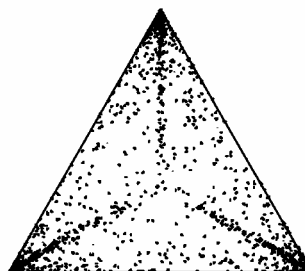
Cytochrome b séquences protéiques



12S 393 bp



12S Zones Loop



Cytochrome b + 12S rRNA

Figure 13 : Test 'Likelihood mapping' pour les différents types de séquences

### *Test 'Likelihood mapping'*

Pour les séquences totales le test est mené en utilisant les distances de Tamura-Nei, alors que pour les séquences partielles le modèle HKY85 a été choisi car le modèle K80 n'est pas disponible dans le logiciel TREE-PUZZLE. Les diagrammes obtenus avec les deux types de séquences sont présentés à la **Figure 13**. Les résultats chiffrés du test apparaissent dans le **Tableau 25**.

On peut constater que pour les séquences totales (cytochrome b 357 bp) et pour les séquences partielles (cytochrome b positions 1 et 3) le bassin 7, représentant les quartets non résolus contient environ 16 % des 10 000 quartets analysés. D'autre part environ 78 % des quartets se répartissent dans les bassins 1, 2 et 3 qui sont ceux pour lesquels les quartets sont entièrement résolus.

Ces tests indiquent que les deux types de séquences de cytochrome b peuvent être utilisées pour des études phylogénétiques.

	<b>Bassin 1</b>	<b>Bassin 2</b>	<b>Bassin 3</b>	<b>Total bassins 1, 2 et 3</b>	<b>Bassin 7</b>
<b>Séquences totales</b>	25,2	26,7	26,3	78,2	16,2
<b>Séquences partielles</b>	26,1	26,3	26,2	78,6	15,9

**Tableau 25 : Résultats du test 'Likelihood mapping' pour les séquences cytochrome b**

## **Séquences protéiques**

### *Composition*

Le **Tableau 26** indique les pourcentages des vingt acides aminés présents dans l'ensemble des séquences étudiées, ainsi que dans les différents groupes. Dans ce tableau les valeurs indiquées en rouge représentent les acides aminés dont les pourcentages s'écartent de l'intervalle (moyenne  $\pm$  écart type). On peut constater que les acides aminés les plus représentés sont Gly et Leu. Deux acides aminés, Pro et Glu montrent des pourcentages constants dans tous les groupes. Il existe également des acides aminés qui ne présentent que de faibles variations entre les différents groupes: Arg, Asn, Asp, Gln, Glu, His et Trp. Enfin Lys apparaît comme l'acide aminé le moins représenté, il n'est trouvé que dans les groupes IND et MED. L'acide aminé présentant le plus de valeurs en dehors de l'intervalle (moyenne  $\pm$  écart type) est Met. Enfin le groupe IND est celui présentant le plus d'acides aminés s'écartant de la moyenne (7 sur 20, soit 35 %), puis on trouve dans l'ordre EUL et DMA (6), TUP (5), MED (4), HAP et CHE (3) et VAR et LEP (2).

Acide aminé	Total	EUL	HAP	VAR	IND	LEP	MED	CHE	DMA	TUP
Ala	7,92	7,84	8,28	6,72	8,12	8,26	8,40	7,90	<b>5,04</b>	7,94
Arg	2,54	2,52	2,52	2,52	<b>2,66</b>	2,52	2,52	2,52	2,52	5,52
Asn	4,09	4,20	4,08	4,20	<b>3,64</b>	4,20	<b>3,36</b>	4,20	4,20	4,20
Asp	1,66	1,68	1,68	1,68	<b>1,54</b>	1,68	1,68	1,68	1,68	1,68
Cys	2,52	<b>2,43</b>	2,52	2,52	2,52	<b>2,66</b>	2,52	2,52	2,52	2,52
Gln	1,85	1,68	1,68	1,68	1,68	1,68	1,68	<b>2,44</b>	1,68	1,68
Glu	0,82	0,75	0,84	0,84	0,84	0,84	0,84	0,84	0,84	0,84
Gly	10,03	<b>10,27</b>	10,08	10,08	<b>9,52</b>	10,08	<b>9,24</b>	10,08	10,08	10,08
His	3,47	3,36	<b>3,84</b>	3,36	3,36	3,50	3,36	3,45	3,36	3,36
Ile	7,97	<b>9,15</b>	7,80	8,82	6,72	7,14	6,72	8,57	9,24	<b>5,04</b>
Leu	11,54	11,76	11,76	<b>12,61</b>	12,04	11,34	11,76	<b>10,50</b>	<b>13,45</b>	11,76
Lys	0,09	0	0	0	<b>0,56</b>	0	<b>0,84</b>	0	0	0
Met	4,93	<b>4,30</b>	<b>5,52</b>	<b>4,20</b>	4,76	5,32	5,04	5,13	<b>4,20</b>	<b>5,46</b>
Phe	8,20	8,40	8,40	8,40	7,70	8,12	<b>7,56</b>	8,40	<b>7,56</b>	<b>7,56</b>
Pro	1,68	1,68	1,68	1,68	1,68	1,68	1,68	1,68	1,68	1,68
Ser	7,90	7,56	7,68	7,56	8,40	<b>7,28</b>	8,40	8,24	<b>10,08</b>	7,56
Thr	8,25	<b>8,78</b>	<b>7,80</b>	9,24	<b>7,84</b>	8,54	8,40	7,98	8,40	7,98
Trp	5,06	<b>5,14</b>	5,04	5,04	5,04	5,04	5,04	5,04	5,04	5,04
Tyr	5,00	5,04	4,68	5,04	5,46	5,04	5,88	<b>4,37</b>	5,88	<b>6,72</b>
Val	4,46	3,45	4,08	3,78	<b>5,88</b>	5,04	5,04	4,45	<b>2,52</b>	<b>6,30</b>

Tableau 26 : Composition des séquences protéiques de cytochrome b

### *Distances*

Les distances pour ces séquences protéiques ont été calculées en utilisant le modèle de Poisson avec corrections. Dans la matrice présentée au **Tableau 27**, les valeurs correspondent aux calculs effectués sur les séquences des différents groupes (partie inférieure de la matrice), les séquences consensus de chaque groupe (partie supérieure de la matrice) et les moyennes intra groupe (diagonale de la matrice).

	EUL	HAP	VAR	IND	LEP	MED	CHE	DMA	TUP
EUL	<b>0,024</b>	0,052	0,034	0,079	0,079	0,106	0,070	0,097	0,135
HAP	0,055	<b>0,033</b>	0,061	0,043	0,079	0,106	0,079	0,135	0,144
VAR	0,043	0,055	<b>0,020</b>	0,088	0,079	0,106	0,079	0,079	0,125
IND	0,094	0,069	0,099	<b>0,060</b>	0,089	0,061	0,097	0,154	0,154
LEP	0,103	0,098	0,091	0,115	<b>0,035</b>	0,026	0,106	0,144	0,154
MED	0,114	0,108	0,102	0,030	0,043	-	0,135	0,134	0,184
CHE	0,078	0,083	0,081	0,109	0,123	0,123	<b>0,030</b>	0,144	0,144
DMA	0,104	0,127	0,092	0,147	0,160	0,174	0,141	-	0,154
TUP	0,138	0,145	0,125	0,150	0,174	0,184	0,137	0,149	<b>0,017</b>

Tableau 27 : Distances génétiques pour les séquences protéiques de cytochrome b

En utilisant pour chaque groupe la totalité des séquences disponibles la distance la plus faible est trouvée pour la comparaison entre MED et LEP (0,043) et la plus élevée pour la comparaison entre TUP et MED (0,184). Les distances sont toujours élevées lorsque la comparaison implique le groupe TUP. Lorsque les séquences consensus sont utilisées les distances calculées sont beaucoup plus faibles que précédemment, mais les variations observées sont les mêmes. Dans le cas des distances intra groupe la valeur la plus faible est trouvée pour le groupe TUP et la plus élevée pour le groupe IND.

Il faut remarquer que la traduction des séquences nucléotidiques fournit un certain nombre de séquences protéiques identiques. Ces identités sont trouvées à l'intérieur de certains groupes, mais jamais entre deux groupes différents. Dans le groupe EUL la traduction des séquences ECO, EFC et EMO conduit à des séquences protéiques identiques, de même que celle des séquences EFA, EFF et EFR. Dans le groupe HAP ce sont les séquences HGA, HGG et HGO qui sont identiques. Dans le groupe IND ce sont les séquences PTA et PVC; dans le groupe LEP les séquences LED et LLE; et enfin dans le groupe CHE les séquences MRU et MTA.

### *Test 'Likelihood mapping'*

Ce test est mené en utilisant le modèle de substitution mtREV24. Le diagramme obtenu par l'analyse dans TREE-PUZZLE est présenté **Figure 13** (cytochrome b séquences protéiques). A partir de l'analyse de 10 000 quartets on constate que 69 % de ceux-ci se répartissent dans les bassins 1, 2 et 3, alors que 24,9 % sont répartis dans le bassin 7. Ces séquences protéiques sont donc à la limite d'une utilisation pour des reconstructions phylogénétiques.

## 12S rRNA

### Composition

Les caractéristiques de composition des séquences de 12S rRNA sont indiquées dans les **Tableaux 28** (séquences de 393 bp), **29** (zones 'Loop') et **30** (zones 'Steem'). Dans le cas des séquences de 393 bp on peut noter une faible représentation de la Guanine, et une représentation assez forte de l'Adénine. Les pourcentages de Thymine et de Cytosine sont, quant à eux, sensiblement identiques. Cela se traduit par un biais général de composition, pour l'ensemble des séquences, de 0,1700.

Groupe	T	C	A	G	G + C	Biais
EUL	23,1	21,2	38,1	17,7	38,9	0,1740
HAP	27,2	21,1	39,0	16,7	37,8	0,1867
VAR	22,8	21,3	39,7	16,2	37,5	0,1960
IND	22,9	22,1	37,1	17,9	40,0	0,1613
LEP	21,1	24,2	37,3	17,3	41,5	0,1647
MED	21,0	24,1	36,7	18,1	42,2	0,1567
CHE	22,9	22,1	37,2	17,9	40,0	0,1620
DMA	22,2	23,5	35,0	19,3	42,8	0,1333
TUP	22,2	23,9	34,5	14,9	43,3	0,1267
<b>Total</b>	22,6	22,1	37,7	17,5	39,6	0,1700

**Tableau 28 : Composition nucléotidique des séquences 12S rRNA (393 bp)**

Groupe	T	C	A	G	G + C	Biais
EUL	18,2	20,4	52,7	8,7	29,1	0,3693
HAP	17,9	20,3	54,3	7,5	27,8	0,3907
VAR	18,9	18,6	55,1	7,4	26,0	0,4013
IND	18,7	20,4	50,7	10,2	30,6	0,3427
LEP	17,1	23,1	51,3	8,5	31,6	0,3507
MED	17,5	22,3	50,5	9,7	32,0	0,3400
CHE	20,3	20,0	43,3	10,5	30,5	0,3233
DMA	20,1	19,6	49,0	11,3	30,7	0,3200
TUP	18,8	22,5	40,3	12,5	35,0	0,2433
Total	18,3	20,8	51,9	9,0	29,8	0,3587

**Tableau 29 : Composition nucléotidique des séquences 12S rRNA (zones 'Loop')**

Groupe	T	C	A	G	G + C	Biais
EUL	28,6	22,1	21,5	27,8	49,9	0,0853
HAP	29,2	21,9	21,8	27,1	49,0	0,0560
VAR	27,4	24,3	22,1	26,3	50,6	0,0487
IND	27,6	24,1	21,4	26,8	48,9	0,0487
LEP	25,8	25,5	21,1	27,6	53,1	0,0520
MED	25,1	26,3	20,6	28,0	54,3	0,0587
CHE	25,8	24,5	23,4	26,4	50,5	0,0287
DMA	24,6	27,9	19,0	28,5	56,4	0,0853
TUP	26,1	25,6	21,0	27,3	52,9	0,0533
Total	27,5	23,7	21,5	27,3	51,0	0,0640

**Tableau 30 : Composition nucléotidique des séquences 12S rRNA (zones 'Steem')**

Dans le cas des zones 'Loop' la répartition est identique à celle trouvée pour les séquences de 393 bp, mais les différences sont plus marquées faisant que le biais, pour l'ensemble des séquences, a une valeur de 0,3587. Enfin pour les zones 'Steem' on trouve des pourcentages sensiblement égaux pour les quatre types de nucléotides avec une faible valeur de biais (0,064 pour l'ensemble des séquences), ce qui confirme le fait que l'on est bien en présence d'un ensemble de zones complémentaires deux à deux.

Lorsque les valeurs sont analysées en fonction des groupes on peut constater l'existence de peu de différences pour les séquences de 393 bp. Dans le cas des zones 'Loop' quelques petites différences peuvent être notées: une augmentation du pourcentage de Thymine chez CHE et DMA; une diminution de la Cytosine chez VAR, une diminution de l'Adénine chez TUP, et une augmentation de la Guanine chez DMA et TUP.

La répartition des sites variables, parcimonieux informatifs et singletons, est indiquée au **Tableau 31** pour les trois types de séquences. En tenant compte de l'ensemble des séquences analysées le pourcentage de sites variables le plus élevé est trouvé pour les zones 'Loop' et le plus faible pour les zones 'Steem'. La répartition est identique en ce qui concerne les sites parcimonieux informatifs. Les singletons, quant à eux, ne présentent que de faibles variations entre les trois types de séquences. Le groupe dans lequel est trouvé le plus fort pourcentage de sites variables et parcimonieux informatifs est IND et celui présentant les plus faibles valeurs est le groupe CHE.



Groupe	Séquences de 393 bp			Zones 'Loop'			Zones 'Steem'		
	% V	% Pi	% S	% V	% Pi	% S	% V	% Pi	% V
EUL	6,36	3,31	3,05	8,92	4,69	4,22	3,33	1,67	1,67
HAP	5,85	3,82	2,03	7,51	5,16	2,35	3,89	2,22	1,67
VAR	2,03	-	-	2,35	-	-	1,67	-	-
IND	13,74	6,11	7,63	17,84	8,45	9,39	8,89	3,33	5,55
LEP	8,65	3,82	4,83	8,45	3,75	4,69	8,89	3,89	5,00
CHE	1,78	-	1,53	2,35	-	1,88	1,11	-	1,11
TUP	6,11	-	-	6,57	-	-	5,55	-	-
<b>Total</b>	<b>41,47</b>	<b>34,35</b>	<b>6,61</b>	<b>47,89</b>	<b>38,97</b>	<b>7,98</b>	<b>33,89</b>	<b>28,89</b>	<b>5,00</b>

Tableau 31 : Caractéristiques des différents types de séquences 12S rRNA

## Transitions et transversions

Le **Tableau 32** montre l'existence de variations importantes du rapport  $T_i / T_v$  en fonction des différents groupes analysés. Cette constatation est valable pour les trois types de séquences. Dans tous les cas, sauf en ce qui concerne le groupe CHE, les transitions sont plus ou tout aussi fréquentes que les transversions. La valeur la plus élevée du rapport  $T_i / T_v$  est trouvée pour le groupe TUP avec les séquences de 393 bp (5,0), pour les groupes EUL et TUP avec les zones 'Loop' (3,8 et 3,7), et pour le groupe TUP avec les zones 'Steem' (9,0).

Groupe	Séquences de 393 bp			Zones 'Loop'			Zones 'Steem'		
	% $T_i$	% $T_v$	R	% $T_i$	% $T_v$	R	% $T_i$	% $T_v$	R
EUL	1,78	0,51	3,9	2,35	0,47	3,8	1,11	0,00	3,9
HAP	1,27	1,02	1,2	1,41	1,41	1,1	1,11	0,55	1,8
VAR	1,02	1,02	1,0	0,94	1,41	0,7	1,11	0,55	2,0
IND	4,83	1,27	3,6	6,10	1,88	3,5	2,78	0,55	4,0
LEP	2,80	0,76	3,3	2,35	0,94	2,1	3,33	0,55	6,8
CHE	0,25	0,76	0,3	0,47	0,94	0,5	0,00	0,55	0,0
TUP	5,09	1,02	5,0	5,16	1,41	3,7	5,00	0,55	9,0
<b>Total</b>	<b>7,63</b>	<b>3,56</b>	<b>2,1</b>	<b>8,92</b>	<b>4,69</b>	<b>1,8</b>	<b>6,11</b>	<b>2,22</b>	<b>2,8</b>

Tableau 32 : Transitions et transversions pour les différents types de séquences 12S rRNA

Les différents types de transitions (valeurs observées et calculées) sont indiqués aux **Tableaux 33** (séquences de 393 bp), **34** (zones 'Loop') et **35** (zones 'Steem'). Pour ce qui est des transversions, les résultats sont reportés dans les **Tableaux 36** (séquences de 393 bp), **37** (zones 'Loop') et **38** (zones 'Steem'). D'une manière générale les deux types d'évènements surviennent plus fréquemment que prévu. Toutefois dans le groupe HAP les transitions impliquant l'Adénine et la Guanine sont moins fréquentes que prévu pour les séquences de 393 bp, ainsi que pour les zones 'Loop'. Il faut également remarquer que les transitions impliquant la cytosine et la thymine surviennent plus fréquemment que celles impliquant l'Adénine et la Guanine. Dans le cas des transversions les valeurs observées sont nettement plus faibles que les valeurs qui peuvent être calculées. D'autre part les transversions impliquant l'Adénine sont plus fréquentes que celles impliquant la Cytosine.

Groupe	AG		CT	
	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>
EUL	93	49	111	34
HAP	20	46	121	33
VAR	2	1	0	1
IND	84	57	169	43
LEP	37	31	98	23
CHE	0	2	2	1
TUP	4	4	14	3
<b>Total</b>	7125	4921	11420	3562

Tableau 33 : Transitions observées et calculées pour les séquences 12S rRNA (393 bp)

Groupe	AG		CT	
	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>
EUL	64	29	91	23
HAP	5	23	90	20
VAR	2	1	0	1
IND	63	34	126	28
LEP	19	12	47	11
CHE	0	1	2	1
TUP	2	2	8	2
<b>Total</b>	4689	2483	6957	2030

Tableau 34 : Transitions observées et calculées pour les séquences 12S rRNA (zones 'Loop')

Groupe	AG		CT	
	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>
EUL	29	9	20	10
HAP	15	11	31	12
VAR	0	0	0	0
IND	21	12	43	14
LEP	18	13	51	13
CHE	0	1	0	1
TUP	2	1	6	2
<b>Total</b>	2436	1424	4463	1532

Tableau 35 : Transitions observées et calculées pour les séquences 12S rRNA (zones 'Steem')

Groupe	AC		AT		CG		GT	
	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>
EUL	25	56	12	61	6	27	9	30
HAP	14	56	42	62	41	25	12	28
VAR	1	1	3	2	0	1	0	1
IND	24	70	29	72	7	35	0	36
LEP	8	42	15	36	10	20	0	17
CHE	0	2	6	2	2	1	0	1
TUP	2	5	2	5	0	3	0	3
<b>Total</b>	3651	5986	3024	6127	615	2860	549	2928

Tableau 36 : Transversions observées et calculées pour les séquences 12S rRNA (393 bp)

Groupe	AC		AT		CG		GT	
	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>
<b>EUL</b>	25	66	11	61	6	11	2	10
<b>HAP</b>	14	64	35	54	34	9	0	7
<b>VAR</b>	0	2	3	2	0	0	0	0
<b>IND</b>	19	76	24	70	2	14	0	12
<b>LEP</b>	8	33	10	24	5	5	0	4
<b>CHE</b>	0	2	4	2	0	0	0	0
<b>TUP</b>	2	4	1	3	0	1	0	1
<b>Total</b>	2831	5901	2244	5187	423	972	283	854

**Tableau 37 : Transversions observées et calculées pour les séquences 12S rRNA (zones 'Loop')**

Groupe	AC		AT		CG		GT	
	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>	<i>Observé</i>	<i>Calculé</i>
<b>EUL</b>	0	7	1	9	0	10	7	13
<b>HAP</b>	0	8	7	12	7	12	12	16
<b>VAR</b>	1	0	0	0	0	0	0	0
<b>IND</b>	5	10	5	12	5	14	0	16
<b>LEP</b>	0	11	5	11	5	15	0	15
<b>CHE</b>	0	1	2	1	2	1	0	1
<b>TUP</b>	0	1	1	1	0	2	0	2
<b>Total</b>	820	1166	780	1361	192	1603	266	1872

**Tableau 38 : Transversions observées et calculées pour les séquences 12S rRNA (zones 'Stem')**

En fonction des caractéristiques observées, seules les séquences de 393 bp et les zones 'Loop' seront utilisées dans les traitements et, dans les deux cas, le modèle de substitution utilisé sera le modèle TN93.

## Distances

Le **Tableau 39** présente les distances de Tamura-Nei pour les séquences de 393 bp (prenant en compte les transitions et les transversions (avec l'option 'complete deletion') calculées à partir des séquences consensus de chaque groupe (partie inférieure de la matrice) et de l'ensemble des séquences disponibles pour chaque groupe (partie supérieure de la matrice). Les moyennes intra groupe sont, quant à elles, indiquées sur la diagonale de la matrice. En utilisant les séquences consensus, les distances sont comprises entre 0,027 (comparaison entre LEP et MED) et 0,279 (comparaison entre IND et TUP). L'utilisation de l'ensemble des séquences disponibles pour chaque groupe fournit des résultats similaires: de 0,035 (entre LEP et MED) à 0,255 (entre IND et TUP). Les moyennes des distances intra groupe vont, quant à elles, de 0,009 (groupe CHE) à 0,066 (groupe TUP).

	EUL	HAP	VAR	IND	LEP	MED	CHE	DMA	TUP
EUL	<b>0,020</b>	0,057	0,074	0,115	0,140	0,149	0,111	0,131	0,222
HAP	0,054	<b>0,026</b>	0,088	0,118	0,147	0,164	0,133	0,147	0,247
VAR	0,081	0,084	<b>0,017</b>	0,113	0,131	0,140	0,126	0,143	0,224
IND	0,121	0,110	0,108	<b>0,063</b>	0,155	0,167	0,137	0,144	0,255
LEP	0,143	0,157	0,130	0,159	<b>0,033</b>	0,035	0,153	0,139	0,208
MED	0,153	0,168	0,149	0,164	0,027	-	0,158	0,145	0,210
CHE	0,116	0,136	0,135	0,150	0,162	0,170	<b>0,009</b>	0,131	0,209
DMA	0,150	0,162	0,145	0,159	0,148	0,163	0,151	-	0,226
TUP	0,227	0,260	0,244	0,279	0,217	0,219	0,222	0,247	<b>0,066</b>

Tableau 39 : Distances génétiques pour les séquences 12S rRNA(393 bp)

L'étude réalisée avec les zones 'Loop' (Tableau 40) montre qu'avec les séquences consensus, la plus faible distance (0,025) est trouvée entre les groupes MED et LEP, alors que la plus importante est celle entre IND et TUP (0,367). Lorsque sont prises en compte toutes les séquences disponibles pour chaque groupe les valeurs vont de 0,034 (entre MED et LEP) à 0,328 (entre IND et TUP). La plus faible valeur de distance intra groupe se trouve chez CHE (0,011) et la plus élevée pour le groupe IND (0,095).

	EUL	HAP	VAR	IND	LEP	MED	CHE	DMA	TUP
EUL	<b>0,030</b>	0,072	0,081	0,186	0,184	0,193	0,138	0,175	0,245
HAP	0,073	<b>0,035</b>	0,101	0,192	0,178	0,200	0,168	0,192	0,289
VAR	0,083	0,082	<b>0,027</b>	0,181	0,173	0,183	0,159	0,182	0,238
IND	0,197	0,170	0,167	<b>0,095</b>	0,227	0,236	0,194	0,214	0,328
LEP	0,178	0,188	0,175	0,236	<b>0,033</b>	0,034	0,203	0,202	0,254
MED	0,195	0,205	0,193	0,242	0,026	-	0,223	0,209	0,237
CHE	0,151	0,181	0,163	0,219	0,220	0,247	<b>0,011</b>	0,176	0,269
DMA	0,197	0,213	0,217	0,223	0,200	0,227	0,204	-	0,279
TUP	0,243	0,301	0,253	0,367	0,260	0,283	0,302	0,297	<b>0,075</b>

Tableau 40 : Distances génétiques pour les séquences 12S rRNA (zones 'Loop')

### Test 'Likelihood mapping'

Les diagrammes obtenus à partir des séquences de 393 bp et celles correspondant aux zones 'Loop' sont présentées à la Figure 13. Le Tableau 41 résume les résultats chiffrés des tests. On peut remarquer que les séquences de 393 bp présentent un plus fort pourcentage de quartets résolus (83,6 %) que les zones 'Loop' (74,2 %). Si les séquences totales sont utilisables pour des reconstructions phylogénétiques le pourcentage relativement peu élevé de quartets résolus avec les zones 'Loop' fait que les reconstructions effectuées à partir de celles-ci devront être examinées avec précautions.

	Bassin 1	Bassin 2	Bassin 3	Total bassins 1, 2 et 3	Bassin 7
Séquences totales	27,8	27,9	27,9	83,6	12,6
Zones 'Loop'	24,5	24,7	25,0	74,2	21,6

Tableau 41 : Résultats du test 'Likelihood mapping' pour les séquences 12S rRNA

## Données mixtes

Pour ces données mixtes les séquences de cytochrome b 357 bp et les séquences des zones "Loop" du 12S rRNA ont été combinées, ce qui permet d'obtenir des séquences de 570 bp.

## Caractéristiques

Pour l'ensemble des séquences combinées, les pourcentages des nucléotides sont les suivants: 25,5 % pour la Thymine; 24,5 % pour la cytosine, 36,3 % pour l'Adénine et 13,8 % pour la Guanine, ce qui donne un biais général de composition de 0,1567. Les pourcentages des différents types de sites, de même que les pourcentages des transitions et des transversions, ainsi que le rapport Ti / Tv pour les différents groupes et l'ensemble des séquences sont indiqués au **Tableau 42**. On peut constater que près de la moitié des 570 sites sont variables et environ 41 % des sites sont de type parcimonieux informatifs. Le rapport Ti / Tv pour l'ensemble des séquences est de 2,0, c'est à dire comparable à celui trouvé tant pour les séquences de cytochrome b que pour les séquences de 12S rRNA.

Groupe	% V	% Pi	% S	% Ti	% Tv	R
EUL	14,21	8,24	5,96	4,74	0,70	7,6
HAP	17,02	11,05	5,96	5,44	1,40	3,8
VAR	1,75	-	-	1,23	0,53	2,3
IND	22,28	12,98	9,30	8,60	2,10	4,1
LEP	7,89	2,81	5,09	1,93	1,23	1,6
CHE	14,56	-	14,38	7,19	2,98	2,5
TUP	5,96	-	-	4,21	1,75	2,4
Total	49,65	40,88	8,42	10,17	5,09	2,0

Tableau 42 : Caractéristiques des séquences mixtes

Le modèle de substitution TN93 sera utilisé pour tous les traitements réalisés à partir de ces données mixtes.

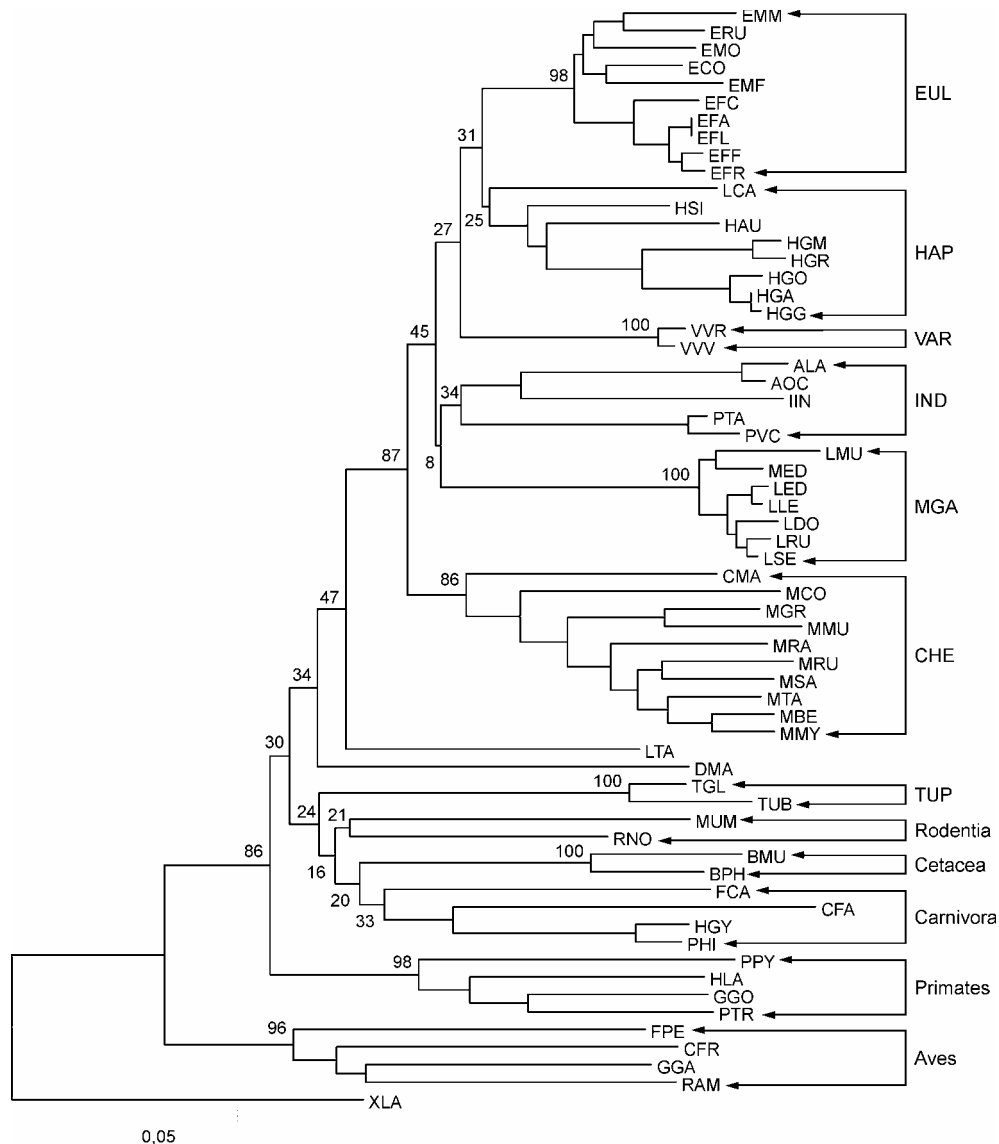
## Test 'Likelihood mapping'

Le diagramme obtenu par l'analyse dans TREE-PUZZLE est présenté à la **Figure 13**. Pour ces séquences mixtes, 96,2 % des 10 000 quartets analysés sont entièrement résolus et seulement 1,4 % ne sont pas du tout résolus. Ceci fait que ces séquences devraient être celles fournissant les meilleurs résultats lors des reconstructions phylogénétiques.

## Test de la monophylie des Lemuriformes

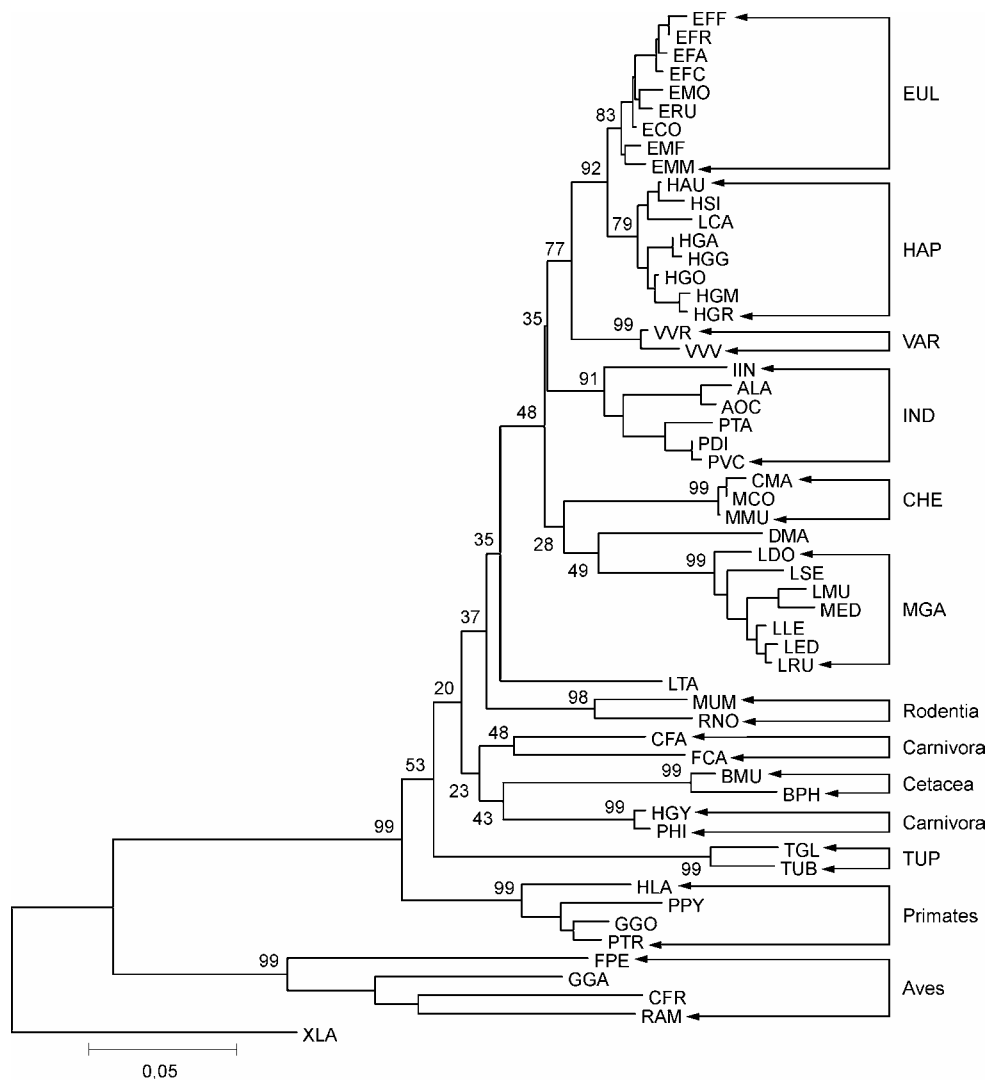
Ce test de monophylie a été mené d'une part avec les séquences de cytochrome b 357 bp et d'autre part avec les séquences de 12S rRNA de 393 bp. Les reconstructions sont effectuées en utilisant la méthode NJ avec des distances de Tamura-Nei prenant en compte les transitions et les transversions. Dans le cas des séquences 12S l'option 'complete deletion' est utilisée. La crédibilité des nœuds est testée par un bootstrap avec 5 000 répliques. Les arbres sont enracinés en utilisant la séquence XLA (*Xenopus laevis*).

Dans le cas du cytochrome b (**Figure 14**) le groupe des Lemuriformes (dans lequel ne sont incluses ni la séquence DMA, ni celles du groupe TUP) est supporté par une valeur de bootstrap de 87 %. L'inclusion de la séquence LTA dans le groupe des Lemuriformes fait passer la valeur de bootstrap à 47 %, ou à 34 % si les séquences LTA et DMA sont incluses. Quant aux séquences du groupe TUP elles sont trouvées dans un cluster renfermant les représentants des Rongeurs, Cétacés et Carnivores.



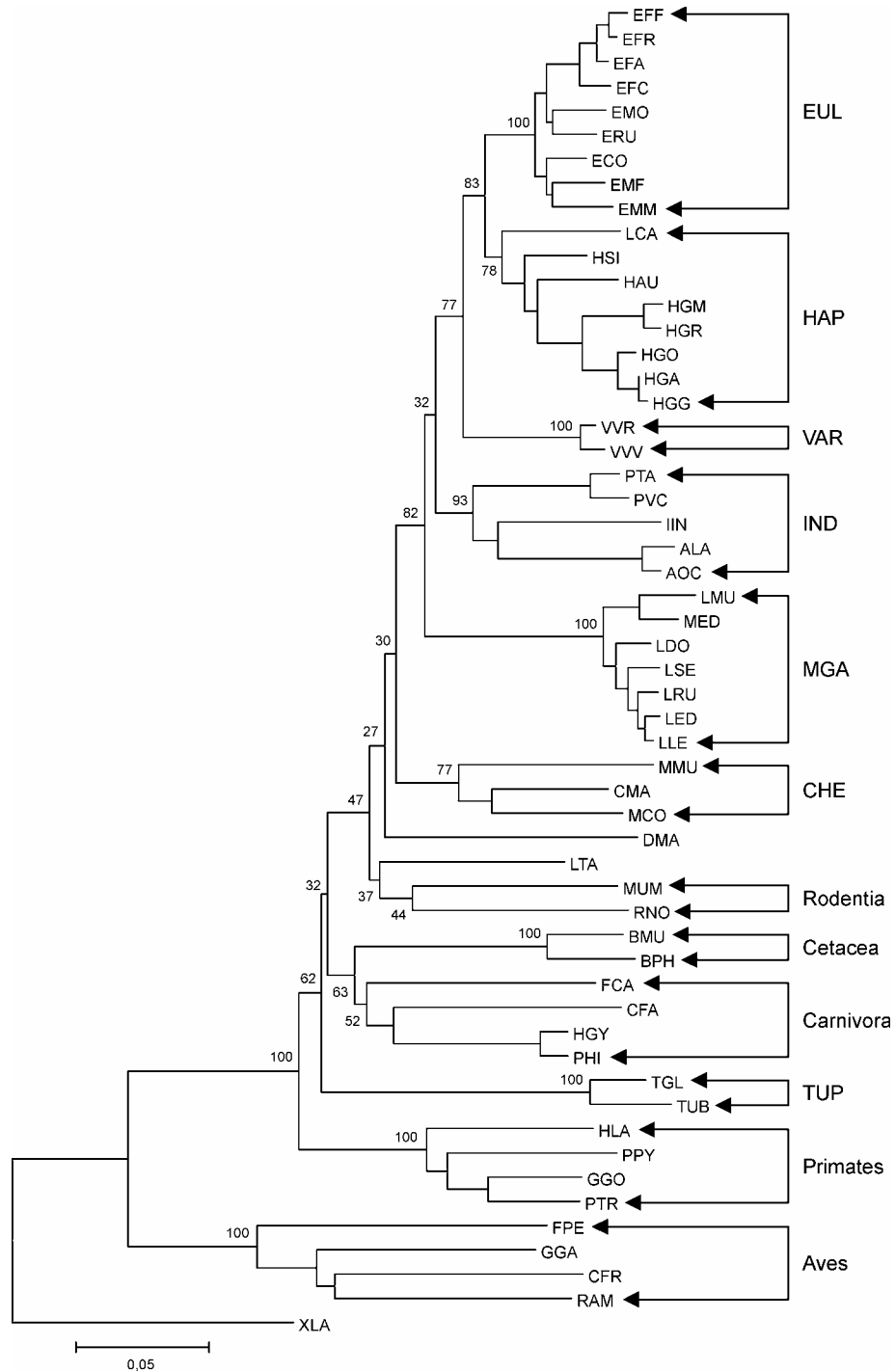
**Figure 14 : Reconstruction phylogénétique à partir des séquences de cytochrome b (357 bp) pour les tests de monophylie**

Dans le cas des séquences de 12S rRNA (**Figure 15**), le groupe des Lemuriformes, dans lequel se trouve incluse la séquence DMA est supporté par une valeur de bootstrap relativement faible (48 %). Si la séquence LTA est incluse la valeur de bootstrap passe à 36 %, ce qui est comparable avec les résultats obtenus pour le cytochrome b. Les séquences du groupe TUP sont ici situées entre le cluster des Primates et les séquences correspondant aux représentants des Rongeurs, Carnivores et Cétacés.



**Figure 15 : Reconstruction phylogénétique à partir des séquences de 12S rRNA (393 bp) pour les tests de monophylie**

Un troisième test a été mené en utilisant la combinaison des séquences de cytochrome b et de 12S rRNA (**Figure 16**). Dans ce cas le groupe des Lemuriformes est supporté par une faible valeur de bootstrap (30 %) sans inclusion de la séquence DMA et de 27 % avec inclusion de cette séquence. Comme dans le cas des séquences 12S rRNA seules, le groupe TUP est ici situé entre le cluster des Primates et les séquences des Rongeurs, Cétacés et Carnivores.



**Figure 16 : Reconstruction phylogénétique à partir des séquences mixtes (cytochrome b et 12S rRNA zones 'Loop') pour les tests de monophylie**

Ces trois topologies montrent que les séquences de Lemuriformes utilisées dans cette étude se trouvent groupées dans un cluster si l'on ne tient pas compte de la séquence DMA et de celles du groupe TUP dont des analyses plus précises pourront préciser la position.



## Détermination de la séquence outgroupe optimale

La détermination d'une séquence outgroupe optimale est essentielle afin d'obtenir la crédibilité maximale des études phylogénétiques. Si de nombreux critères ont pu être proposés pour cette définition (69, 203, 227, 305, 364), celui défini par Lyons-Weiler et coll. (194) semble être en mesure de fournir les meilleurs résultats. Ce dernier stipule que le signal phylogénétique présent dans une analyse fournissant un arbre enraciné (comprenant donc l'outgroupe) doit être supérieur à celui trouvé lors d'une analyse fournissant un arbre non enraciné, ne prenant en compte que les séquences de l'ingroupe.

Une mesure directe du contenu phylogénétique consiste en la détermination de la proportion des caractères présentant l'état ancestral dans une séquence donnée. Toutefois cette mesure nécessitant la connaissance de l'évolution de l'ensemble des séquences du groupe étudié, il n'est généralement possible d'effectuer qu'une mesure approximative. Dans un outgroupe trois facteurs vont déterminer la distribution de la pleisomorphie:

- Le nombre de caractères ayant changé.
- La position des caractères ayant changé.
- La manière dont les caractères ont changé.

On admet que la distance d'évolution entre l'outgroupe et le MRCA (More Recent Common Ancestor) de l'ingroupe est une fonction du temps et permet ainsi une bonne évaluation du nombre de caractères ayant changé. Si cette distance est importante le contenu pleisomorphique de l'outgroupe est généralement faible.

Pour réaliser cette étude les p-distances ont été calculées entre les différents groupes et 5 séquences uniques, phylogénétiquement plus ou moins proches des Lemuriformes: BMU, GGO, LTA, PPY et XLA. Les résultats obtenus avec les séquences de cytochrome b 357 bp sont présentés dans le **Tableau 43** et ceux obtenus avec les séquences de 12S rRNA 393 bp dans le **Tableau 44**.

	Outgroupes				
	BMU	GGO	LTA	PPY	XLA
<b>EUL</b>	0,229	0,230	0,175	0,240	0,256
<b>HAP</b>	0,239	0,248	0,199	0,243	0,266
<b>VAR</b>	0,225	0,214	0,169	0,223	0,277
<b>IND</b>	0,242	0,248	0,199	0,252	0,275
<b>LEP</b>	0,253	0,222	0,187	0,226	0,294
<b>MED</b>	0,252	0,232	0,204	0,232	0,291
<b>CHE</b>	0,245	0,261	0,196	0,261	0,287
<b>DMA</b>	0,230	0,218	0,196	0,238	0,266
<b>TUP</b>	0,231	0,249	0,207	0,255	0,269
<b>Moyenne</b>	0,238	0,236	0,192	0,241	0,276
<b>± SE</b>	0,010	0,015	0,012	0,012	0,012

Tableau 43 : Test pour la séquence outgroupe optimale (cytochrome b 357 bp)

	<b>Outgroupes</b>				
	<b>BMU</b>	<b>GGO</b>	<b>LTA</b>	<b>PPY</b>	<b>XLA</b>
<b>EUL</b>	0,161	0,158	0,121	0,171	0,287
<b>HAP</b>	0,172	0,151	0,126	0,177	0,285
<b>VAR</b>	0,157	0,154	0,132	0,179	0,286
<b>IND</b>	0,169	0,157	0,124	0,179	0,292
<b>LEP</b>	0,174	0,154	0,162	0,183	0,303
<b>MED</b>	0,183	0,162	0,172	0,193	0,315
<b>CHE</b>	0,157	0,178	0,150	0,205	0,305
<b>DMA</b>	0,160	0,165	0,139	0,179	0,292
<b>TUP</b>	0,202	0,212	0,197	0,232	0,316
<b>Moyenne</b>	0,171	0,167	0,147	0,183	0,298
<b>± SE</b>	0,014	0,017	0,024	0,018	0,011

**Tableau 44 : Test pour la séquence outgroupe optimale (12S rRNA 393 bp)**

Dans le cas du cytochrome b, la moyenne des distances entre l'outgroupe et les Lemuriformes va de  $0,192 \pm 0,012$  pour la séquence LTA à  $0,276 \pm 0,012$  pour la séquence XLA. Quant au 12S rRNA les valeurs vont de  $0,147 \pm 0,024$  pour la séquence LTA à  $0,298 \pm 0,011$  pour la séquence XLA.

La séquence LTA est donc celle qui présente, pour les deux types de fragments, les meilleures caractéristiques, et sera donc utilisée dans toutes les analyses.

## **Traitements sur les séquences cytochrome b**

### **Séquences nucléotidiques**

#### **Substitutions nucléotidiques par site**

Le logiciel GZ-GAMMA permet d'estimer le nombre de substitutions par site, ainsi que le paramètre de forme (appelé  $\alpha$ ) de la distribution du taux de substitution. Ces estimations sont effectuées par une méthode de maximum de vraisemblance en utilisant une topologie définie pour les séquences étudiées.

Nous utilisons ici des topologies obtenues par la méthode NJ en utilisant les distances de Tamura-Nei (prenant en compte les transitions et les transversions) et un test de bootstrap comportant 5 000 répliques. En fonction de cette topologie, les séquences ancestrales peuvent être estimées à chaque nœud par une approche Bayesian (392) en utilisant le modèle de Kimura-2-paramètres. Puis le nombre de substitutions par site est estimée en suivant le modèle de Jukes-Cantor. Enfin le paramètre  $\alpha$  est estimé à partir de la distribution du nombre minimal de substitutions par site.

Le logiciel TREE-PUZZLE a également été utilisé pour l'estimation du paramètre  $\alpha$ , et de plus dans ce cas il est possible d'obtenir une estimation de la variance de la distribution gamma.

Les résultats de ces différentes estimations sont reportés dans le **Tableau 45**.

Groupe	Alpha 1*	Alpha 2**	Variance de la distribution gamma
EUL	0,236783	0,20 ± 0,05	5,09
HAP	0,341681	0,30 ± 0,06	3,34
IND	0,504956	0,38 ± 0,13	2,60
LEP	0,091595	0,13 ± 0,11	7,45
CHE	0,301728	0,21 ± 0,04	4,75
MGA	0,180461	0,14 ± 0,08	7,37
EHP	0,279447	0,22 ± 0,03	4,45
EHV	0,120210	0,13 ± 0,03	4,32
IMG	0,604270	0,53 ± 0,07	2,34
GRA	0,352313	0,32 ± 0,02	3,16
GRB	0,304454	0,31 ± 0,02	3,28
GRC	0,346072	0,32 ± 0,02	3,10

\* estimation dans GZ-GAMMA

\*\* estimation dans TREE-PUZZLE

**Tableau 45 : Estimations du paramètre alpha pour les séquences cytochrome b**

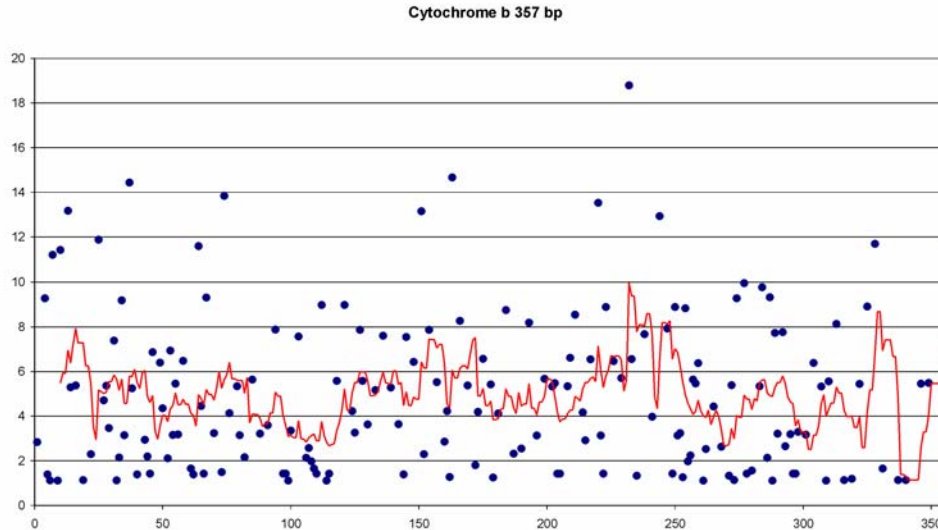
On peut constater qu'il existe d'importantes variations dans la valeur du paramètre de forme de la distribution gamma entre les différents groupes de Lemuriformes étudiés. La valeur la plus faible (0,091595 dans GZ-GAMMA et 0,13 dans TREE-PUZZLE) est trouvée pour le groupe des Lepilemuridae alors que la plus élevée (0,604270 dans GZ-GAMMA et 0,53 dans TREE-PUZZLE) pour le groupe constitué par l'association des Indriidae et Megaladapidae (IMG).

Le paramètre  $\alpha$  permet de rendre compte du taux de substitution par site ( $\lambda$ ), ces deux valeurs étant liées par:

$$\text{Coefficient de variation de } \lambda = \frac{1}{\sqrt{\alpha}}$$

Ceci fait que lorsque la valeur de  $\alpha$  est très élevée, le taux de substitution est uniforme ou quasi uniforme pour tous les sites considérés. A l'inverse de faibles valeurs de  $\alpha$  indiquent l'existence d'une variabilité du taux de substitution.

Dans le cas du cytochrome b, le groupe des Lepilemuridae est celui montrant la plus grande variabilité du taux de substitution et le groupe des Indriidae est celui pour lequel cette variabilité est la plus faible. On peut par ailleurs remarquer que l'inclusion du sub-fossile MED dans le groupe des Lepilemuridae (formant ainsi le groupe MGA) entraîne un doublement de la valeur de  $\alpha$  estimée dans GZ-GAMMA, et seulement une légère augmentation de cette valeur lorsqu'elle est estimée dans TREE-PUZZLE.



**Figure 17 : Répartition des substitutions par site (cytochrome b 357 bp)**

Pour l'ensemble des séquences étudiées (groupe GRC), la valeur de alpha est de 0,346072 (estimée dans GZ-GAMMA) ou de 0,32 (estimée dans TREE-PUZZLE). Elle traduit une hétérogénéité du taux de substitution par site. Ces données sont confirmées par le diagramme de répartition des nombres supposés de substitutions par site (**Figure 17**). Dans ce diagramme la courbe représentée en rouge représente la moyenne mobile calculée avec une période de 10.

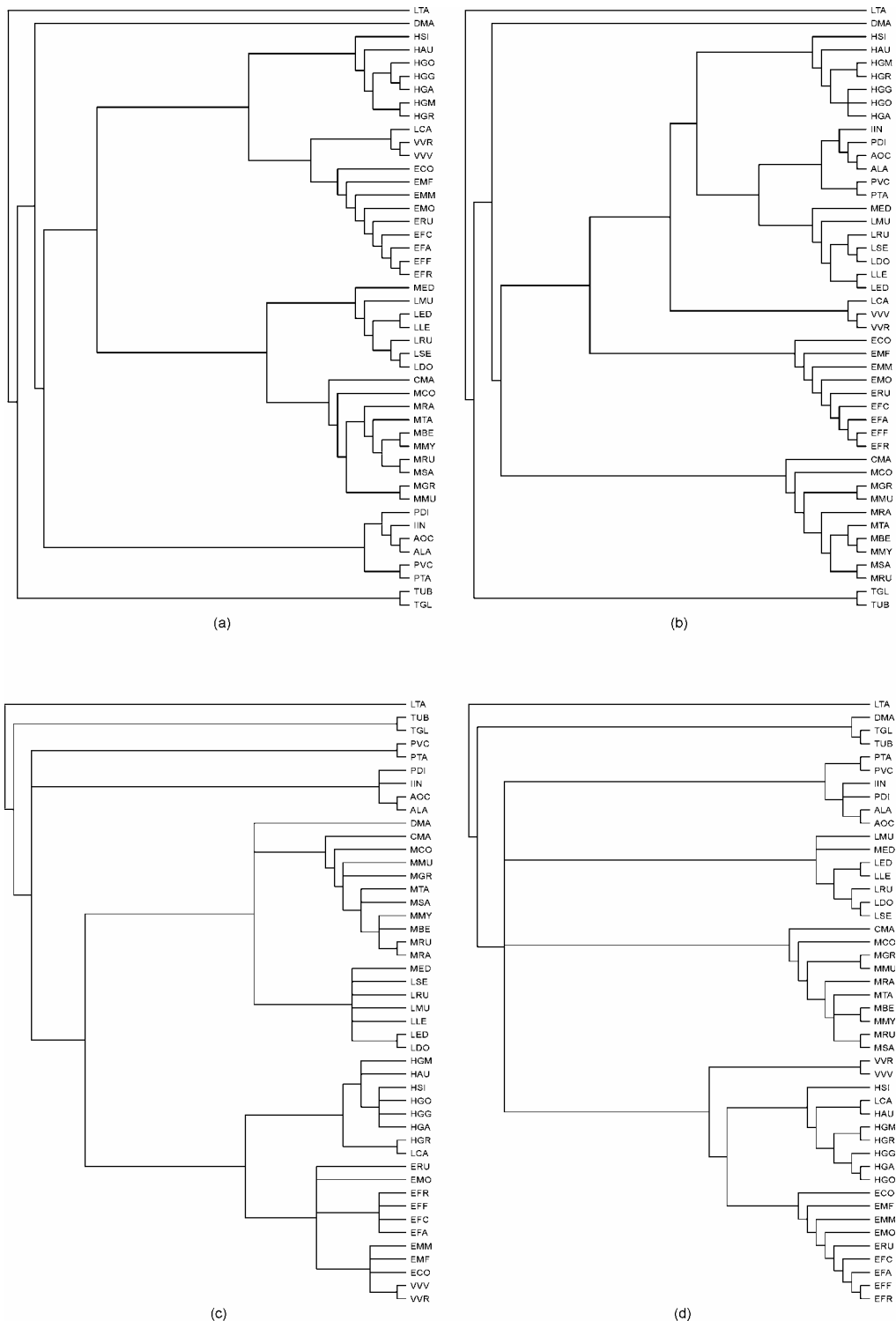
## Reconstructions phylogénétiques

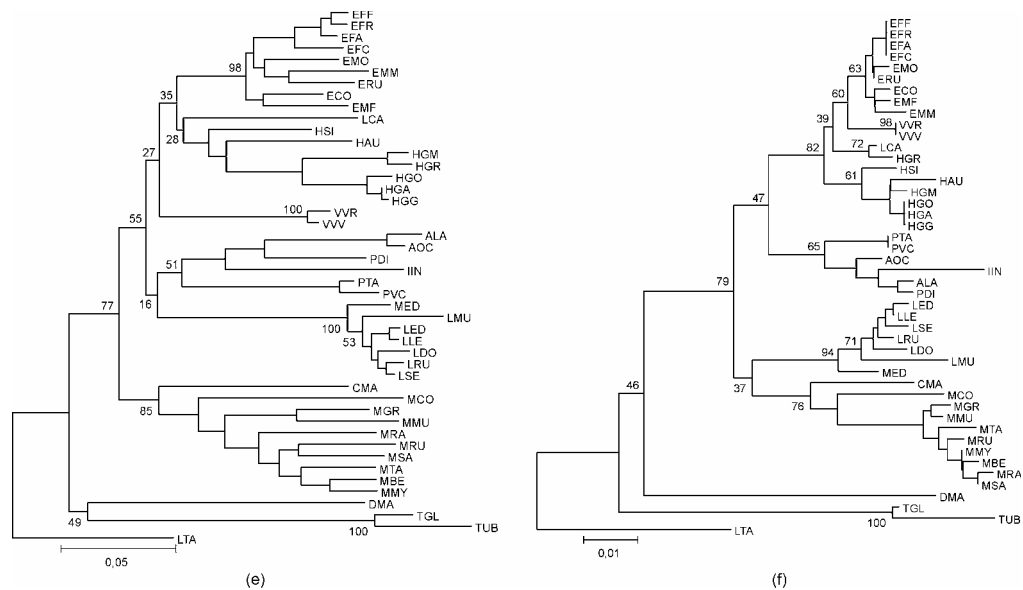
Pour les deux types de séquences de cytochrome b (séquences totales de 357 bp et séquences regroupant les positions 1 et 3 des codons) les méthodes de reconstruction utilisées sont les suivantes:

- **Maximum Likelihood** (ML): le programme DNAML du pack PHYLIP est utilisé avec une valeur du rapport Transitions / Transversions de 2,0 pour les séquences totales et de 1,9 pour les séquences partielles. L'arbre final est obtenu après réarrangements globaux. L'analyse est répétée 10 fois en utilisant à chaque fois un ordre aléatoire d'introduction des séquences.
- **Maximum Parsimony**: le programme DNAPARS du pack PHYLIP est utilisé soit en prenant en compte tous les types de substitutions (MP), soit en ne considérant que les transversions (MPV). Le ou les arbres les plus parcimonieux sont obtenus après réarrangements globaux. L'arbre final est le consensus strict des arbres les plus parcimonieux (traitement par le programme CONSENSE du pack PHYLIP). Comme dans le cas de l'analyse ML le traitement est répété 10 fois avec un ordre aléatoire d'introduction des séquences.
- **Parsimony Ratchet** (PR): le traitement est effectué à l'aide du logiciel WINCLADA en utilisant 500 répliques et le rééchantillonnage, à chaque réplique, de 20 % des caractères. Un seul arbre est retenu à chaque itération. L'arbre final est obtenu par consensus strict des arbres les plus parcimonieux.

- **Neighbor-Joining**: le traitement est effectué dans le logiciel MEGA. Chaque arbre est calculé à partir de 5 000 répliques, en utilisant les distances de Tamura-Nei pour les séquences totales, et celles de Kimura-2-paramètres pour les séquences partielles. Une reconstruction est effectuée en prenant en compte les transitions et les transversions (NJ) et une seconde en ne prenant en compte que les transversions (NJV).

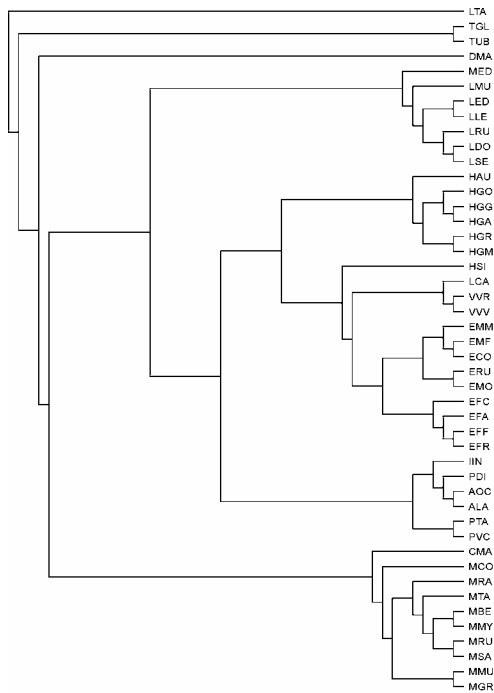
Les résultats obtenus à partir des séquences de 357 bp sont présentés à la **Figure 18**: (a) pour l'arbre ML, (b) pour l'arbre MP, (c) pour l'arbre MPV, (d) pour l'arbre PR, (e) pour l'arbre NJ et (f) pour l'arbre NJV.



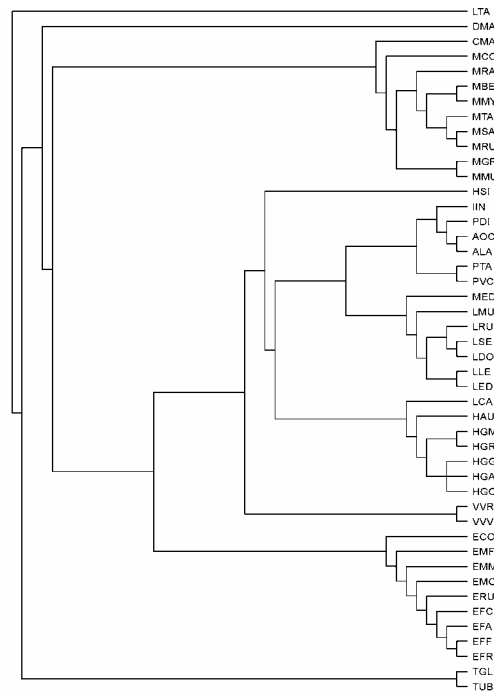


**Figure 18 : Reconstructions phylogénétiques utilisant le cytochrome b (357 bp)**

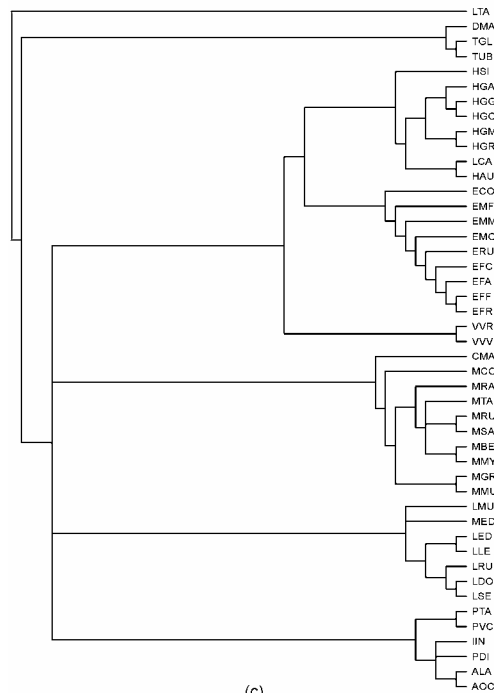
Ceux obtenus à partir des séquences partielles sont reproduits dans la **Figure 19**: **(a)** pour l'arbre ML, **(b)** pour l'arbre MP, **(c)** pour l'arbre PR, **(d)** pour l'arbre NJ et **(e)** pour l'arbre NJV. Pour ce type de séquences l'analyse MPV n'a pas été possible.



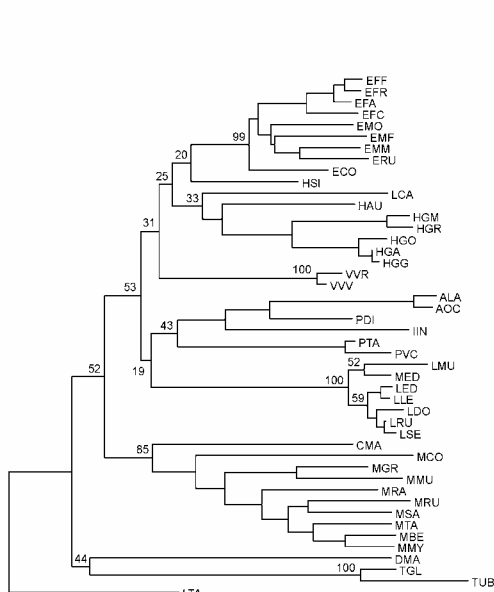
(a)



(b)



(c)



(d)

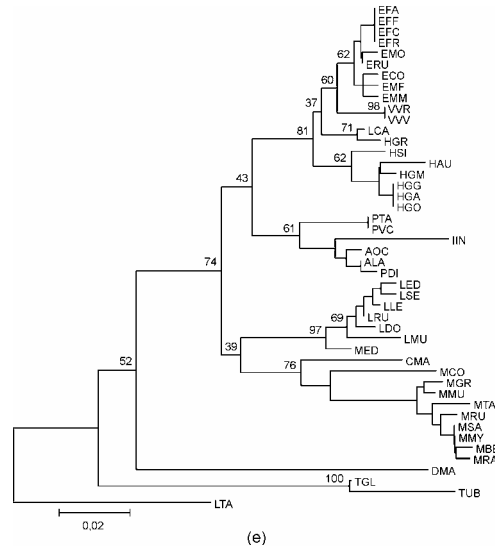


Figure 19 : Reconstructions phylogénétiques utilisant le cytochrome b (positions 1 et 3)

Pour les séquences de 357 bp, l'arbre MP est le consensus de 244 arbres et celui reconstruit à partir des séquences partielles est le consensus de seulement 6 arbres. L'arbre MPV, obtenu à partir des séquences totales, est unique. Dans le cas des reconstructions PR, pour les deux types de séquences, les arbres obtenus correspondent au consensus de 6 arbres.

Ces différentes topologies ont été comparées en utilisant le programme BASEML du pack PAML avec un test de bootstrap comprenant 10 000 répliquions, d'une part par la méthode REML (170), et d'autre part en utilisant la méthode de Shimodaira et Hasegawa (1999) avec une correction pour les comparaisons multiples. Ces tests ont été menés en considérant, ou non, l'existence d'une horloge moléculaire globale. Les résultats sont présentés aux **Tableaux 46** (séquences totales) et **47** (séquences partielles).

Topologie	Sans horloge			Avec horloge		
	<i>LnL</i>	<i>pSH</i>	<i>pRELL</i>	<i>LnL</i>	<i>pSH</i>	<i>pRELL</i>
ML	- 4083,994	- 1,000	0,629	- 4138,413	- 1,000	0,494
MP	- 4090,506	0,737	0,282	- 4152,653	0,570	0,103
MPV	- 4232,034	0,000	0,000	- 4369,876	0,000	0,000
PR	- 4100,311	0,484	0,042	- 4154,823	0,481	0,018
NJ	- 4100,754	0,482	0,048	- 4140,480	0,840	0,386
NJV	- 4202,209	0,000	0,000	- 4261,763	0,000	0,000

- 1,000 pour une valeur pSH indique que le calcul n'est pas possible

Tableau 46 : Comparaison des topologies (cytochrome b 357 bp)

Topologie	Sans horloge			Avec horloge		
	<i>LnL</i>	<i>pSH</i>	<i>pRELL</i>	<i>LnL</i>	<i>pSH</i>	<i>pRELL</i>
ML	- 3684,718	- 1,000	0,648	- 3734,547	- 1,000	0,815
MP	- 3692,586	0,675	0,173	- 3763,185	0,195	0,006
PR	- 3696,875	0,543	0,113	- 3758,191	0,263	0,016
NJ	- 3706,251	0,320	0,060	- 3667,901	0,491	0,163
NJV	- 3818,869	0,000	0,000	- 3876,883	0,000	0,000

- 1,000 pour une valeur pSH indique que le calcul n'est pas possible

Tableau 47 : Comparaison des topologies (cytochrome b positions 1 et 3)



Pour les deux types de séquences, que l'on prenne ou non en compte l'existence d'une horloge moléculaire, la topologie ML apparaît comme étant celle représentant l'arbre 'correct'. Dans celle reconstruite à partir des séquences de 357 bp (**Figure 18a**) les séquences des groupes EUL, HAP et VAR se trouvent réunies dans un cluster qui se présente comme étant le cluster frère d'un second dans lequel sont regroupées les séquences MGA et CHE. Les séquences du groupe IND forment un cluster situé à la base de l'ensemble précédent. La séquence DMA se trouve à la base des Lemuriformes et enfin le groupe TUP occupe la position la plus basale dans cette topologie.

Par contre pour la reconstruction effectuée à partir des séquences partielles (**Figure 19a**) on trouve, comme précédemment, un cluster formé des séquences EUL, HAP et VAR. Le groupe IND se trouve à la base de ce cluster, puis on trouve dans l'ordre, les groupes MGA, CHE et enfin la séquence DMA. Comme pour la topologie ML avec les séquences totales, dans le cas des séquences partielles le groupe TUP occupe la position la plus basale.

Lorsque l'on prend en compte l'existence d'une horloge moléculaire la topologie NJ est la plus probable après la topologie ML. Dans le cas des séquences de 357 bp (**Figure 18e**) on trouve un cluster formé par les groupes EUL, HAP et VAR qui est le cluster frère d'un second dans lequel sont regroupées les séquences des groupes IND et MGA. Le groupe CHE se trouve à la base de l'ensemble précédent. Enfin la séquence DMA et le groupe TUP forment un cluster dont la position est la plus basale dans la topologie. Le cluster regroupant Lemuridae, Indriidae et Megaladapidae est supporté par une valeur de bootstrap de 77 % et celui formé de DMA et TUP par une valeur de bootstrap de 48 %. Dans le cas des séquences partielles (**Figure 19d**) la topologie générale est identique avec toutefois des valeurs de bootstrap légèrement plus faibles.

Dans les reconstructions qui sont totalement résolues, c'est à dire celles ne présentant aucune multifurcation, le sub-fossile MED est toujours trouvé associé au cluster formé par les séquences de Lepilemuridae et occupe une position basale par rapport à ce cluster. Une seule exception à cette disposition a été trouvée pour la topologie NJ reconstituée à partir des séquences partielles, dans laquelle le sub-fossile est associé à la séquence LMU.

On peut également constater que les reconstructions NJV fournissent des topologies dont la disposition générale est identique à celle des topologies ML. Mais la crédibilité de ces topologies NJV est très faible.

Le choix entre la topologie ML et la topologie NJ est guidé par les résultats obtenus précédemment par d'autres auteurs, ainsi que par l'étude des longueurs des branches principales de chacune de ces topologies. Ces longueurs sont estimées pour les reconstructions effectuées tant avec les séquences totales qu'avec les séquences partielles (**Tableau 48**).

Branches	Séquences totales		Séquences partielles	
	Arbre ML	Arbre NJ	Arbre ML	Arbre NJ
Longueur totale	3,13447	3,22088	4,72417	4,78532
IND – MGA	0,152 (4,85 %)	0,206 (6,39 %)	0,275 (5,82 %)	0,311 (6,50 %)
IMG – EHV	0,067 (2,14 %)	0,044 (1,37 %)	0,051 (1,08 %)	0,079 (1,65 %)
CHE – GRA	0,067 (2,14 %)	0,092 (2,86 %)	0,178 (7,76 %)	0,160 (3,34 %)
DTU – GRB	0,114 (3,64 %)	0,117 (3,63 %)	0,136 (2,88 %)	0,100 (3,09 %)

Les valeurs entre parenthèses représentent le pourcentage de la longueur de la branche par rapport à la longueur totale de l'arbre.

**Tableau 48 : Longueurs des branches principales des topologies ML et NJ (cytochrome b 357 bp)**

On peut constater qu'en général les branches sont mieux définies, c'est à dire que leur longueur représente un pourcentage plus important de la longueur totale de l'arbre, avec les séquences partielles.

Le logiciel PHYLTEST est également utilisé afin de tester, au moyen de l'analyse quatre clusters, la crédibilité des différentes topologies constituées de quatre ensembles de séquences. Les principaux résultats de ces test sont indiqués dans le **Tableau 49**.

Topologie	Séquences	Branche interne	CP	Probabilité
(IND,HAP), (MGA,VAR)	357 bp	0,00442525	0,251	NS
	238 bp	0,00104022	0,0876	NS
(IMG,CHE+DTU), (HAP,EUL+VAR)	357 bp	0,00533065	0,281	NS
	238 bp	0,00974687	0,528	NS
(EUL,VAR), (HAP,IMG)	357 bp	0,00507973	0,0956	NS
	238 bp	0,0100814	0,205	NS
(TUP,DMA), (CHE,IMG+EHV)	357 bp	0,0276895	0,959	< 5 %
	238 bp	0,031438	0,847	< 20 %
(MED,LEP), (IND,EHV)	357 bp	0,0801632	1,000	< 0,1 %
	238 bp	0,121091	1,000	< 0,1 %
(IMG,EHV), (CHE,DAM+TUP)	357 bp	0,0134725	0,959	< 5 %
	238 bp	0,0253644	0,957	< 5 %

**Tableau 49 : Test des branches internes pour les topologies cytochrome b (357 bp)**

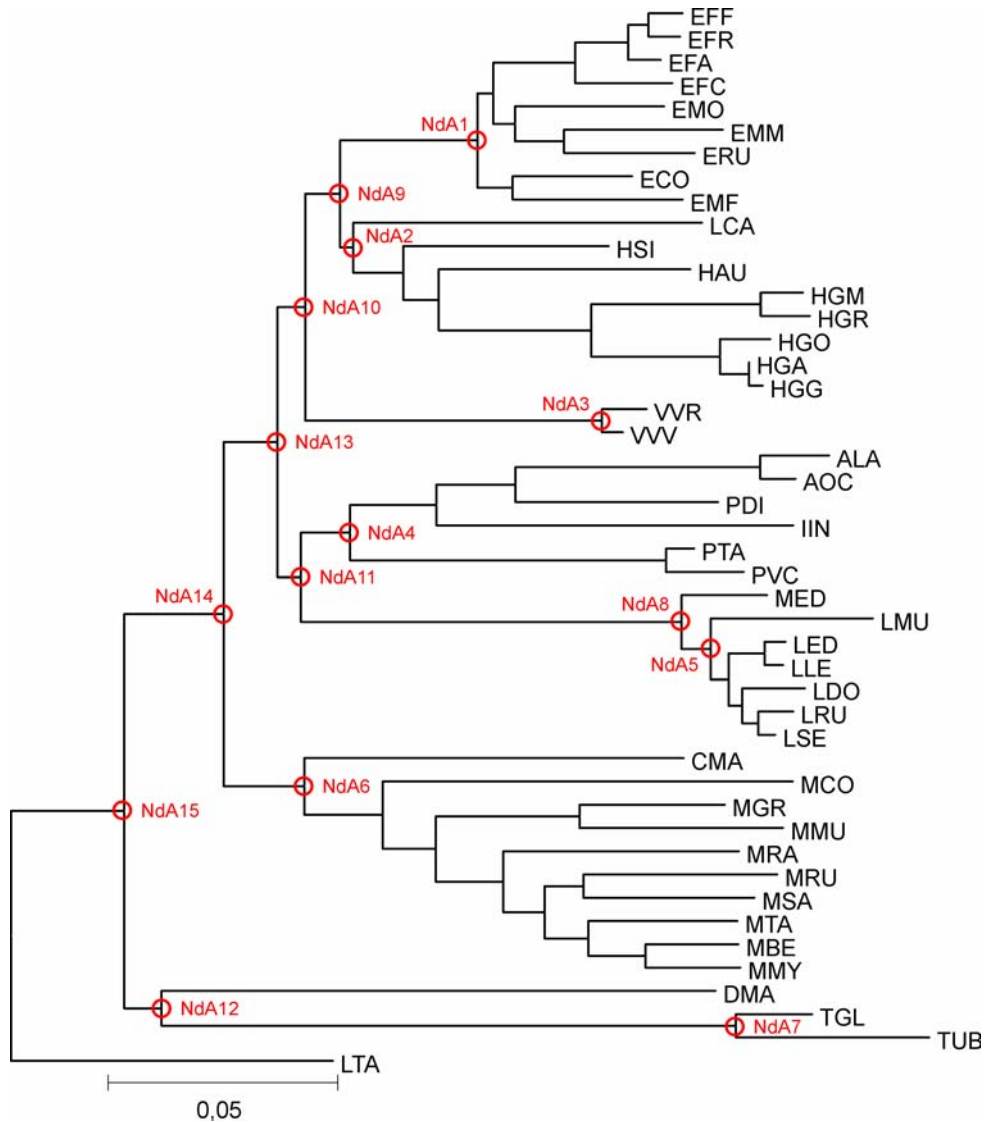
Ces résultats indiquent d'une part que l'association du sub-fossile MED et des séquences de Lepilemuridae est crédible avec une très forte probabilité et d'autre part que les séquences des groupes composés des Indriidae et Megaladapidae et des Eulemurs, Hapalemurs et Varecia ont une importante probabilité de former deux clusters associés. Parmi les groupes dont l'association est fortement improbable on trouve: Indriidae et Hapalemurs; Megaladapidae et Varecia; l'ensemble Indriidae, Megaladapidae et l'ensemble Cheirogaleidae, Daubentonia, Tupaia.

En fonction de ces résultats la topologie considérée comme la plus probable doit être représentée par celle fournie en utilisant la méthode NJ. Elle peut être représentée de la manière suivante (sans prendre en compte la séquence LTA utilisée comme outgroupe):

**((DMA, TUP), (CHE, ((IND, MGA), (VAR, (HAP, EUL))))))**

## Reconstruction des séquences ancestrales

La reconstruction des séquences nucléotidiques ancestrales est effectuée à l'aide des logiciels ANC-GENE et BASEML du pack PAML. Ces deux programmes travaillent à partir d'une topologie qui leur est fournie. Celle utilisée dans ce travail est la topologie NJ obtenue avec les séquences totales (**Figure 18e**). Les nœuds pour lesquels sont estimées les séquences ancestrales sont indiqués au **Tableau 50** et à la **Figure 20**.



**Figure 20 : Nœuds ancestraux étudiés pour le cytochrome b**

Noeud	Groupe de séquences
NdA1	EUL
NdA2	HAP
NdA3	VAR
NdA4	IND
NdA5	LEP
NdA6	CHE
NdA7	TUP
NdA8	MGA
NdA9	EHP
NdA10	EHV
NdA11	IMG
NdA12	DTU
NdA13	GRA
NdA14	GRB
NdA15	GRC

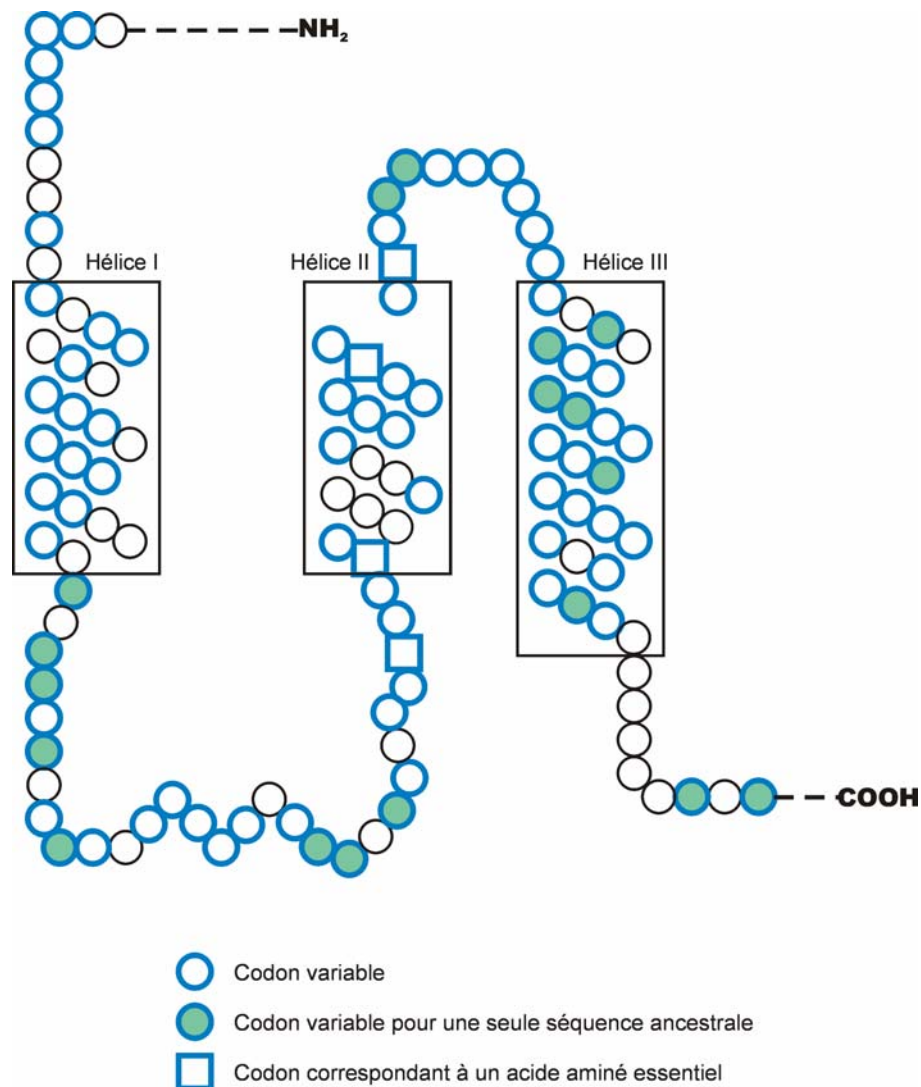
**Tableau 50 : Correspondances entre nœuds ancestraux et groupes de séquences pour le cytochrome b**

Si l'on compare la composition nucléotidique de ces séquences ancestrales (**Tableau 51**) avec celle des groupes actuels correspondants (**Tableau 14**) on peut constater que les valeurs du biais général de composition sont toujours plus élevées pour les séquences ancestrales que pour celles des groupes actuels, sauf en ce qui concerne le groupe LEP, que celui-ci soit considéré seul, ou avec inclusion de la séquence MED. Cette variation est due, dans les séquences ancestrales, à une augmentation des pourcentages en Cytosine et Adénine et à une diminution du pourcentage en Thymine. Pour la Guanine, les variations sont plus faibles que celles affectant les trois autres nucléotides, avec des pourcentages légèrement plus faibles dans les séquences ancestrales que dans les séquences actuelles.

Séquence	Groupe	T	C	A	G	Biais
NdA1	EUL	27,7	29,6	27,7	16,0	0,1267
NdA2	HAP	29,4	26,9	28,3	15,4	0,1280
NdA3	VAR	30,0	27,5	27,7	14,8	0,1360
NdA4	IND	25,2	30,3	28,6	15,7	0,1227
NdA5	LEP	28,0	28,3	25,8	17,3	0,0947
NdA6	CHE	28,9	26,9	29,1	15,1	0,1320
NdA7	TUP	28,6	25,2	29,7	16,5	0,1140
NdA8	MGA	28,0	28,9	25,5	17,6	0,0987
NdA12	DTU	26,9	26,6	31,4	15,1	0,1320
NdA15	Total	26,9	27,2	30,8	15,1	0,1320

**Tableau 51 : Compositions nucléotidiques des séquences ancestrales de cytochrome b**

En utilisant comme référence la séquence ancestrale estimée pour le nœud le plus basal (NdA15), l'analyse des 14 autres séquences ancestrales montre l'existence de 115 sites modifiés (soit 32,21 % des sites étudiés). Ces sites sont répartis dans 87 codons (79,11 % des codons étudiés). La répartition de ces sites est indiquée à la **Figure 21**. Parmi ces sites modifiés 18 (20,69 %) correspondent une modification sur une seule séquence ancestrale, 11 de ceux-ci (61,11 %) sont trouvés dans la séquence NdA7 (nœud ancestral des Tupaia), 3 (9,62 %) pour la séquence NdA6 (nœud ancestral des Cheirogaleidae), 2 (11,11 %) pour la séquence NdA1 (nœud ancestral des EUL) et 1 (5,55 %) pour les séquences NdA2 (nœud ancestral des Hapalemurs) et NdA3 (nœud ancestral des Varecia). Les sites modifiés sont localisés à 70,43 % en position 3 des codons, à 20,87 % en position 1 et à 9,70 % en position 3.



**Figure 21 : Codons variables pour les séquences ancestrales de cytochrome b**

Les transitions et transversions ont été calculées entre les nœuds ancestraux consécutifs et, pour les nœuds terminaux, entre la séquence ancestrale et la séquence consensus du groupe (**Figure 22**). On peut alors constater que le nombre de transitions est égal ou supérieur à celui des transversions pour toutes les comparaisons effectuées.

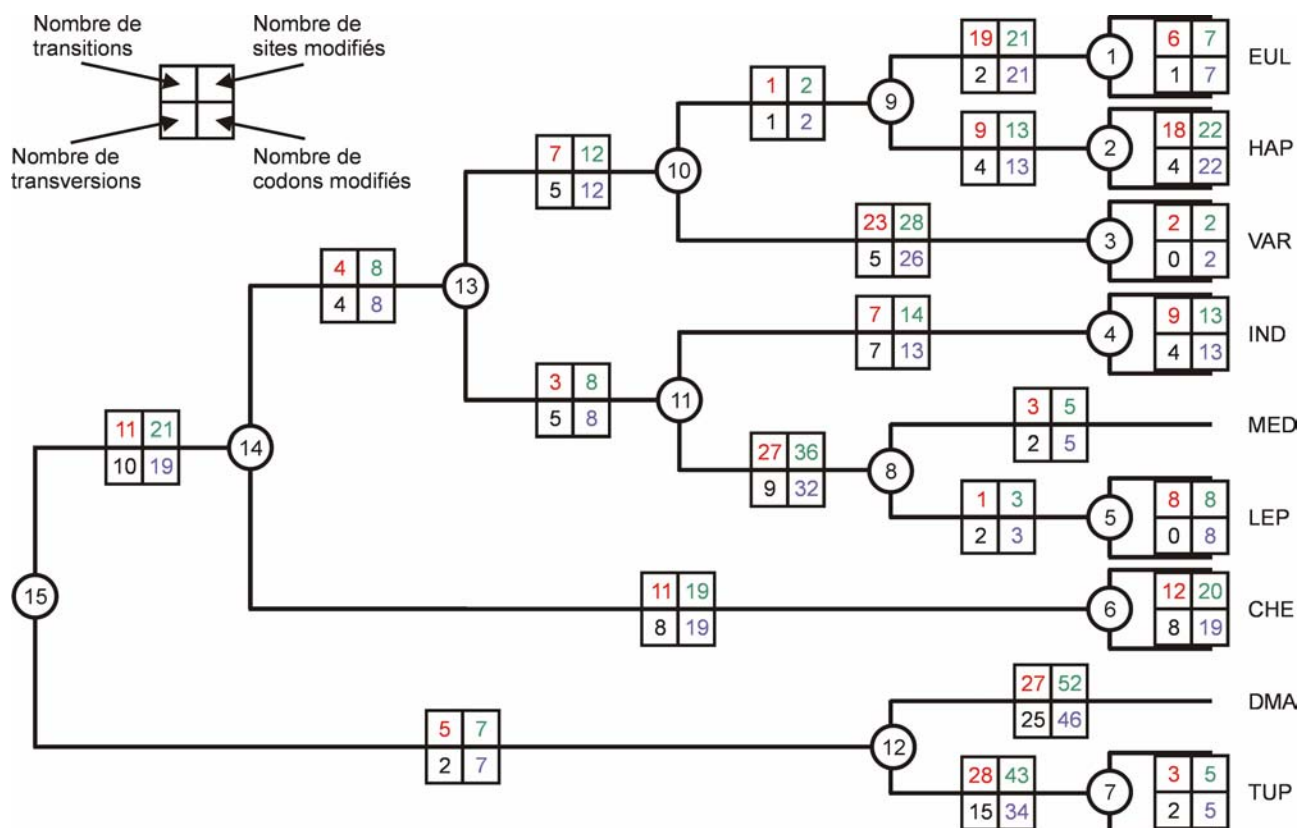


Figure 22 : Transitions et transversion pour les séquences ancestrales de cytochrome b

Le détail de ces différentes substitutions est indiqué dans le **Tableau 52**.

Entre	Transitions		Transversions			
	AG	CT	AC	AT	CG	GT
NdA15 – NdA12	0	5	1	1	0	0
NdA15 – NdA14	0	11	7	3	0	0
NdA12 – DMA	8	19	18	6	0	1
NdA12 – NdA7	6	22	8	6	1	0
NdA14 – NdA6	0	11	4	4	0	0
NdA14 – NdA13	1	3	2	2	0	0
NdA13 – NdA10	0	7	5	0	0	0
NdA10 – NdA3	2	21	1	2	2	0
NdA10 – NdA9	0	1	1	0	0	0
NdA9 – NdA1	4	15	1	1	0	0
NdA9 – NdA2	0	9	2	0	1	1
NdA13 – NdA11	1	2	4	0	0	1
NdA11 – NdA4	1	6	4	1	1	1
NdA11 – NdA8	9	18	6	2	1	0
NdA8 – MED	1	2	1	1	0	0
NdA8 – NdA5	0	1	1	0	0	1
NdA1 – CnsEUL	0	6	1	0	0	0
NdA2 – CnsHAP	1	17	2	2	0	0
NdA3 – CnsVAR	1	1	0	0	0	0
NdA4 – CnsIND	3	6	1	1	1	1
NdA5 – CnsLEP	3	5	0	0	0	0
NdA6 – CnsCHE	2	10	5	2	1	0
NdA7 – CnsTUP	1	2	0	2	0	0

Tableau 52 : Répartition des transitions et transversion pour les séquences ancestrales de cytochrome b

Du fait du biais de composition des séquences (proportion de Guanine inférieure à 25 %) les substitutions dans lesquelles est engagée la Guanine sont très peu représentées.

Substitution	Pourcentage observé	Pourcentage calculé
AG	14,13	37,65
CT	85,87	62,35
AC	67,28	33,23
AT	28,11	32,63
CG	2,76	17,22
GT	1,85	16,92
Transitions	55,98	44,02
Transversions	32,86	67,14

**Tableau 53 : Substitutions observées et calculées pour les séquences ancestrales de cytochrome b**

Une simulation, effectuée sur l'ensemble des séquences ancestrales, permet de comparer les valeurs réelles, et celles qui peuvent être calculées, pour les différents types de substitution (**Tableau 53**). Ce sont les transversions C  $\leftrightarrow$  G et G  $\leftrightarrow$  T qui présentent les plus fortes différences entre valeurs observées et calculées. Par rapport aux valeurs qui peuvent être calculées, les transitions sont augmentées d'environ 25 % et les transversions diminuées de 50 %.

Lorsque l'on analyse les modifications de sites survenant le long des branches menant aux groupes actuels, du nœud ancestral (NdA15) jusqu'au nœud ancestral du groupe (par exemple NdA1 pour le groupe EUL), il est possible de comparer les évolutions des différents groupes. Ces analyses sont effectuées à partir des alignements des séquences ancestrales, en prenant comme référence la séquence NdA15. Les résultats sont présentés au **Tableau 54**.

Branche menant à	Sites modifiés (1)	Codons modifiés (2)	Positions 1 (3)	Positions 2 (3)	Positions 3 (3)	Modifications 1 (4)	Modifications 2 (5)
EUL	13,16	36,97	17,02	6,38	76,60	34,09	23,40
HAP	12,04	33,61	16,28	6,98	76,74	40,00	18,60
VAR	14,84	39,49	20,75	5,66	73,59	23,40	33,96
IND	11,48	31,93	14,63	7,32	78,05	44,74	21,95
MED	16,81	42,02	20,00	8,33	71,67	50,00	46,67
LEP	17,08	42,86	21,31	6,56	72,13	27,45	1,64
CHE	9,52	26,89	17,05	8,82	74,13	46,87	38,23
DMA	1,96	5,88	0,00	14,28	85,72	-	-
TUP	13,44	32,77	20,83	14,58	64,53	12,82	85,42

(1) Pourcentage par rapport au nombre total de sites (357)

(2) Pourcentage par rapport au nombre de codons (119)

(3) Pourcentage par rapport au nombre de sites modifiés

(4) Pourcentage par rapport au nombre de sites modifiés. Modifications du type ACCCCC

(5) Pourcentage par rapport au nombre de sites modifiés. Modifications du type AAAAAC

**Tableau 54 : Modifications des sites sur les séquences ancestrales de cytochrome b en fonction des groupes**

Les branches présentant le pourcentage le plus élevé de sites modifiés sont celles menant à MED et LEP (respectivement 16,81 % et 17,08 %). Le pourcentage le plus faible est trouvé pour la branche menant à DMA (1,96 %). Pour les autres branches les valeurs varient de 14,84 % (branche menant à VAR) à 11,48 % (branche menant à IND). En terme de codons, les variations sont

identiques. Si l'on analyse les modifications en fonction de la position sur les codons on peut remarquer qu'environ 75 % des positions modifiées sont les troisièmes positions; les secondes positions étant celles pour lesquelles on retrouve le moins de modifications. Deux types particuliers de changements ont été analysés: d'une part ceux pour lesquels le nucléotide de la séquence NdA15 est modifié dès le premier nœud, cette modification restant constante jusqu'au nœud terminal; d'autre part celles pour lesquelles le site n'est modifié que pour le nœud terminal. Pour le premier type de modification les pourcentages les plus importants se trouvent pour les branches menant à MED (50,0 %) et CHE (46,87 %); et les plus faibles pour les branches menant à TUP (12,82 %) et à LEP (27,45 %). Ces modifications impliquent en grande majorité la substitution d'une Adénine en une Cytosine, ou d'une Thymine en une Cytosine, sauf dans le cas de la branche menant à TUP pour laquelle la modification la plus fréquente est la substitution d'une Cytosine en une Thymine. Pour le second type de changement, les variations entre groupes sont très importantes: de 85,42 % (pour la branche menant à TUP) à 1,64 % (pour la branche menant à LEP); pour les autres branches les valeurs vont de 21,95 % (branche menant à IND) à 46,67 % (branche menant à MED). Dans ce cas, pour tous les groupes, la substitution la plus fréquente est celle d'une Cytosine en une Thymine.

Si l'on étudie maintenant les modifications survenues entre la séquence du nœud terminal d'un groupe et la séquence consensus de ce groupe (**Tableau 55**) on peut constater que les groupes EUL, VAR, MED et TUP renferment peu de sites variables dans la branche et à nouveau modifiés dans la séquence consensus (entre 0,28 % et 1,40 %). Pour trois groupes les valeurs sont un peu plus élevées: de 2,24 % à 4,48 % pour HAP, IND et CHE. Enfin la valeur la plus élevée est trouvée pour le groupe DMA (13,44 %).

Groupe	Sites déjà variables				Nouveaux sites variables			
	Sites (1)	Position 1 (2)	Position 2 (2)	Position 3 (2)	Sites (3)	Position 1 (4)	Position 2 (4)	Position 3 (4)
EUL	0,56	0,00	0,00	100,00	10,64	40,00	0,00	60,00
HAP	3,64	15,38	7,69	76,93	18,60	0,00	0,00	100,00
VAR	0,28	100,00	0,00	0,00	1,89	100,00	0,00	0,00
IND	2,24	25,00	12,50	62,50	12,19	40,00	0,00	60,00
MED	1,40	20,00	0,00	80,00	8,33	0,00	0,00	100,00
LEP	1,12	0,00	0,00	100,00	6,56	0,00	0,00	100,00
CHE	4,48	31,25	0,00	68,75	11,76	0,00	0,00	100,00
DMA	13,44	22,92	14,58	62,50	42,86	0,00	33,33	66,67
TUP	1,12	25,00	0,00	75,00	2,08	0,00	0,00	100,00

(1) Pourcentage par rapport au nombre total de sites (357)

(2) Pourcentage par rapport au nombre de sites variables sur la branche

(3) Pourcentage par rapport au nombre de sites variables

(4) Pourcentage par rapport au nombre de nouveaux sites variables

**Tableau 55 : Modifications des sites sur les séquences ancestrales de cytochrome b entre nœud terminal et groupe actuel de séquences**

En ce qui concerne les sites devenant variables entre la séquence ancestrale et la séquence consensus les résultats vont dans le même sens que les précédents.



## Horloge moléculaire et arbres linéarisés

Une partie de cette étude est réalisée au moyen du logiciel LINETREE. Dans un premier temps les séquences de cytochrome b totales (357 bp) sont utilisées pour la reconstruction d'un arbre NJ en utilisant les distances de Tamura-Nei. La topologie obtenue (**Figure 23**) est en tous points conforme à celle qui peut être calculée dans le logiciel MEGA (**Figure 18e**).

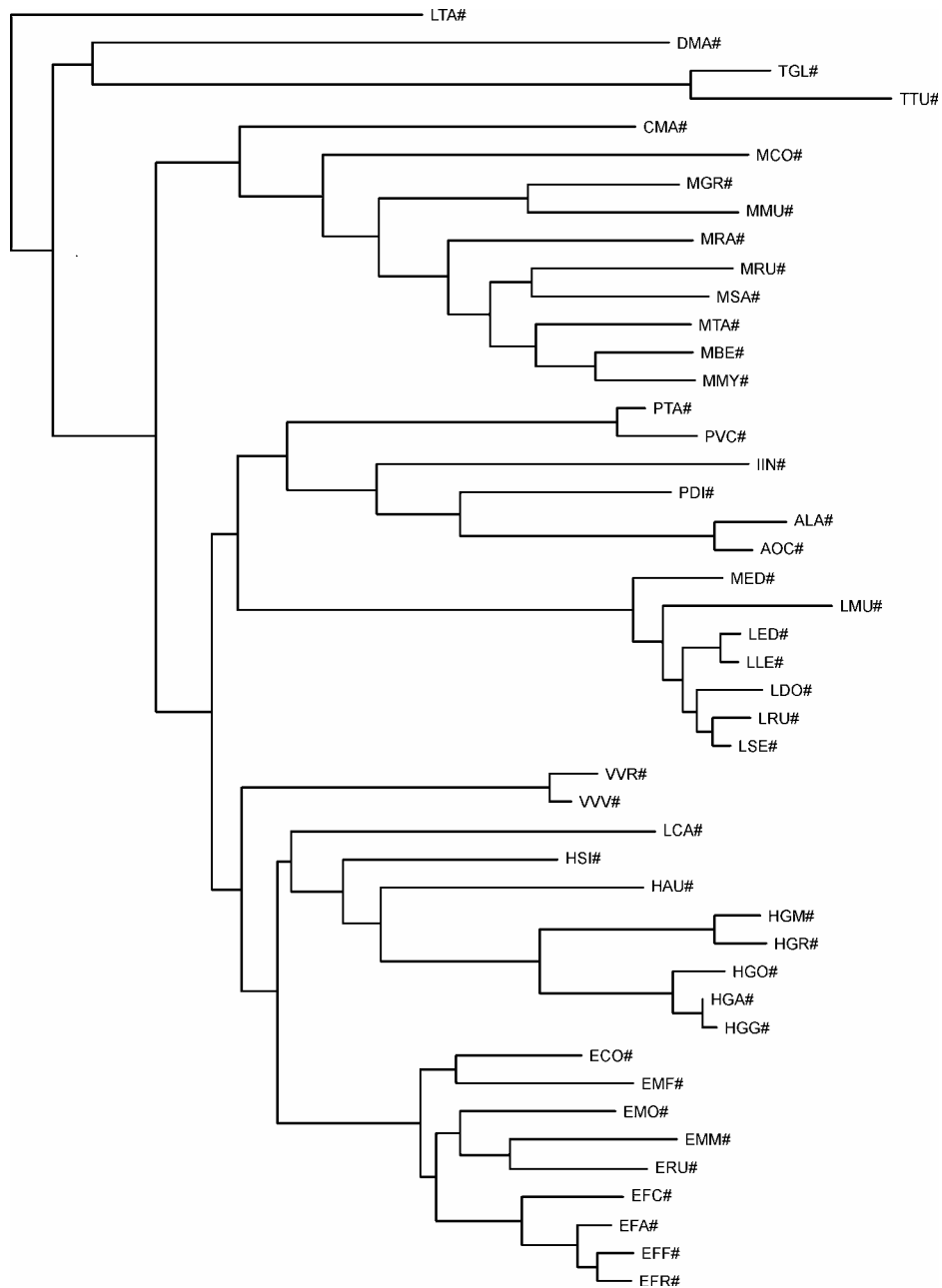


Figure 23 : Reconstruction NJ dans LINETREE pour le cytochrome b (357 bp)

Les valeurs de bootstrap, ainsi que les résultats de l'analyse 'deux clusters' effectuée dans LINETREE sont indiquées dans le **Tableau 56**.

Cluster A	Cluster B	Bootstrap (%)	b <sub>A</sub>	b <sub>B</sub>	Z	CP (%)
IND	MGA	16	0,098932	0,113920	1,112271	73,30
EUL	HAP	35	0,078483	0,088082	1,242538	78,50
DMA	TUP	49	0,119987	0,154928	1,546446	87,40
VAR	EHP	27	0,077974	0,087172	1,160942	75,40
IMG	EHV	55	0,110462	0,087822	2,136028	96,68
CHE	GRA	77	0,116985	0,110419	0,509352	38,30
DTU	GRB	-	0,150661	0,134222	0,835495	59,34

**Tableau 56 : Test 'deux clusters' pour la reconstruction NJ à partir du cytochrome b (357 bp)**

Une étude détaillée du test 'deux clusters' permet de montrer que la séquence HSI du groupe HAP présente une évolution significativement plus lente que celle de toutes les autres séquences analysées (avec une valeur de CP de 99,96 %, soit une significabilité du test de 1 %). D'autre part le test effectué sur les longueurs des branches de la topologie indique qu'en plus de la séquence HSI, la séquence ECO du groupe EUL montre une évolution significativement plus lente que celle de toutes les autres séquences (les valeurs de CP pour HSI et ECO, dans ce test, sont de 100 %, soit une significabilité de 1 %).

Compte tenu de ces résultats la reconstruction de l'arbre linéarisé est effectuée sans prendre en compte les séquences HSI et ECO. Les résultats du test 'deux clusters' avec ce nouvel ensemble de séquences sont reportés au **Tableau 57**. Ce test, de même que celui portant sur les longueurs des branches, ne montre plus de séquences ayant des évolutions significativement plus lentes que la valeur moyenne.

Cluster A	Cluster B	Bootstrap (%)	b <sub>A</sub>	b <sub>B</sub>	Z	CP (%)
IND	MGA	16	0,098409	0,114443	1,199733	76,60
EUL	HAP	35	0,081727	0,093236	1,343234	81,98
DMA	TUP	49	0,119933	0,154982	1,550597	87,64
VAR	EHP	27	0,078841	0,090427	1,393298	83,54
IMG	EHV	55	0,110853	0,089830	1,964961	95,00
CHE	GRA	77	0,117038	0,111926	0,392912	30,34
DTU	GRB	-	0,151213	0,135115	0,817175	59,82

**Tableau 57 : Test 'deux clusters' pour la reconstruction NJ à partir du cytochrome b (357 bp) après élimination des séquences aberrantes**

L'arbre linéarisé obtenu à partir de ces données est celui de la **Figure 24**.

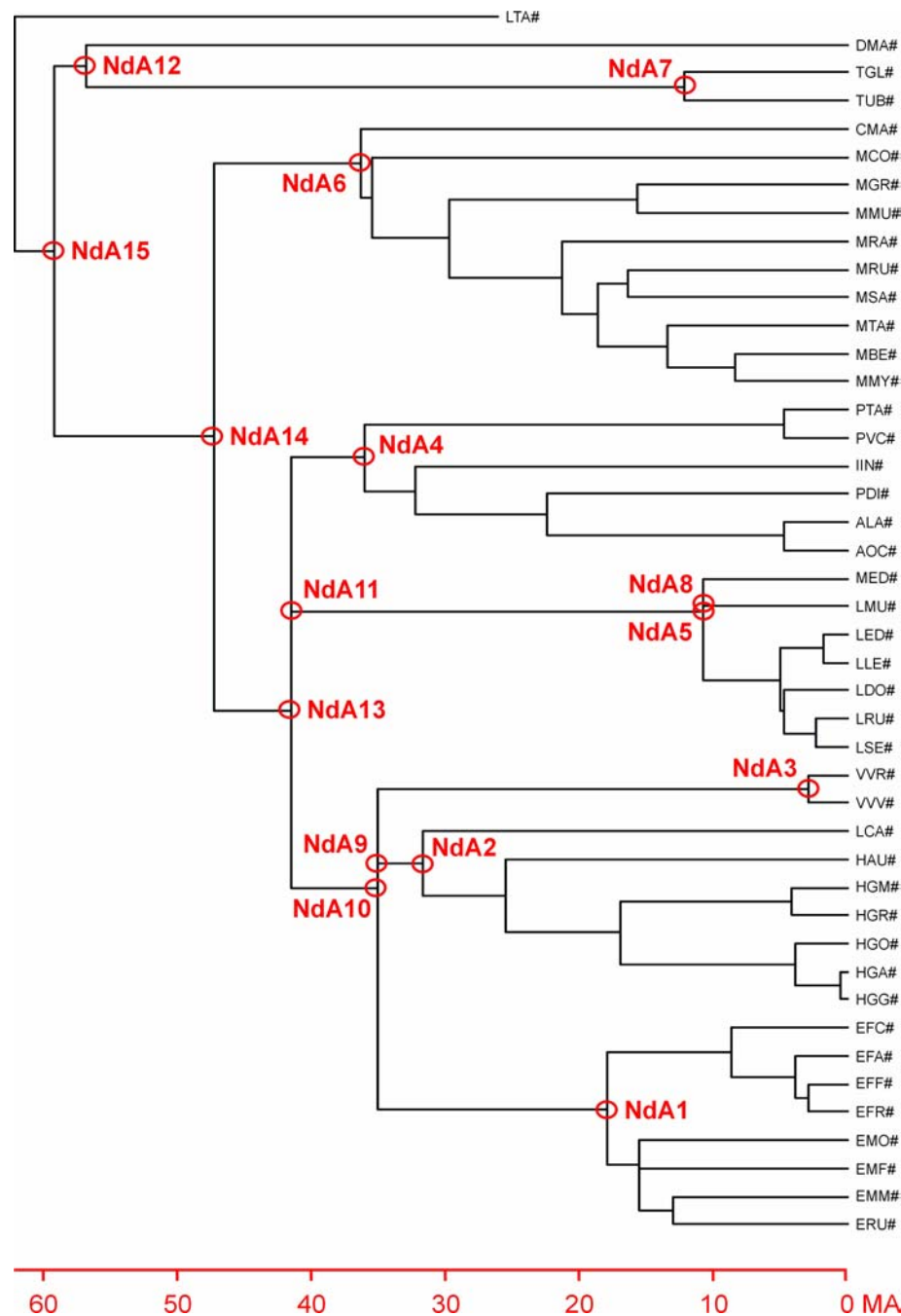


Figure 24 : Arbre linéarisé obtenu à partir des séquences de cytochrome b (357 bp)

En considérant que la divergence entre les Strepsirhini et les Lorisidae est intervenue il y a 62 millions d'années (MA), l'arbre linéarisé permet d'estimer les temps de divergence correspondant aux nœuds principaux de la topologie. Les estimations obtenues sont reportées dans le **Tableau 58** (arbre linéarisé 1). Un certain nombre de divergences ne peuvent être ici précisées correctement: séparation LEP – MED (nœud NdA8); séparation (IND,MGA) – (EUL,HAP,VAR) (nœud NdA13) et séparation VAR – (EUL,HAP) (nœud NdA10).

Nœud	Divergence	Arbre linéarisé 1	Arbre linéarisé 2	Arbre linéarisé 3	Arbre linéarisé 4	Moyenne ± SD
NdA1	EUL	18,27	-	18,36	-	18,31 ± 0,04
NdA2	HAP	31,53	-	33,91	-	32,72 ± 1,19
NdA3	VAR	2,94	-	2,93	-	2,93 ± 0,01
NdA4	IND	35,97	-	38,55	-	37,26 ± 1,29
NdA5	LEP	10,68	-	8,26	-	9,47 ± 1,21
NdA6	CHE	36,22	-	41,37	-	38,79 ± 2,57
NdA7	TUP	12,08	-	12,25	-	12,16 ± 0,08
NdA8	MED – LEP	10,97	10,28	9,49	9,86	10,15 ± 0,56
NdA9	EUL – HAP	34,88	27,44	39,95	29,20	32,87 ± 4,93
NdA10	VAR – EHP	36,05	37,59	40,75	43,56	39,49 ± 2,90
NdA11	IND – MGA	43,86	35,98	44,76	39,32	40,98 ± 3,55
NdA12	DMA – TUP	56,65	57,64	56,25	57,44	56,99 ± 0,57
NdA13	IMG – EHV	41,35	40,34	43,49	44,57	42,44 ± 1,68
NdA14	CHE – GRA	47,18	43,55	53,15	44,62	47,12 ± 3,72
NdA15	DTU – GRB	59,00	59,10	59,22	60,31	59,41 ± 0,53

**Tableau 58 : Temps de divergence estimés à partir des arbres linéarisés pour le cytochrome b**

La constance du taux d'évolution pour l'ensemble de la topologie peut être testée au moyen d'un test du  $X^2$ . Ce dernier fournit une valeur de 60,179028 avec un nombre de degrés de liberté (ddl) de 43. Ce test est significatif à 5 % (la valeur critique est de 59,3035). On peut donc supposer que le taux d'évolution n'est pas constant sur la totalité de l'arbre analysé.

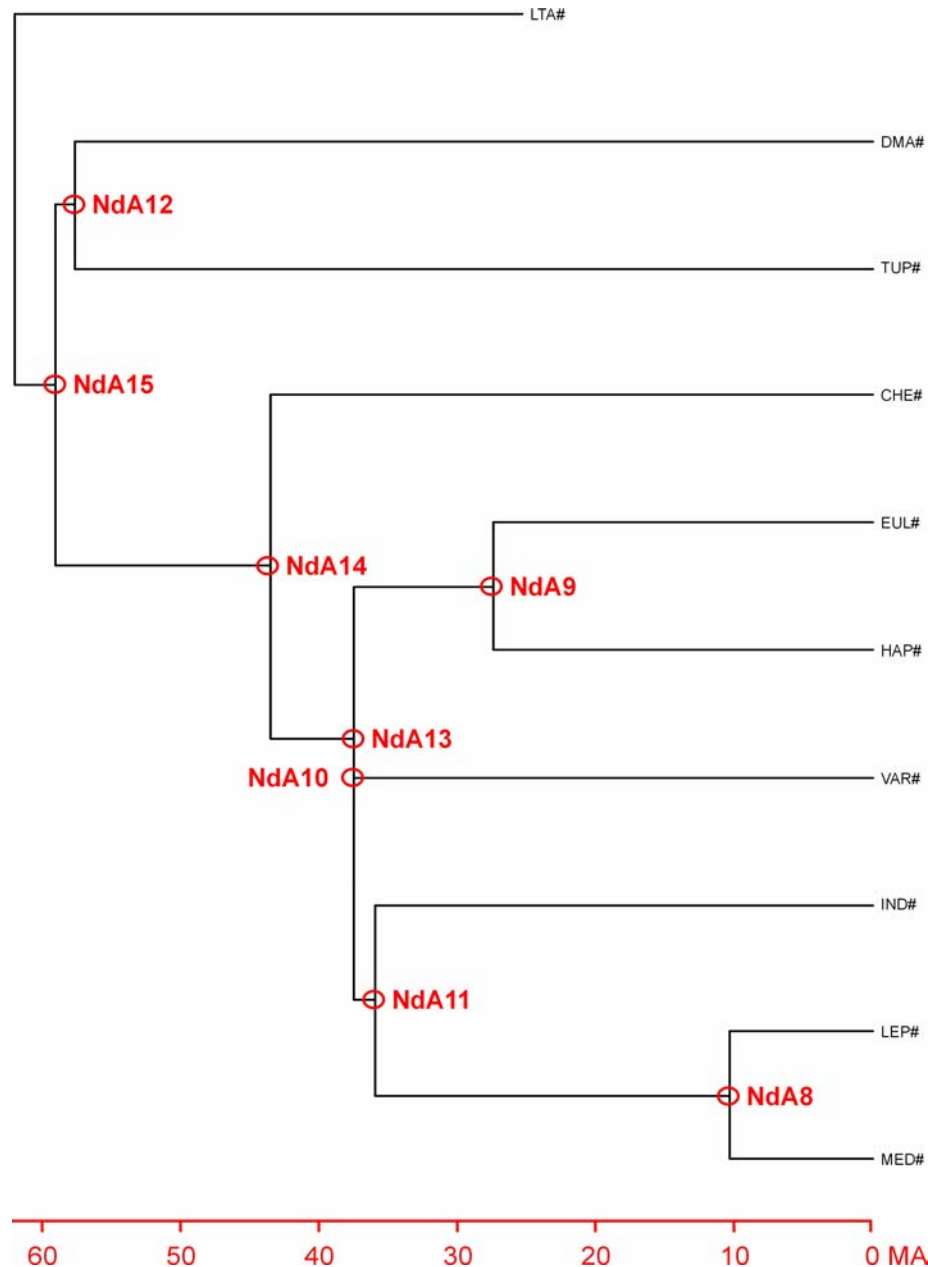


Figure 25 : Arbre linéarisé obtenu à partir des séquences consensus de cytochrome b pour les différents groupes

Les trois multifurcations qui sont présentes dans l'arbre linéarisé sont à rapprocher des faibles valeurs de bootstrap obtenues pour l'arbre NJ de la **Figure 18e**: 16 % pour ce qui est de la séparation IND – MGA et 21 % pour la séparation (EUL,HAP) – VAR. Afin de tenter de résoudre ces multifurcations les séquences consensus correspondant aux différents groupes sont utilisées. L'arbre linéarisé obtenu à partir de celles-ci est présenté à la **Figure 25**. On peut noter qu'il ne reste plus ici qu'une seule multifurcation. Comme dans le cas précédent cet arbre est utilisé pour l'estimation des temps de divergence, les valeurs obtenues sont indiquées dans le **Tableau 58** (arbre linéarisé 2). Le test des longueurs des branches (**Tableau 59**) montre que les séquences des groupes EUL, VAR et IND ont une évolution plus lente que les séquences des autres groupes.

Groupe	Distance à la racine	Delta *	Z	CP (%)
DMA	0,130308	0,000590	0,029400	2
TUP	0,152547	0,022828	0,998769	68
CHE	0,134868	0,005150	0,317123	24
EUL	0,105067	0,024057	2,296664	98
VAR	0,111552	0,017572	1,795886	93
HAP	0,136935	0,007215	0,526045	40
IND	0,110112	0,019605	1,913910	94
LEP	0,139271	0,009553	0,796515	57
MED	0,145617	0,015899	1,267977	79

\* Différence entre la distance à la racine et la moyenne générale des distances à la racine

**Tableau 59 : Test 'deux clusters' pour la topologie à partir des séquences consensus de cytochrome b**

Pour la séquence EUL, la valeur de CP est de 98 % ce qui fournit une probabilité voisine de 5 % pour le rejet de l'hypothèse de constance du taux d'évolution. Pour les séquences VAR et IND, cette même probabilité est comprise entre 5 % et 10 %. Ces données sont confirmées par le test du taux d'évolution effectué sur la topologie: celui-ci fournit une valeur de 14,755541 (avec 9 ddl), soit une probabilité de 10 % (la valeur critique est de 14,6837). Ceci confirme le fait qu'avec les séquences consensus, comme avec les séquences individuelles, le taux d'évolution ne peut être considéré comme constant pour les différents groupes de Lemuriformes.

Une étude similaire est effectuée en utilisant les séquences partielles de cytochrome b. L'analyse portant sur l'ensemble des séquences montre, comme précédemment, que les séquences HSI et ECO présentent une évolution significativement plus lente que celle des autres séquences et doivent être retirées de l'analyse. L'arbre linéarisé obtenu en utilisant les distances de Kimura-2-paramètres est présenté à la **Figure 26**. Comme dans le cas des séquences de 357 bp on constate ici l'existence de multifurcations ne permettant pas de préciser l'émergence relative des groupes IND et MGA d'une part, et des groupes VAR, EUL et HAP d'autre part. Par contre la divergence MED – LEP est individualisée.

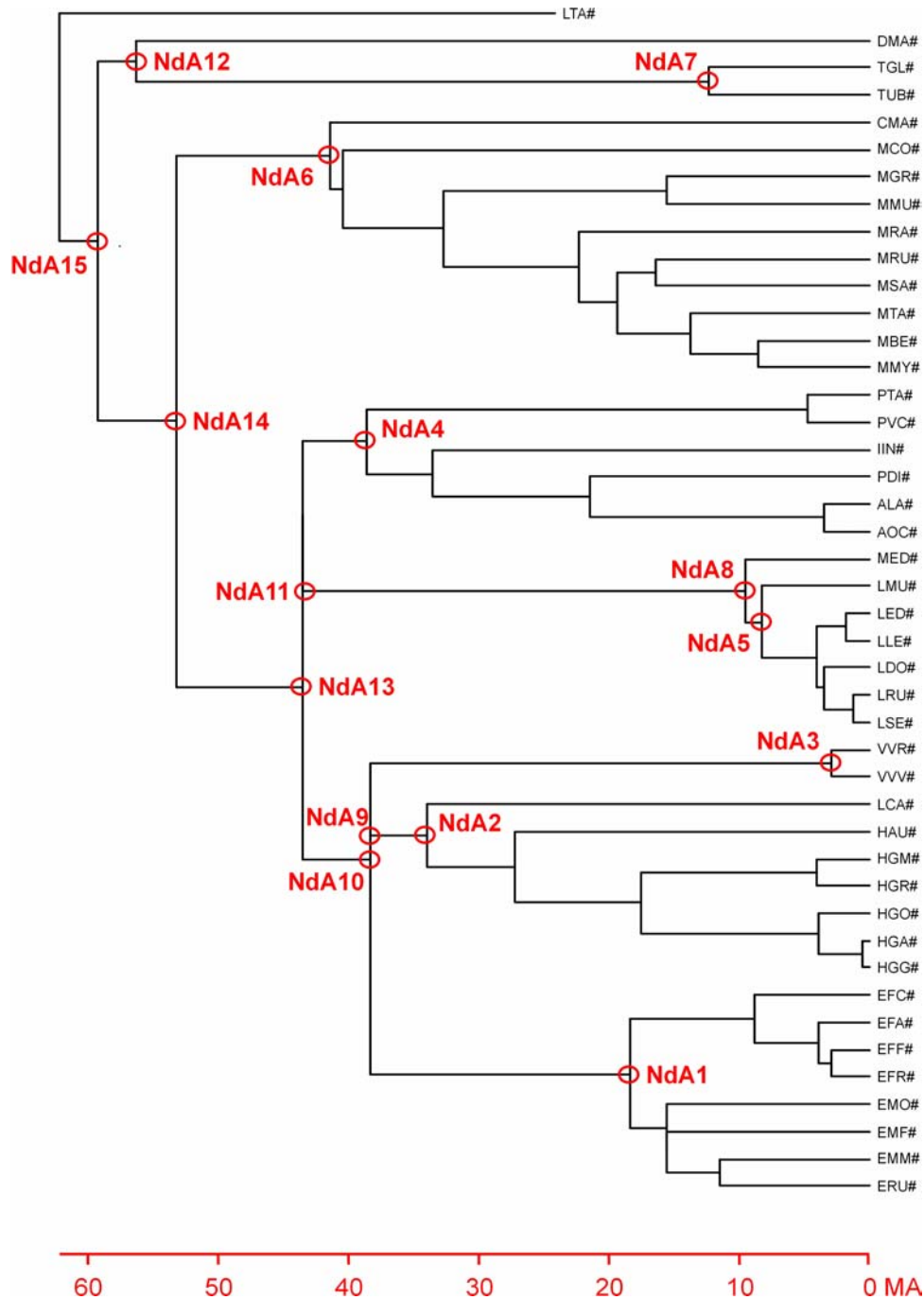


Figure 26 : Arbre linéarisé obtenu à partir des séquences partielles de cytochrome b (positions 1 et 3)

L'utilisation des séquences consensus pour les différents groupes permet la reconstruction de l'arbre linéarisé de la **Figure 27**. Cet arbre ne présente plus qu'une seule multifurcation faisant que la divergence (EUL,VAR,HAP) – (IND,MGA) ne peut être précisée. Les estimations des temps de divergence, effectuées à partir de cette topologie, sont indiquées au **Tableau 58** (arbre linéarisé 4). Le test de la constance du taux d'évolution donne ici une valeur de 16,651343 ce qui avec 9 ddl, montre que le test est significatif à 5 % (la valeur critique est de 16,9190) et donc que la constance du taux d'évolution ne peut être supportée. Ce résultat est en contradiction avec celui obtenu lors de l'utilisation de l'ensemble des séquences partielles pour chaque groupe. Il semble ici que les séquences consensus, d'une certaine manière, ont tendance, pour les séquences partielles du cytochrome b, à amplifier les différences intervenant dans les calculs de taux d'évolution.

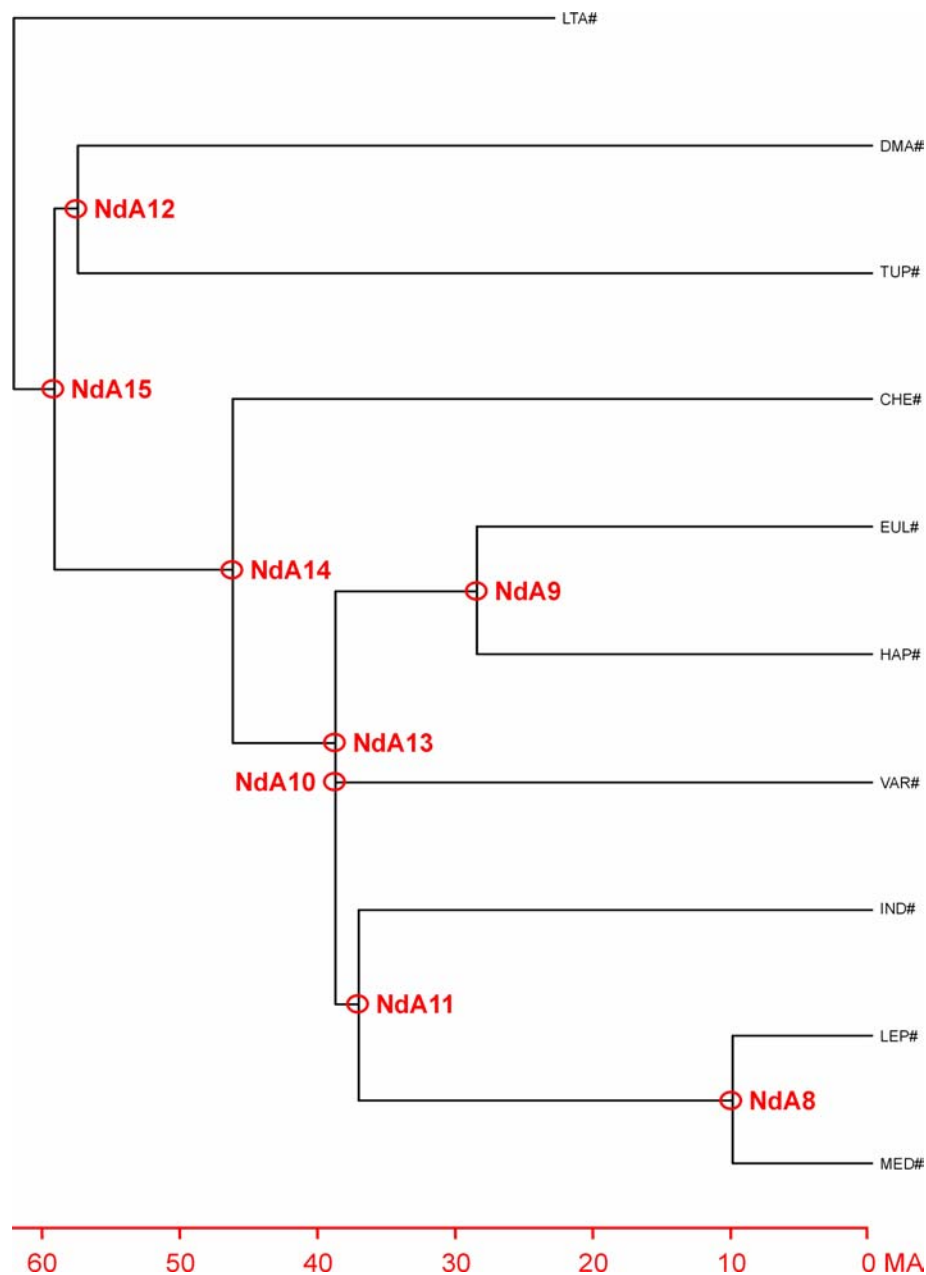


Figure 27 : Arbre linéarisé obtenu à partir des séquences consensus partielles de cytochrome b (positions 1 et 3)

Les temps de divergence ont également été estimés en utilisant le logiciel BASEML du pack PAML avec la topologie NJ de la **Figure 18e**. Dans ce cas différents modèles de substitutions ont été testés: JC69, K80, F81, F84, HKY85, TN93 et REV. Ces études ont été menées avec les séquences totales et avec les séquences partielles. Les résultats sont regroupés dans le **Tableau 60**.



Nœud	JC69	K80	F81	F84	HKY85	TN93	REV
NdA1	10,16 (12,94)	9,33 (11,92)	10,93 (12,57)	9,07 (11,02)	9,03 (11,03)	9,11 (11,37)	9,30 (11,62)
NdA2	19,53 (25,22)	18,38 (24,08)	21,50 (25,11)	18,39 (23,89)	18,35 (23,91)	18,91 (25,04)	19,30 (25,58)
NdA3	1,46 (1,90)	1,27 (1,64)	1,58 (1,85)	1,24 (1,49)	1,23 (1,49)	1,21 (1,49)	1,24 (1,52)
NdA4	21,61 (26,78)	21,95 (27,81)	23,75 (26,99)	22,01 (28,76)	22,06 (28,80)	23,13 (30,42)	23,14 (30,59)
NdA5	4,88 (5,11)	4,39 (4,57)	5,30 (4,97)	4,24 (4,19)	4,21 (4,20)	4,24 (4,32)	4,32 (4,41)
NdA6	24,95 (32,96)	25,93 (35,03)	27,72 (33,33)	26,63 (36,81)	26,56 (36,76)	27,43 (38,21)	27,25 (38,24)
NdA7	5,72 (7,73)	5,49 (7,43)	6,20 (7,64)	5,53 (7,41)	5,59 (7,46)	5,91 (8,09)	6,06 (8,36)
NdA8	5,78 (6,41)	5,28 (5,81)	6,10 (6,13)	4,99 (5,20)	4,97 (5,21)	5,01 (5,31)	5,17 (5,45)
NdA9	21,55 (27,77)	20,53 (26,83)	23,71 (27,97)	20,49 (26,97)	20,46 (26,91)	21,14 (28,35)	21,44 (28,95)
NdA10	21,55 (27,77)	20,53 (26,83)	23,71 (27,97)	20,49 (26,97)	20,46 (26,91)	21,14 (28,35)	21,44 (28,95)
NdA11	27,14 (33,83)	30,36 (39,79)	29,85 (33,95)	30,35 (41,63)	30,37 (41,71)	31,83 (44,14)	31,67 (44,10)
NdA12	37,47 (47,21)	43,41 (55,58)	42,15 (50,02)	45,48 (63,79*)	45,15 (63,69*)	46,05 (66,24*)	44,92 (64,71*)
NdA13	29,69 (37,72)	30,36 (39,79)	32,82 (38,12)	30,35 (41,63)	30,37 (41,71)	31,83 (44,14)	31,67 (44,10)
NdA14	32,75 (43,11)	36,66 (49,75)	36,19 (43,76)	37,26 (53,83)	37,27 (53,85)	38,95 (56,54)	38,36 (55,96)
NdA15	41,75 (51,92)	49,72 (62,38)	46,06 (53,23)	51,29 (69,57*)	51,07 (69,55*)	52,46 (73,03*)	50,55 (70,82*)

Les valeurs entre parenthèses correspondent au traitement des séquences partielles, les autres au traitement des séquences totales.

**Tableau 60 : Temps de divergences calculés dans BASEML à partir des séquences cytochrome b (357 bp) avec différents modèles de substitution**

On peut constater que pour les faibles valeurs de temps de divergence il n'existe que de petites variations en fonction du modèle de substitution utilisé. Par contre pour les valeurs élevées de temps de divergence les variations sont importantes et, de plus, un certain nombre de ces valeurs sont définies avec un écart type très important (elles sont signalées dans le tableau par '\*'), conduisant à leur retrait du traitement. D'autre part les estimations effectuées à partir des séquences partielles sont généralement un peu plus élevées que celles obtenues à partir des séquences totales.

L'ensemble de ces données permet de construire le schéma d'évolution de la **Figure 28** en utilisant les valeurs moyennes et les écart types des dates de divergence des différents nœuds (**Tableau 61**).

Nœud	Date moyenne de divergence ( ± SD )
NdA1	11,62 ± 2,80
NdA2	23,29 ± 4,49
NdA3	1,65 ± 0,52
NdA4	27,02 ± 4,93
NdA5	5,14 ± 1,72
NdA6	32,21 ± 5,28
NdA7	7,39 ± 2,06
NdA8	6,52 ± 2,00
NdA9	26,36 ± 5,09
NdA10	27,83 ± 7,01
NdA11	36,37 ± 5,84
NdA12	48,95 ± 6,37
NdA13	37,45 ± 5,47
NdA14	44,60 ± 7,47
NdA15	53,43 ± 5,67

Tableau 61 : Dates moyennes de divergence des nœuds ancestraux dans le cas des séquences de cytochrome b

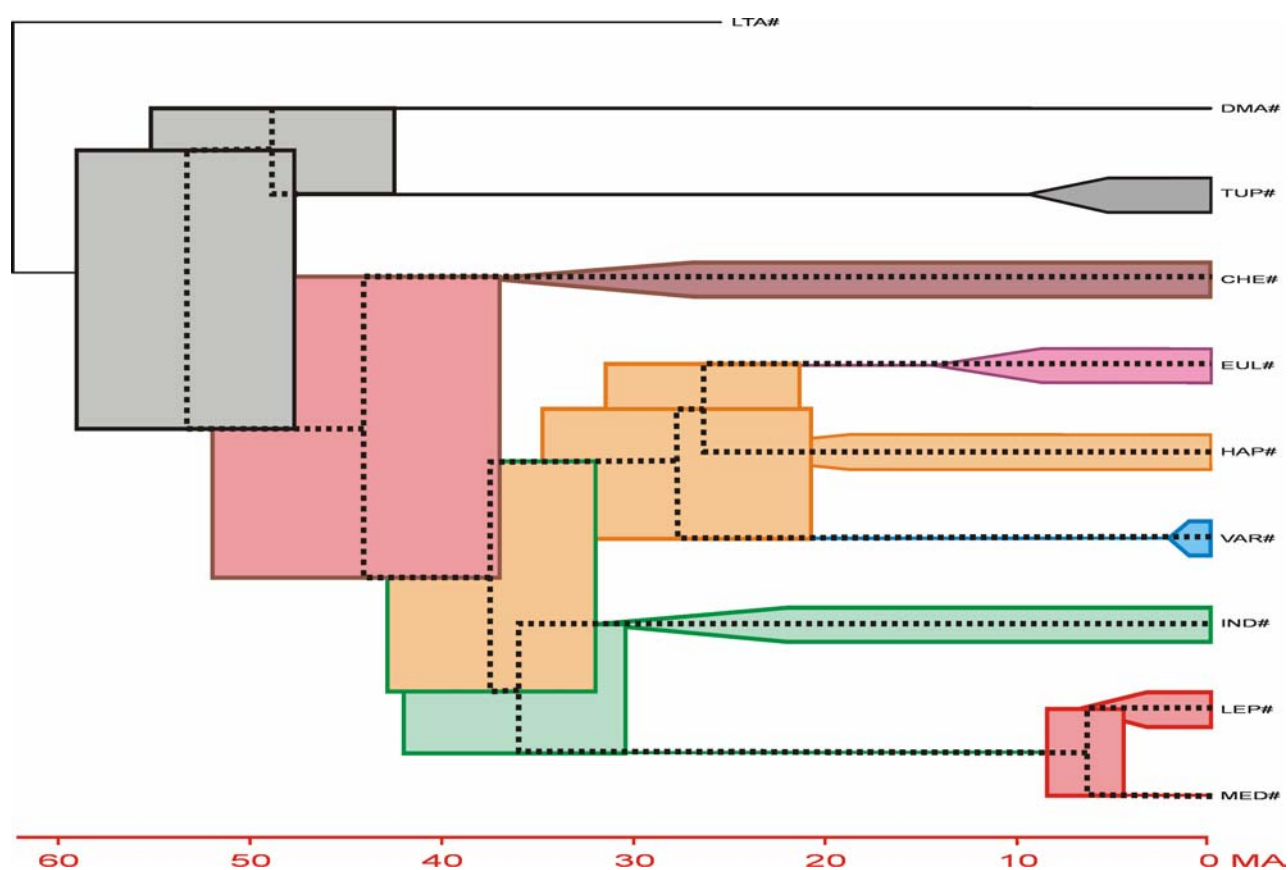


Figure 28 : Schéma d'évolution des différents groupes en fonction des données obtenues à partir des séquences de cytochrome b

Comme l'étude des taux d'évolution a montré qu'il existait des variations non négligeables entre les différents groupes, le programme BASEML est utilisé avec la topologie NJ de la **Figure 18e** afin de déterminer les taux de substitution pour les branches de l'arbre correspondant aux différents groupes. Les résultats de cette étude sont résumés dans le **Tableau 62**.

Branches du groupe	Taux de substitution
EUL	0,31598
HAP	0,46546
VAR	0,45055
IND	0,56356
MGA	0,26899
CHE	0,63983
TUP	1,34884
DMA	1,01675

**Tableau 62 : Taux de substitution pour les différents groupes (cytochrome b)**

Ces valeurs sont comparables à celles qui avaient été calculées précédemment (**Tableau 45**). Les valeurs pour les groupes VAR, DMA et TUP, qui n'avaient pu être déterminées avec les logiciels GZ-GAMMA ou TREE-PUZZLE ont pu être chiffrée en utilisant BASEML. On peut alors constater que le taux de substitution pour les Varecia est sensiblement identique à celui trouvé pour les Hapalemurs. En ce qui concerne le Daubentonia et le groupe TUP, les taux sont nettement plus élevés que celui trouvé pour les autres groupes. Enfin la seule différence entre les deux évaluations concerne le groupe des Vheirogaleidae qui avec BASEML montre le taux de substitution le plus élevé, alors qu'avec GZ-GAMMA ou TREE-PUZZLE la valeur obtenue est l'une des plus faibles de celles trouvées dans le groupe des Lemuriformes.

Il apparaît donc ici que l'existence d'horloges moléculaires locales rende mieux compte des données que l'hypothèse d'une horloge moléculaire globale.

Le logiciel BASEML permet également d'estimer le taux de mutation pour l'ensemble des séquences de la topologie étudiée. En utilisant la topologie NJ et différents modèles de substitution nucléotidique ces taux de mutation ont été estimés à partir des séquences totales, ainsi qu'à partir des séquences partielles. Les valeurs obtenues sont regroupées dans le **Tableau 63**.

Modèle de substitution	Séquences totales	Séquences partielles
JC69	4,19 10 <sup>-3</sup>	5,19 10 <sup>-3</sup>
K80	4,97 10 <sup>-3</sup>	6,24 10 <sup>-3</sup>
F81	3,92 10 <sup>-3</sup>	5,32 10 <sup>-3</sup>
F84	5,13 10 <sup>-3</sup>	6,96 10 <sup>-3</sup>
HKY85	5,11 10 <sup>-3</sup>	6,95 10 <sup>-3</sup>
TN93	5,25 10 <sup>-3</sup>	7,30 10 <sup>-3</sup>
REV	5,06 10 <sup>-3</sup>	7,08 10 <sup>-3</sup>
Moyenne	4,80 10 <sup>-3</sup>	6,43 10 <sup>-3</sup>
Ecart type	0,49	0,80

**Tableau 63 : Taux de mutation pour les différents groupes (cytochrome b)**

Les estimations effectuées à partir des séquences totales sont toujours inférieures à celles obtenues avec l'utilisation des séquences partielles. Cette différence confirme le fait que les positions 1 et 3 des codons sont les plus sujettes à des modifications. La valeur moyenne trouvée pour les séquences totales est de  $4,80 \pm 0,49 \cdot 10^{-3}$ , elle est du même ordre de grandeur que celle qui avait été précédemment définie pour l'ADN des Mammifères.

## Substitutions synonymes et non synonymes

Une partie de l'étude des substitutions synonymes et non synonymes est réalisée au moyen du logiciel Bn-Bs. Celui-ci permet, pour un groupe donné de séquences, d'une part de déterminer les nombres moyens des deux types de substitution et d'autre part, en fonction d'une topologie donnée, d'estimer les longueurs des branches en termes de substitutions synonymes ou en termes de substitutions non synonymes. Les résultats obtenus pour les différents groupes définis dans cette étude sont récapitulés au **Tableau 64**.

Groupe	Nombre moyen de substitutions synonymes	Nombre moyen de substitutions non synonymes
<b>EUL</b>	241,034	115,966
<b>HAP</b>	246,245	110,755
<b>IND</b>	245,893	111,107
<b>LEP</b>	255,174	101,826
<b>CHE</b>	247,065	109,935
<b>MGA</b>	254,429	102,571
<b>EHP</b>	245,304	111,636
<b>EHV</b>	245,291	111,709
<b>IMG</b>	250,364	106,636
<b>DTU</b>	254,268	102,732
<b>GRA</b>	249,525	107,475
<b>GRB</b>	251,066	105,934
<b>GRC</b>	252,167	104,833

**Tableau 64 : Substitutions synonymes et non synonymes pour les différents groupes (cytochrome b)**

On peut constater qu'il n'existe que de faibles variations entre les différents groupes. Il y a sensiblement deux fois plus de substitutions synonymes que de substitutions non synonymes. Ce sont les groupes LEP, MED et DTU qui présentent le plus petit nombre de sites avec des substitutions non synonymes, alors que la valeur la plus élevée est trouvée pour le groupe EUL.

Sur les 234 substitutions non synonymes possibles, 131 correspondent à la modification d'une seule des trois positions du codon. De telles substitutions sont appelées SSNCS (Single Step Nonsynonymous Codon Substitutions). L'étude de ces substitutions particulières est effectuée dans le logiciel DAMBE qui permet à la fois le calcul du nombre réel de SSNCS en fonction de la position sur les codons, et l'estimation de ces mêmes valeurs en fonction des caractéristiques des séquences. Il est alors possible, en utilisant un test du  $X^2$ , de déterminer si les valeurs observées dévient significativement ou non de celles qui peuvent être calculées. Cette étude est réalisée en utilisant es séquences ancestrales estimées pour chaque branche menant à un groupe. L'analyse est effectuée soit en prenant en compte la séquence consensus du groupe terminal, soit en omettant celle-ci. Le **Tableau 65** résume les résultats obtenus sans prise en compte de la séquence consensus du groupe terminal; l'inclusion de ces séquences consensus fournit des valeurs dont les variations sont identiques, bien qu'un peu moins prononcées.

Branche	Pourcentages observés			Pourcentages calculés			Probabilité
	Pos. 1	Pos. 2	Pos. 3	Pos. 1	Pos. 2	Pos. 3	
EUL	0,367	0,300	0,333	0,429	0,435	0,135	NS
HAP	0,219	0,281	0,500	0,429	0,435	0,136	NS
VAR	0,278	0,333	0,389	0,429	0,435	0,136	NS
IND	0,450	0,400	0,150	0,430	0,438	0,131	< 1 %
MGA	0,568	0,243	0,189	0,431	0,438	0,131	< 10 %
LEP	0,610	0,203	0,186	0,431	0,438	0,131	10 %
CHE	0,500	0,333	0,167	0,429	0,439	0,132	< 10 %
DTU	0,000	1,000	0,000	0,429	0,439	0,132	NS
TUP	0,333	0,667	0,000	0,430	0,441	0,130	NS

**Tableau 65 : Pourcentage des SSNCS observées et calculées pour les différents groupes (cytochrome b)**

Le test du  $X^2$  montre une bonne significabilité uniquement dans le cas du groupe IND avec une répartition sensiblement égale des SSNCS dans les positions 1 et 3, et des valeurs beaucoup plus faibles pour la position 3. Pour les groupes EUL et VAR, la répartition est sensiblement identique pour les trois positions. Pour le groupe HAP, les modifications localisées en position 3 sont les plus fréquentes, leur pourcentage est le double de celui trouvé pour les positions 1 et 2. Pour les groupes MGA et LEP ce sont les positions 1 qui renferment plus de la moitié des SSNCS. Enfin pour les groupes DTU et TUP aucun SSNCS n'est trouvé pour les positions 3 et ce sont les positions 2 qui présentent les pourcentages les plus élevés.

Si l'on examine maintenant les moyennes des distances de Grantham pour les trois positions du codon (**Tableau 66**), on peut constater que les substitutions non synonymes survenant à la seconde position impliquent des remplacements d'acides aminés avec des effets plus importants que pour les positions 1 et surtout que pour les positions 3.

Branche	Position 1	Position 2	Position 3
EUL	51,45	99,56	12,40
HAP	77,00	99,56	12,25
VAR	43,60	104,00	11,71
IND	56,22	104,00	10,00
MGA	62,05	99,00	10,00
LEP	62,31	96,50	10,00
CHE	34,00	104,00	10,00
DTU	-	64,00	-
TUP	22,00	43,00	-

**Tableau 66 : Distances de Grantham pour les différents groupes (cytochrome b)**

Ceci est une indication du fait que les substitutions non synonymes à la seconde position sont sujettes à une sélection de purification relativement importante, expliquant le fait qu'elles sont moins fréquentes que prévu dans la majorité des groupes analysés.

Les résultats du **Tableau 65** permettent d'évaluer l'effet pouvant être attribué à un processus de sélection en examinant la différence entre pourcentages observés et calculés. Ces résultats confirment l'existence d'une sélection de purification pour les positions 2 dans tous les groupes, sauf DTU et TUP. Par contre, pour les positions 3 la situation est inverse avec l'existence d'une sélection positive pour tous les groupes sauf DTU et TUP. Pour les positions 1 les résultats sont plus partagés avec une sélection de purification marquée pour DTU.

## Séquences protéiques

### Modifications des acides aminés

Sur les 119 acides aminés étudiés et pour l'ensemble des séquences protéiques, 42 (soit 35,29 %) sont variables et leur répartition est indiquée à la **Figure 29**. Les nombres d'acides aminés variables pour chaque groupe en fonction de l'ensemble des séquences, ou des seules séquences composant le groupe, sont reportés dans le **Tableau 67**. Les faibles valeurs intra groupe trouvées pour VAR et TUP sont sans doute à relier au faible nombre (2) de séquences composant chacun de ces groupes.

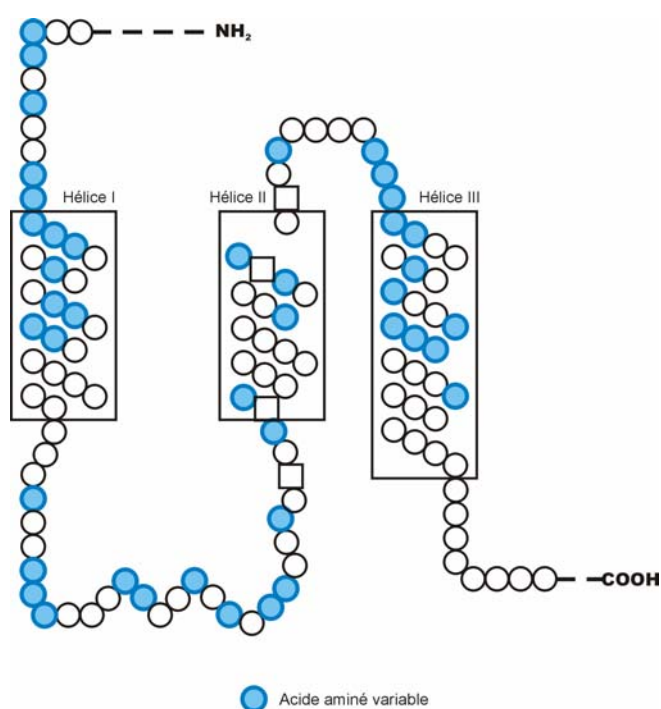


Figure 29 : Acides aminés variables pour les séquences cytochrome b

Groupe	Acides aminés variables dans le groupe	Acides aminés variables pour l'ensemble
EUL	7 (5,88 %)	7 (5,88 %)
HAP	8 (6,72 %)	10 (8,40 %)
VAR	3 (2,52 %)	6 (5,04 %)
IND	14 (11,76 %)	17 (14,28 %)
LEP	9 (7,56 %)	16 (13,44 %)
MGA	11 (9,24 %)	18 (15,13 %)
CHE	11 (9,24 %)	14 (11,76 %)
TUP	2 (1,68 %)	15 (12,60 %)
DTU	17 (14,28 %)	19 (15,97 %)

Tableau 67 : Répartition des acides aminés variables en fonction des différents groupes (cytochrome b)

L'alignement de ces acides aminés variables est donné à la **Figure 30**. Sur les 42 sites un seul, celui se trouvant à la position 20, est variable uniquement lorsque sont prises en compte toutes les séquences; cette variation apparaît uniquement dans les groupes MGA et CHE et à l'intérieur de ces groupes il n'est pas variable. Entre les groupes LEP et MGA il existe deux acides aminés (positions 9 et 11) qui sont caractéristiques de la séquence du sub-fossile MED. Par contre pour les groupes TUP et DTU il existe seulement deux acides aminés variables qui soient communs au groupe TUP et à la séquence DMA (aux positions 16 et 99), ce sont également les seuls acides aminés variables trouvés pour les séquences du groupe TUP.

	111111	2223333444	4555566777	8888899999	00
	3469012369	0234789347	9125919359	4567824789	04
ECO	NISNFGSLAA	LIIATTAVAC	DNYILALIIL	FTLTEIILFT	VA
EFR	.....	.....	.....T.	.....	.....
EMF	.....	.....	.....	.....M.	.....
EMM	.....	...G.V.W	.....	...G.	.....
ERU	.....T	.....	.....	.....	.....
LCA	.....	.....IT.	...V.	...S.	.....
HAU	.MA.....	.....T.	...V.	...S.V.	.....
HGG	.MA.....	...M...T.	..H.	...S.	.....
HGM	HMA.....	...M...T.	...V.	...S.V.	.....
HGR	.....	...M...T.	...V.	...S.V.	.....
HSI	.MA.....	.....T.	.....	...S.	.....
VVR	.....T	.V....IT.	.....	...L.	.....
VVV	.....T	.....T.	.....	...L.V.	.....
ALA	.MAKYS...	..V...T.	V.V...V.	...S...L.	.....
AOC	.MA.R....	..V...T.	...V...V.	...S...L.	.....
IIN	.MAKYS...I	.....	...V.S.V.	...S.	.....
PDI	.MAKYS...	.....T.	.K.V...V.	.V.S.A.	.....
PTA	.....	.....T.	...V...V.	.V.S.	.....
LDO	.MA...CM.T	I.....	...T...M.	...L.V...S.	.....
LED	.MA...I.T	I.....	...T...V.	...L.V...S.	.....
LMU	.NA....T	I.....	...T...V.	...L.V.HIS	.V
LRU	TMA.S...T	I.....	...T...V.	...L.V...S.	.....
LSE	.MA...M.T	I.....	...T...V.	...L.V...S.	.....
MED	.MAKYS...T	I.....	...T...V.	...L.V...S.	.....
CMA	.....	I.....T.	..H....VM	.MS	.....
MBE	.....	I...A...T.	..Q...LVM	.S.S	.....
MCO	.....	I...A...T.	..Q..I.M.VM	...S.	.....
MGR	.....T	I.....T.	..Q...LVM	...S.	.....
MMU	.....V	I.....T.	..Q...LVM	...S.	.....
MMY	.....	I.V.A...T.	..Q...LVM	...S.	.....
MRA	.....	I.....T.	..Q...LVM	...Y.	.....
MRU	.....	I...A...T.	..Q...LVM	...S.	.....
MSA	.....	I...A...T.	..Q...LVV	...Y.	.....
DMA	.....TI	..LS.S.IS.	.....T.	..YL...L.	.....
TGL	.....VV	..A...T.	...V..ALV.	YMYL..V..A	T.
TUB	.....MV	..A...T.	...V..ALV.	YMYL..V...T.	.....

Z : acide aminé caractéristique d'un groupe

Figure 30 : Alignement des acides aminés variables pour les différentes séquences (cytochrome b)

Groupe	Position des acides aminés variables caractéristiques	Nature de l'acide aminé et séquence
EUL	39	V (EMM)
	47	W (EMM)
	88	G (EMM)
HAP	-	-
VAR	22	V (VVR)
IND	49	V (ALA)
	51	K (PDI)
	61	S (IIN)
LEP – MGA	12	C (LDO) – I (LED)
	13	M (LDO et LSE)
	97	H (LMU)
	104	V (LMU)
TUP – DTU	16	T (DMA) – V (TGL) – M (TUB)
	38	S (DMA)
	84	Y (TGL et TUB)
	94	V (TGL et TUB)
	99	A (TGL)
	104	T (TGL et TUB)
CHE	59	I (MCO)
	79	V (MSA) – M (autres)
	86	M (CMA)
	87	S (CMA et MBE et MCO et MMY et MRU) – Y (MRA et MSA) – T (MMU)

**Tableau 68 : Acides aminés caractéristiques (cytochrome b)**

Un certain nombre d'acides aminés variables le sont uniquement dans un groupe ou pour un groupe, leurs caractéristiques sont indiquées dans le **Tableau 68**.

Du point de vue de la composition globale des séquences protéiques (**Tableau 69**) on peut constater que sept acides aminés: Ala, Gly, Ile, Leu, Phe, Ser et Thr sont les plus représentés, que ce soit sur l'ensemble des séquences ou dans les groupes pris indépendamment. La Lysine est l'acide aminé le moins représenté, on ne le trouve que dans la séquence MED et le groupe IND. Par ailleurs la Proline est en proportion constante dans tous les groupes.



Acide aminé	EUL	HAP	VAR	IND	LEP	MED	MGA	CHE	DMA	TUP	DTU	Total
Ala	7,7	8,1	6,7	8,2	8,2	8,4	8,3	7,8	5,0	8,0	7,0	7,9
Cys	2,4	2,5	2,5	2,5	2,7	2,5	2,7	2,5	2,5	2,5	2,5	2,5
Asp	1,7	1,7	1,7	1,5	1,7	1,7	1,7	1,7	1,7	1,7	1,7	1,7
Glu	0,7	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8
Phe	8,4	8,4	8,4	7,6	8,1	7,6	8,0	8,4	7,6	7,6	7,6	8,1
Gly	10,4	10,1	10,1	9,4	10,1	9,2	9,9	10,1	10,1	10,1	10,1	10,0
His	3,4	3,6	3,4	3,4	3,5	3,4	3,5	3,5	3,4	3,4	3,4	3,5
Ile	9,1	7,8	8,8	6,6	7,1	6,7	7,0	8,6	9,2	5,0	6,4	7,8
Lys	0,0	0,0	0,0	0,7	0,0	0,8	0,1	0,0	0,0	0,0	0,0	0,1
Leu	11,8	11,8	12,6	12,1	11,3	11,8	11,3	10,5	13,4	11,8	12,3	11,5
Met	4,4	5,2	4,2	4,9	5,4	5,0	5,3	5,1	4,2	5,5	5,0	5,0
Asn	4,2	4,1	4,2	3,5	4,2	3,4	4,1	4,2	4,2	4,2	4,2	4,1
Pro	1,7	1,7	1,7	1,7	1,7	1,7	1,7	1,7	1,7	1,7	1,7	1,7
Gln	1,7	1,7	1,7	1,7	1,7	1,7	1,7	2,4	1,7	1,7	1,7	1,9
Arg	2,5	2,5	2,5	2,7	2,5	2,5	2,5	2,5	2,5	2,5	2,5	2,5
Ser	7,6	7,8	7,6	8,4	7,4	8,4	7,6	8,2	10,1	7,6	8,4	8,0
Thr	8,7	8,0	9,2	7,9	8,6	8,4	8,5	8,0	8,4	8,0	8,1	8,3
Val	3,5	4,2	3,8	5,9	5,0	5,0	5,0	4,5	2,5	6,3	5,0	4,6
Trp	5,2	5,0	5,0	5,0	5,0	5,0	5,0	5,0	5,0	5,0	5,0	5,1
Tyr	5,0	4,9	5,0	5,5	5,0	5,9	5,2	4,4	5,9	6,7	6,4	5,1

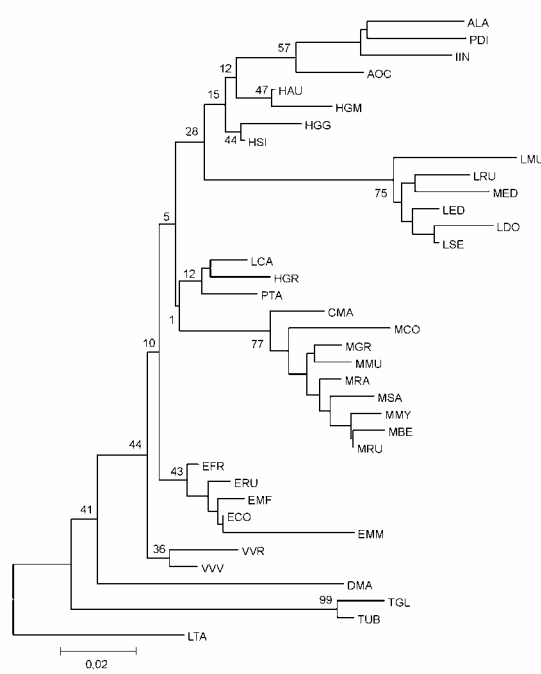
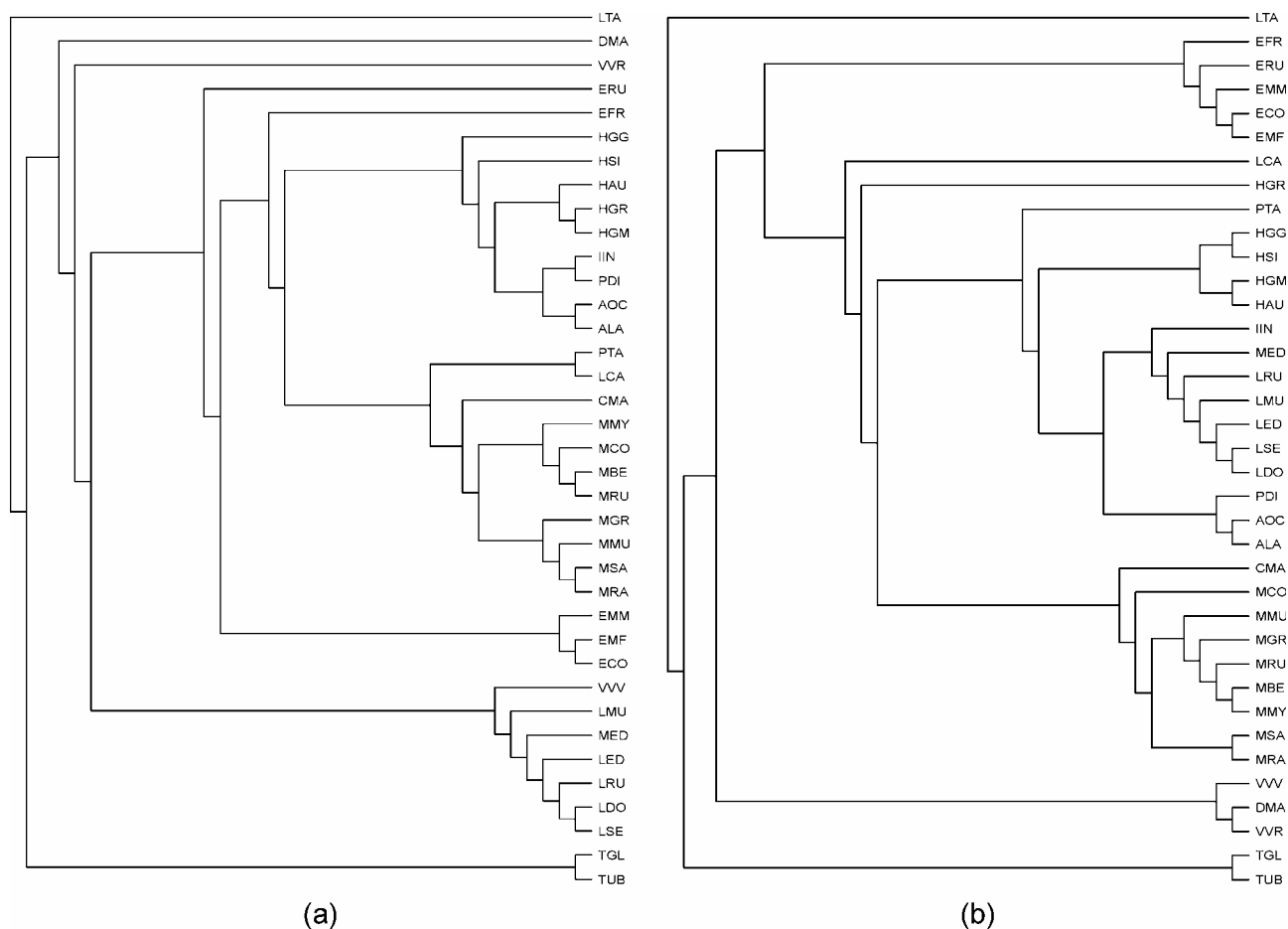
Tableau 69 : Répartition des différents acides aminés (cytochrome b) dans les différents groupes

## Reconstructions phylogénétiques

Trois méthodes de reconstruction ont été utilisées avec les séquences protéiques:

- **Maximum Likelihood (ML)**: le programme PROTML du pack PHYLIP est utilisé. L'arbre final est obtenu après réarrangements globaux. L'analyse est répétée 10 fois en utilisant à chaque fois un ordre aléatoire d'introduction des séquences.
- **Maximum Parsimony (MP)**: le programme PROTPARS du pack PHYLIP est utilisé. Les arbres les plus parcimonieux sont obtenus après réarrangements globaux. L'arbre final est le consensus strict des arbres les plus parcimonieux (calculé dans le programme CONSENSE du pack PHYLIP). Le traitement est répété 10 fois avec un ordre aléatoire d'introduction des séquences.
- **Neighbor-Joining (NJ)**: le traitement est effectué dans le logiciel MEGA en utilisant des distances de Poisson avec corrections et un test de bootstrap de 5 000 répliques.

Les résultats obtenus sont présentés à la **Figure 31**. L'arbre MP correspond au consensus de 100 arbres.



**Figure 31 : Reconstructions phylogénétiques à partir des séquences protéiques de cytochrome b**  
 Ces trois topologies (a) pour ML, (b) pour MP et (c) pour NJ) sont comparées en utilisant le programme CODEML du pack PAML, avec la méthode RELL et celle décrite par Shimodaira et Hasegawa (1999). Dans les deux cas un test de bootstrap avec 10 000 répliquions est pratiqué. Ces

tests ont été réalisés en considérant l'existence ou non d'une horloge moléculaire globale. Les résultats sont présentés au **Tableau 70**.

Topologie	Sans horloge			Avec horloge		
	<i>LnL</i>	<i>pSH</i>	<i>pRELL</i>	<i>LnL</i>	<i>pSH</i>	<i>pRELL</i>
ML	- 1166,699	0,485	0,327	- 1209,245	0,395	0,255
MP	- 1159,489	- 1,000	0,428	- 1205,487	0,489	0,211
NJ	- 1164,510	0,599	0,244	- 1194,845	- 1,000	0,534

**Tableau 70 : Comparaison des topologies obtenues à partir des séquences protéiques de cytochrome b**

Sans horloge moléculaire la crédibilité des trois topologies est, dans l'ordre, MP, NJ et ML, alors que si l'on admet l'existence d'une horloge moléculaire globale on trouve: NJ, MP et ML. Compte tenu des résultats obtenus à partir des séquences nucléotidiques correspondant aux séquences protéiques étudiées ici, on peut considérer que la topologie NJ constitue l'arbre 'correct'.

Dans cette topologie NJ, la séquence DMA et le groupe TUP sont situés à la position la plus basale de l'arbre mais ils ne forment pas un cluster comme dans le cas des séquences nucléotidiques. Les séquences des groupes EUL, VAR, CHE et MGA apparaissent groupées dans les quatre clusters correspondants. Pour le groupe IND, seule la séquence PTA est située en dehors du cluster formé par les séquences ALA, AOC, IIN et PDI. Par contre les séquences du groupe HAP sont plus dispersées dans l'arbre: LCA et HGR sont groupées avec PTA, les autres sont groupées en deux clusters, l'un formé de HAU et HGM, et l'autre de HGG et HSI; ces deux derniers clusters se trouvant à la base de celui formé par le groupe IND. Cet ensemble forme un cluster qui est le cluster frère de celui du groupe MGA. Les séquences LCA, HGR et PTA forment un cluster qui est le cluster frère du groupe CHE. Les clusters des groupes EUL et VAR sont en position basale de l'arbre, VAR étant le plus basal après DMA et TUP. Le sub-fossile MED est bien inclus dans le groupe des Lepilemuridae mais contrairement à la situation trouvée avec les séquences nucléotidiques, il n'est pas ici en position basale puisqu'on le trouve associé à la séquence LRU. Les clusters correspondant aux groupes MGA et IND forment deux clusters frères. Toutefois, pour l'ensemble de l'arbre, les valeurs de bootstrap supportant les différents clusters sont faibles.

La topologie MP est pratiquement identique à la topologie NJ. Les différences essentielles portent sur la séquence IIN qui se trouve à la base du cluster du groupe MGA. Les autres séquences du groupe IND, à l'exception de PTA, forment un cluster frère du précédent. Par contre la séquence DMA se trouve dans le cluster des séquences du groupe VAR.

Dans le cas de l'arbre ML on ne retrouve que très peu des organisations décrites ci-dessus.

Compte tenu de la topologie NJ, le groupe HAP a été subdivisé en trois sous groupes pour la suite des analyses: HAP1 (HAU et HGM), HAP2 (HGG et HSI) et HAP3 (LCA et HGR). Les longueurs des branches principales de cette topologie NJ ont été estimées (Tableau 71). La branche la moins bien définie est celle reliant le cluster du groupe CHE associé au cluster regroupant HAP3 et PTA à l'ensemble formé par MGA, IND, HAP1 et HAP2. La séquence DMA, de même que le groupe TUP sont très bien définis par rapport à l'ensemble des autres séquences.

Branche	Longueur	Pourcentage *
Arbre complet	0,98345	
IND – HAP1	0,025938	2,64 %
HAP1 – HAP2	0,025914	2,63 %
(IND,HAP1,HAP2) – MGA	0,023065	2,34%
HAP3 – PTA	0,017926	1,82 %
(HAP3,PTA) – CHE	0,050193	5,10 %
(HAP3,PTA,CHE) – (IND,HAP1,HAP2,MGA)	0,006406	0,65 %
EUL – (MGA,IND,HAP,CHE)	0,026347	2,68 %
VAR – (EUL,MGA,IND,HAP,CHE)	0,036837	3,74 %
DMA – (EHV,MGA,IND,CHE)	0,10389	10,56 %
TUP – (DMA,EHV,MGA,IND,CHE)	0,082716	8,41 %

\* Pourcentage de la longueur de la branche par rapport à la longueur totale de l'arbre.

Tableau 71 : Longueurs des branches principales de la topologie NJ pour les séquences protéiques de cytochrome b

Les tests menés au moyen du logiciel PHYLTEST (**Tableau 72**) confirment la crédibilité de l'association entre DMA et LEP et de l'association de ces deux clusters avec les séquences HAP et IND. Il existe de nombreux quartets pour lesquels la résolution est moyenne, ce qui confirme les résultats du 'Likelihood mapping' sur les séquences protéiques (**Figure 13**). Ce résultat est également à mettre en correspondance avec les faibles valeurs de bootstrap trouvées au cours de l'analyse NJ. Les conclusions concernant les groupes dont l'association est très peu probable sont similaires à celles formulées dans le cas des séquences nucléotidiques.

Topologie	Branche interne	CP	Probabilité
(IND,HAP) , (MGA,VAR)	0,00992459	0,535	NS
(IMG,HAP) , (CHE+DTU,EUL+VAR)	0,00700613	0,522	NS
(EUL,VAR) , (HAP,IMG)	0,0118726	0,806	NS
(TUP,DMA) , (CHE,IMG+EHV)	0,0158315	0,758	NS
(MED,LEP) , (IND,HAP)	0,0428598	0,979	< 5 %
(IMG,EHV) , (CHE,DMA+TUP)	0,00662119	0,616	NS
(HAP,IND) , (MGA,CHE)	0,00382459	0,281	NS
(HAP+IND+MGA,CHE) , (EUL,VAR)	0,00669353	0,673	NS
(DMA,TUP) , (VAR,EUL+HAP+MGA+CHE)	0,0117224	0,803	NS
(DTU,VAR) , (EUL,HAP+MGA+CHE)	0,00667289	0,663	NS

Tableau 72 : Test des branches internes pour différentes topologies à partir des séquences protéiques de cytochrome b

## Reconstruction des séquences ancestrales

La reconstruction des séquences protéiques ancestrales est effectuée en utilisant deux approches, avec les logiciels ANC-GENE et CODEML du pack PAML. La première approche utilise la traduction des séquences nucléotidiques ancestrales calculées à partir de la topologie NJ de la **Figure 18e**. La seconde utilise les séquences protéiques actuelles avec la topologie NJ de la **Figure 31c**. Pour cette dernière les nœuds analysés sont indiqués à la **Figure 32** et au **Tableau 73**.

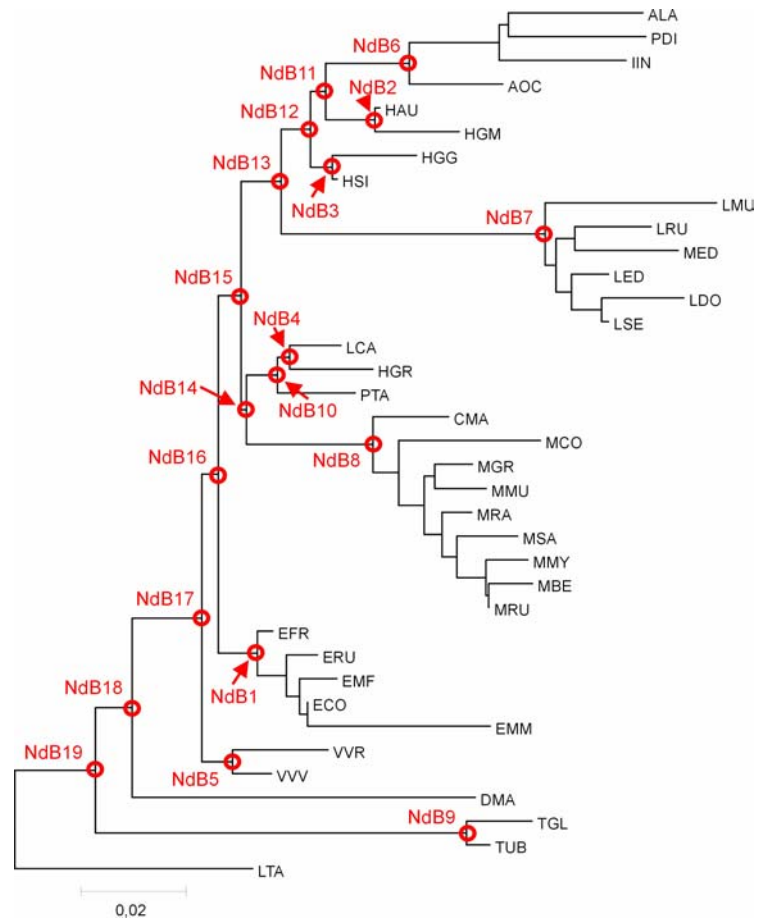


Figure 32 : Nœuds ancestraux analysés pour les séquences protéiques de cytochrome b

Nœud	Groupe de séquences
NdB1	EUL
NdB2	HAP1
NdB3	HAP2
NdB4	HAP3
NdB5	VAR
NdB6	IND
NdB7	MGA
NdB8	CHE
NdB9	TUP
NdB10	PTA + HAP3
NdB11	IND + HAP1
NdB12	IND + HAP1 + HAP2
NdB13	MGA + IND + HAP1 + HAP2
NdB14	PTA + HAP3 + CHE
NdB15	IND + HAP1 + HAP2 + HAP3 + PTA + MGA + CHE
NdB16	EUL + IND + HAP1 + HAP2 + HAP3 + PTA + MGA + CHE
NdB17	VAR + EUL + IND + HAP1 + HAP2 + HAP3 + PTA + MGA + CHE
NdB18	DMA + VAR + EUL + IND + HAP1 + HAP2 + HAP3 + PTA + MGA + CHE
NdB19	TUP + DMA + VAR + EUL + IND + HAP1 + HAP2 + HAP3 + PTA + MGA + CHE

Tableau 73 : Identification des nœuds ancestraux pour les séquences protéiques de cytochrome b

Pour l'ensemble des séquences reconstruites par les deux approches, il y a 21 acides aminés variables (soit 17,65 % des 119 acides aminés étudiés) et ces acides aminés variables occupent les mêmes positions pour les deux séries de séquences. Leur localisation, par rapport aux acides aminés variables des séquences actuelles, est indiquée à la **Figure 33**.

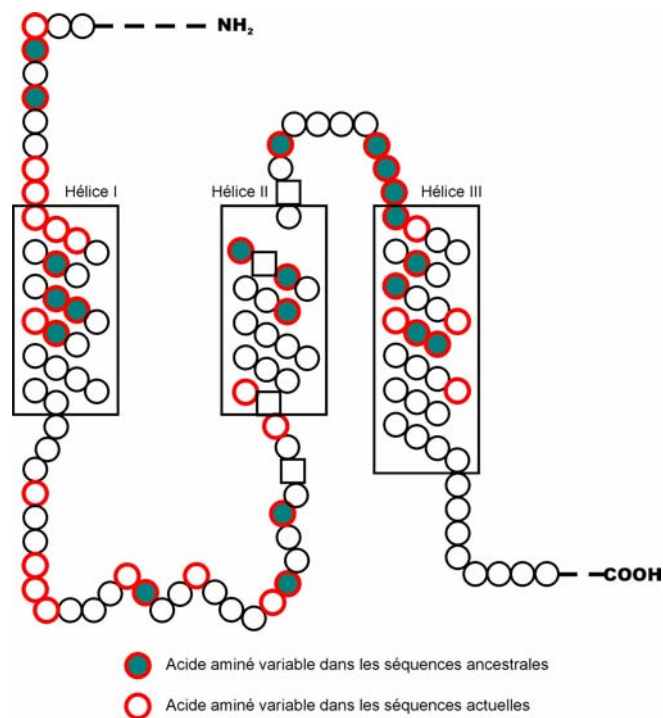


Figure 33 : Acides aminés variables pour les séquences protéiques ancestrales et actuelles de cytochrome b

La comparaison des séquences ancestrales estimées à partir de la traduction des séquences nucléotidiques ancestrales montre les identités suivantes:

- ✓ Séquences NdA5 et NdA8 (ancestrales de LEP et MGA); ce deux séquences étant également identiques à la séquence protéique consensus des Lepilemuridae actuels.
- ✓ Séquences NdA9 et NdA10 (ancestrales de EHP et EHV).
- ✓ Séquence NdA1 (ancestrale de EUL) et consensus des séquences protéiques des Eulemurs actuels.

Pour ce qui concerne les reconstructions effectuées directement à partir des séquences protéiques actuelles on trouve les identités suivantes:

- ✓ Séquences NdB19 et NdB18 (ancestrales de TUP et DMA).
- ✓ Séquences NdB16, NdB15 et NdB14 (ancestrales des Lemuriformes à l'exception des Varecia).
- ✓ Séquences NdB13, NdB12 et NdB3.
- ✓ Séquences NdB10 et NdB4.
- ✓ Séquence NdB7 et consensus des séquences protéiques des Lepilemuridae actuels.

Enfin il existe également des identités entre les deux séries de séquences ancestrales:

- ✓ Séquences NdB13, NdB12, NdB3 et NdA2 (ancestrale du groupe HAP).
- ✓ Séquences NdB5 et NdA3 (ancestrales du groupe VAR).
- ✓ Séquences NdB7 et NdA5, NdA8 (ancestrales du groupe MGA).
- ✓ Séquences NdB16, NdB15, NdB14 et NdA9, NdA10 (ancestrales du groupe EHV).

Parmi les séquences ancestrales correspondant à un groupe identique dans les deux séries, il n'existe en général que de faibles variations de composition:

- ✓ 1 acide aminé variable entre NdB1 et NdA1 (ancestrales de EUL); entre NdB6 et NdA6 (ancestrales de IND); entre NdB3 et NdA7 (ancestrales de TUP).
- ✓ 2 acides aminés variables entre NdB6 et NdA4 (ancestrales de IND).
- ✓ 4 acides aminés variables entre NdB4, NdB9, NdB2 et NdA2 (ancestrales de HAP).

Ces résultats montrent que les deux voies de reconstruction utilisées, même si elles n'utilisent pas des topologies de base identiques, conduisent à des reconstructions de séquences ancestrales très proches l'une de l'autre.

La comparaison des nombres d'acides aminés variables pour les branches menant aux différents groupes (**Tableau 74**) montre que ce sont les branches menant à MGA et TUP qui sont celles présentant le plus de modifications. Ce résultat est identique que soit ou non prise en compte la séquence consensus du groupe terminal. On peut également remarquer qu'il y a toujours un peu plus d'acides aminés variables si la séquence consensus est incluse dans l'analyse. Les deux seules différences marquées entre les deux voies de reconstruction sont trouvées pour les groupes EUL et VAR pour lesquels le nombre d'acides aminés variables est beaucoup plus faible dans le cas de la reconstruction à partir des séquences nucléotidiques ancestrales qu'à partir des séquences protéiques actuelles.

Branche menant à	Reconstruction à partir des séquences nucléotidiques		Reconstruction à partir des séquences protéiques	
	<i>Sans consensus</i>	<i>Avec consensus</i>	<i>Sans consensus</i>	<i>Avec consensus</i>
EUL	7	7	3	4
HAP	8	9	-	-
VAR	6	7	2	8
IND	7	11	7	10
MGA	12	-	11	-
LEP	12	12	-	-
CHE	8	10	7	9
DTU	1	11	-	-
TUP	10	11	10	11
EHP	6	-	-	-
EHV	6	-	-	-
IMG	7	-	-	-
GRA	-	-	-	-
GRB	5	-	-	-
HAP1	-	-	7	-
HAP2	-	-	5	-
HAP3	-	-	4	-

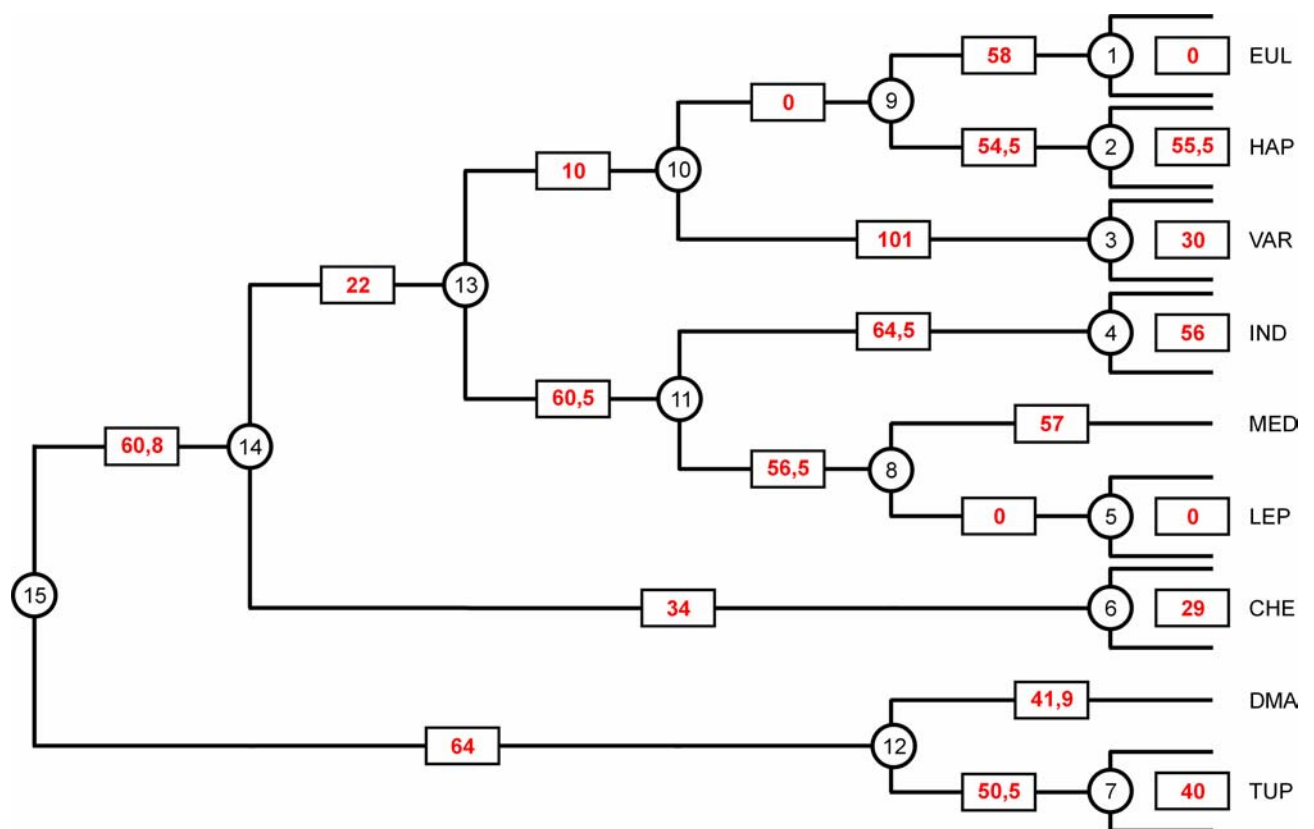
**Tableau 74 : Acides aminés ancestraux variables (cytochrome b) à partir de la traduction des séquences nucléotidiques ancestrales et à partir des séquences protéiques ancestrales**

Les modifications de nœud à nœud ont été examinées, pour les deux voies de reconstruction, en utilisant la valeur moyenne des distances de Grantham pour les changements entre nœuds consécutifs et éventuellement la séquence consensus terminale (séquence actuelle). La théorie de la neutralité de l'évolution moléculaire suppose que le taux de substitution diminue avec l'augmentation de la non similitude des acides aminés modifiés. Si l'on considère que les distances de Grantham sont une bonne mesure de la non similarité des acides aminés, lorsque cette distance augmente il doit y avoir une diminution du taux de substitution. Pour Kimura (168) la plupart des substitutions entre acides aminés implique des formes similaires (pour lesquelles les distances de Grantham sont faibles) et ceci supporte la théorie de la neutralité. Par contre pour Gillespie (106) les substitutions les plus fréquentes ne se produisent pas entre les acides aminés les plus semblables, mais plutôt entre ceux présentant des distances de Miyata (223) voisines de 1, c'est à dire avec des distances de Grantham comprises entre 70 et 80 environ.

Dans le cas de la reconstruction effectuée à partir des séquences nucléotidiques ancestrales (**Figure 34**) on peut remarquer l'existence de différences marquées. Les valeurs nulles indiquent ici que les séquences pour les nœuds considérés sont identiques. La valeur la plus faible de la moyenne des distances de Grantham (10) est obtenue pour NdA13 – NdA10, c'est à dire que le taux de substitution apparaît fortement augmenté au départ de la branche menant à l'ensemble Eulemurs,



Hapalemurs et Varecia; alors que ce taux est plus faible pour les branches menant aux Indriidae et Megaladapidae. On peut également remarquer que le taux de substitution pour la branche menant aux Varecia semble faible (distance de Grantham de 101 pour NdA10 – NdA3); l'évolution de ce groupe serait donc ralentie depuis sa séparation d'avec les Eulemurs et Hapalemurs. Par contre dans le cas des Cheirogaleidae, la distance moyenne de Grantham est faible (34 pour NdA14 – NdA6) indiquant une évolution relativement rapide. Pour ce groupe, comme pour celui des Varecia, l'évolution actuelle est encore importante, les distances de Grantham entre le nœud terminal et la séquence consensus actuelle étant faible (respectivement de 29 et 30). Pour cette reconstruction les substitutions les plus fréquentes, pour l'ensemble de la topologie, impliquent des acides aminés pour lesquels les distances de Grantham sont comprises entre 50 et 65, ce qui semblerait indiquer que l'hypothèse de Gillespie rende mieux compte des données que celle de Kimura pour ce qui est de la théorie de la neutralité de l'évolution.



**Figure 34 : Modifications des acides aminés ancestraux à partir de la traduction des séquences nucléotidiques ancestrales**

Pour la reconstruction effectuée à partir des séquences protéiques actuelles (**Figure 35**), la valeur la plus élevée de la moyenne des distances de Grantham est trouvée pour NdB17 – NdB16 (144), correspondant à la branche menant à l'ensemble des Lemuriformes, hors mis le groupe des Varecia, DMA et le groupe des Tupaia. Comme précédemment la branche menant à CHE (NdB14 – NdB8) montre une évolution relativement rapide avec une distance de Grantham moyenne de 37. Ceci est également le cas pour les branches menant aux trois sous groupes de Hapalemurs. Ce résultat est en apparence contradiction avec celui trouvé dans le cas de la reconstruction précédente; mais le fait que le groupe HAP ait du ici être subdivisé a sans doute eut pour effet de supprimer les différences les plus marquées en regroupant les séquences les plus identiques. Par contre pour le groupe EUL les résultats sont parfaitement concordants pour les deux voies de reconstruction. Ici

encore, mais peut être d'une manière un peu moins prononcée que dans le cas précédent, l'hypothèse de Gillespie semble le mieux rendre compte des données.

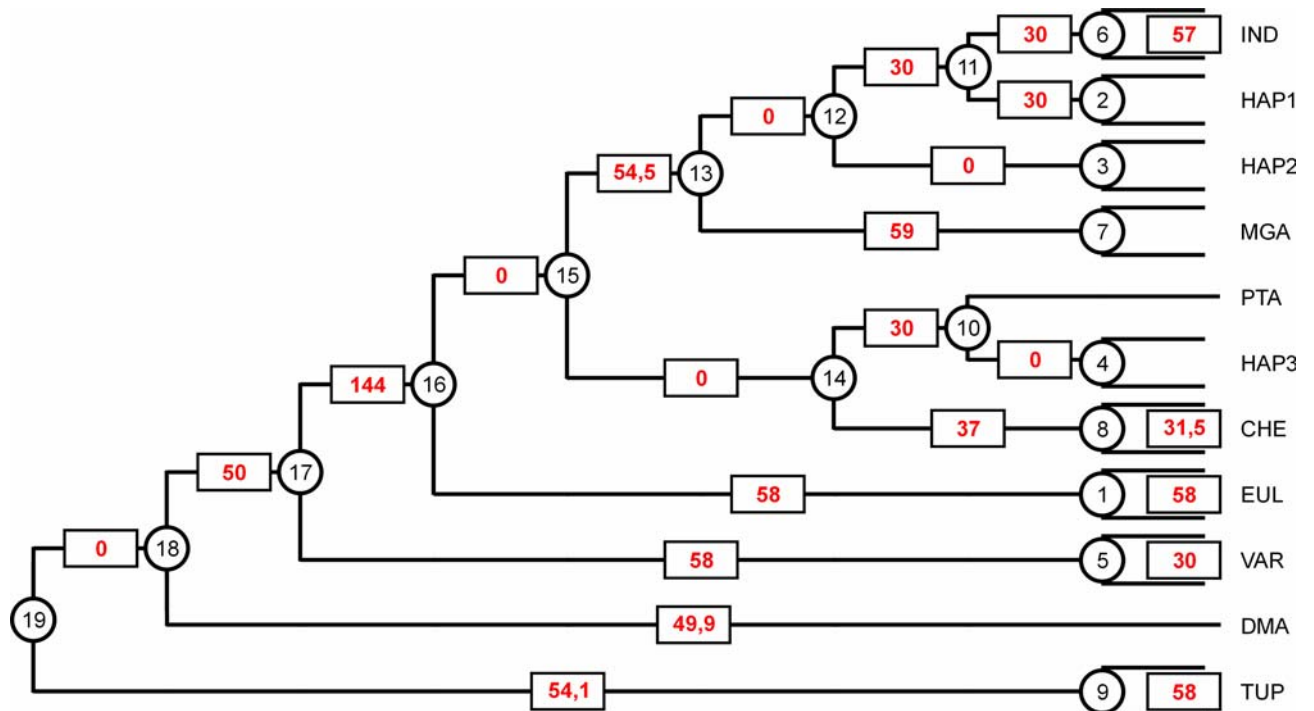


Figure 35 : Modifications des acides aminés ancestraux à partir de la traduction des séquences protéiques ancestrales

## Horloge moléculaire et arbres linéarisés

L'arbre NJ reconstruit dans LINETREE, en utilisant les distances de Poisson avec correction, (**Figure 36**) est conforme à celui obtenu dans MEGA (**Figure 31c**). Les résultats de l'analyse 'deux clusters' effectuée dans LINETREE, ainsi que les valeurs de bootstrap pour cette topologie sont résumés dans le **Tableau 75**. Le test 'deux clusters' ne montre aucune séquence dont l'évolution soit significativement plus lente que la moyenne de l'ensemble des séquences. Le test du  $X^2$  fournit une valeur de 105,960671 ce qui avec 34 ddl donne une significabilité inférieure à 0,1 % (valeur critique de 65,2472). On ne peut donc ici rejeter l'hypothèse de la constance du taux d'évolution au long de l'arbre considéré.

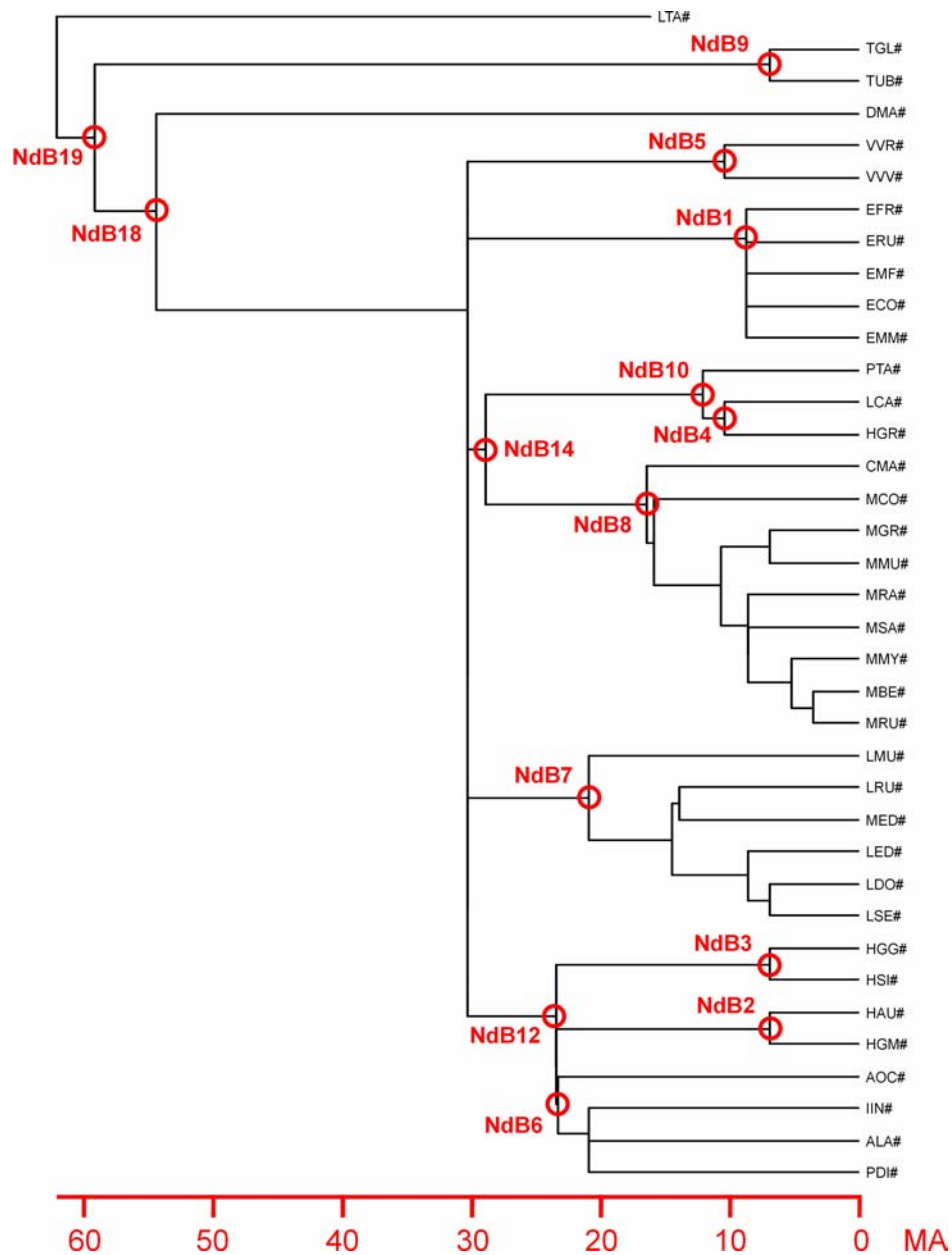


Figure 36 : Arbre NJ obtenu dans LINETREE à partir des séquences protéiques (cytochrome b)

Cluster A	Cluster B	Bootstrap (%)	bA	bB	Z	CP (%)
IND	HAP1	12	0,049130	0,020701	1,783376	92,50
IND, HAP1	HAP2	15	0,042695	0,015342	2,391452	98,32
MGA	IND, HAP1, HAP2	28	0,063060	0,040039	1,604980	89,04
PTA	HAP3	12	0,013062	0,016825	0,384385	29,60
CHE	PTA, HAP3	1	0,050419	0,021353	1,925557	94,52
CHE, PTA, HAP3	MGA, IND, HAP1, HAP2	5	0,041140	0,064188	1,186372	76,20
EUL	CHE, PTA, HAP3, MGA, IND, HAP1, HAP2	10	0,027471	0,059749	2,254265	97,56
VAR	EUL, CHE, PTA, HAP3, MGA, IND, HAP1, HAP2	44	0,022464	0,052739	2,331343	98,02
DMA	VAR, EUL, CHE, PTA, HAP3, MGA, IND, HAP1, HAP2	41	0,069431	0,065846	0,153414	11,92
TUP	DMA, VAR, EUL, CHE, PTA, HAP3, MGA, IND, HAP1, HAP2	-	0,078182	0,068766	0,360296	28,12

Tableau 75 : Test 'deux clusters' sur la topologie NJ (séquences protéiques de cytochrome b)

Ces données permettent la reconstruction de l'arbre linéarisé de la **Figure 37**. Celui-ci présente un certain nombre de nœuds qui ne peuvent être résolus.



**Figure 37 : Arbre linéarisé obtenu à partir des séquences protéiques de cytochrome b**

En utilisant la même référence que pour les séquences nucléotidiques (62 MA pour la divergence entre Lorisidae et Strepsirhini) il est possible de calculer les temps de divergence des nœuds définis (**Tableau 76** – arbre linéarisé 1).

Nœud	Divergence	Arbre linéarisé 1	Arbre linéarisé 2
NdB1	EUL	8,55	-
NdB2	HAP1	6,80	-
NdB3	HAP2	6,80	-
NdB4	HAP3	10,25	-
NdB5	VAR	10,25	-
NdB6	IND	23,21	-
NdB7	MGA	20,79	10,05
NdB8	CHE	16,36	-
NdB9	TUP	6,80	-
NdB10	PTA – HAP3	11,98	-
NdB12	IND, HAP1 – HAP2	23,27	-
NdB14	PTA, HAP3 – CHE	28,78	-
NdB18	DMA – VAR, EUL, CHE, PTA, HAP3, MGA, IND, HAP1, HAP2	54,25	52,32
NdB19	TUP – DMA, VAR, EUL, CHE, PTA, HAP3, MGA, IND, HAP1, HAP2	58,93	59,11

**Tableau 76 : Temps de divergence estimés à partir des arbres linéarisés (séquences protéiques de cytochrome b)**

Une étude similaire est effectuée avec les séquences consensus de chaque groupe. Dans ce cas l'arbre est reconstruit en utilisant les distances de Poisson avec correction et l'arbre linéarisé calculé est reproduit à la **Figure 38**. Le test du  $X^2$  effectué à partir des analyses 'deux clusters' sur cette topologie fournit une valeur de 19,691429, ce qui avec 7 ddl donne une significabilité inférieure à 1 % (valeur critique de 18,4753) ne permettant pas, comme avec l'ensemble des séquences, de rejeter l'hypothèse d'une constance du taux d'évolution. D'autre part l'utilisation des séquences consensus ne permet pas de résoudre des nœuds supplémentaires. Les trois dates de divergence qui peuvent être calculées dans ce cas sont reportées dans le **Tableau 76** (arbre linéarisé 2). On peut constater que, si pour les divergences les plus anciennes il existe peu de différences entre les estimations fournies par les deux arbres linéarisés, par contre pour la divergence du groupe MGA la valeur obtenue à partir de l'ensemble des séquences est double de celle obtenue à partir des séquences consensus. Si maintenant ces valeurs sont comparées à celles obtenues à partir des séquences nucléotidiques (**Tableau 58**) les variations sont très importantes. La seule estimation concordante est celle de la divergence des Megaladapidae qui est estimée à  $10,15 \pm 0,56$  MA avec les séquences nucléotidiques et à 10,05 MA avec les séquences protéiques consensus.

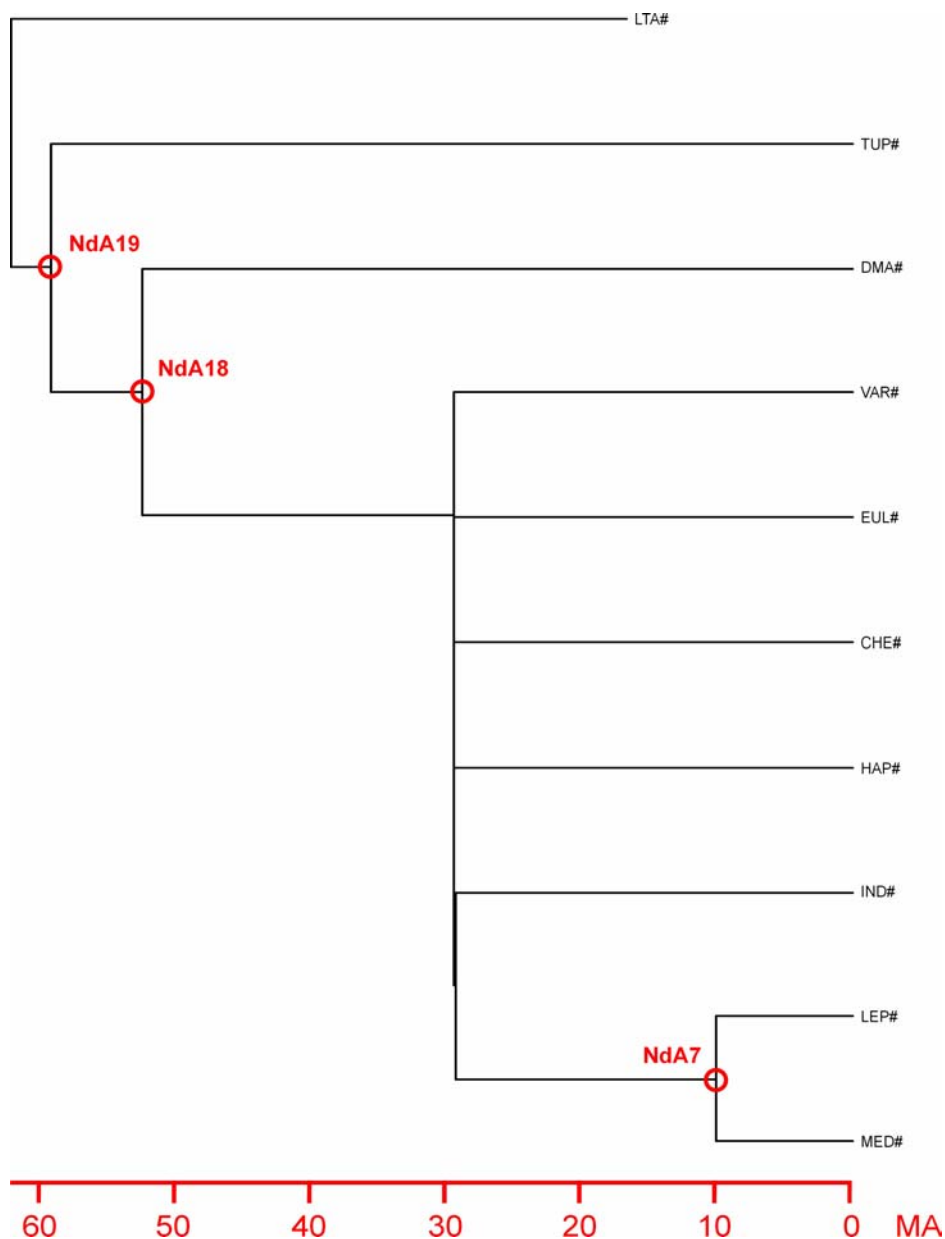


Figure 38 : Arbre linéarisé obtenu à partir des séquences protéiques consensus (cytochrome b)

Les temps de divergence ont également été estimés avec l'option 'TipDate' du programme CODEML. Les résultats de cette analyse sont reportés au **Tableau 77**. On peut remarquer que toutes les estimations sont nettement plus faibles que celles pouvant être effectuées dans LINETREE. Le taux de mutation, estimé avec CODEML, est de  $2,90 \cdot 10^{-3}$  en prenant en compte l'ensemble des séquences protéiques disponibles et la topologie NJ de la **Figure 31c**.

Nœud	Temps de divergence	Ecart type
NdB1	6,00	2,09
NdB2	3,99	2,43
NdB3	5,67	3,89
NdB4	6,18	5,28
NdB5	6,88	3,65
NdB6	9,34	0,00
NdB7	4,92	1,53
NdB8	6,67	1,18
NdB9	2,68	1,78
NdB10	6,18	4,14
NdB11	11,15	0,00
NdB12	12,88	0,00
NdB13	12,88	0,00
NdB14	15,09	0,00
NdB15	15,09	0,00
NdB16	15,09	0,00
NdB17	17,31	0,00
NdB18	26,58	1,57
NdB19	28,99	0,00

**Tableau 77 : Temps de divergence moyens estimés à partir des séquences protéiques de cytochrome b**

Il apparaît donc que les séquences protéiques, du moins celles de 119 acides aminés, utilisées dans ce travail ne sont pas en mesure de fournir des informations fiables concernant les divergences des différents groupes. Ceci est sans doute du, en partie, au fait que les groupes étudiés sont proches l'un de l'autre, faisant que les méthodes d'estimation des divergences deviennent inopérantes.

## ***Traitements sur les séquences 12S rRNA***

### **Substitutions nucléotidiques par site**

A l'aide du logiciel GZ-GAMMA et des topologies NJ (utilisant les distances de Tamura-Nei prenant en compte les transitions et les transversions) reconstituées pour les différents groupes, les taux de substitution par site ont été étudiés pour les séquences 12S totales, ainsi que pour celles représentant les zones 'Loop'. Pour chaque groupe, les 'gaps' présents dans les séquences ont été supprimés en utilisant une option 'complete deletion'. Des estimations identiques ont été réalisées à l'aide de TREE-PUZZLE, mais dans ce cas les gaps n'ont pas été éliminés.

Les résultats de ces analyses sont résumés dans les **Tableaux 78** (séquences 12S totales) et **79** (séquences correspondant aux zones 'Loop'). Les calculs pour le groupe CHE n'ont pas été possibles car seulement trois séquences sont disponibles pour celui-ci.

Dans le cas des séquences 12S totales, on constate une grande variabilité du taux de substitution pour les groupes EUL et HAP. La valeur la plus faible est trouvée pour le groupe IND, comme cela était le cas avec les séquences de cytochrome b, et cette valeur est sensiblement identique à celle qui peut être estimée sur l'ensemble des séquences (groupe GRC).

Groupe	Alpha 1 *	Alpha 2 **	Variance de la distribution gamma
EUL	0,086745	0,02 ± 0,13	42,24
HAP	0,065914	0,09 ± 0,04	11,09
IND	0,365577	0,43 ± 0,30	2,30
LEP	0,116981	0,21 ± 0,16	4,68
MGA	0,100040	0,22 ± 0,13	4,61
EHP	0,194396	0,21 ± 0,07	4,68
EHV	0,207607	0,21 ± 0,05	4,83
IMG	0,330392	0,36 ± 0,07	2,78
GRA	0,363762	0,33 ± 0,04	3,04
GRB	0,358799	0,32 ± 0,04	3,14
GRC	0,373432	0,31 ± 0,03	3,18

\* Estimations effectuées dans GZ-GAMMA

\*\* Estimations effectuées dans TREE-PUZZLE

Tableau 78 : Valeurs du paramètre alpha pour les séquences 12S rRNA (393 bp)

L'inclusion du sub-fossile MED dans le groupe des Lepilemuridae entraîne une légère augmentation de la variabilité. Les valeurs obtenues à partir de l'analyse dans TREE-PUZZLE sont très différentes de celles obtenues avec GZ-GAMMA, mais les variations, dans les deux cas, vont dans le même sens. La plus importante différence entre les deux estimations se trouve pour les groupes EUL et HAP. Ceci est sans doute dû au fait que la distribution gamma présente une grande variance, et peut être aussi au fait que les séquences ne sont pas strictement identiques (du fait de l'élimination ou non des gaps) pour les deux méthodes.

Pour le groupe GRC, représentant l'ensemble des séquences, l'estimation de  $\alpha$  dans GZ-GAMMA est de 0,373432, et de  $0,31 \pm 0,03$  dans TREE-PUZZLE. Ceci traduit une hétérogénéité dans le taux de substitution par site. Celle-ci est très proche de la valeur qui peut être estimée à partir des séquences de cytochrome b. Lorsque les taux de substitution par site sont étudiés au log de la séquence (**Figure 39**) on constate bien qu'il existe une variabilité dans ces taux. Dans cette figure la courbe figurée en rouge représente la moyenne mobile avec une période de 10.

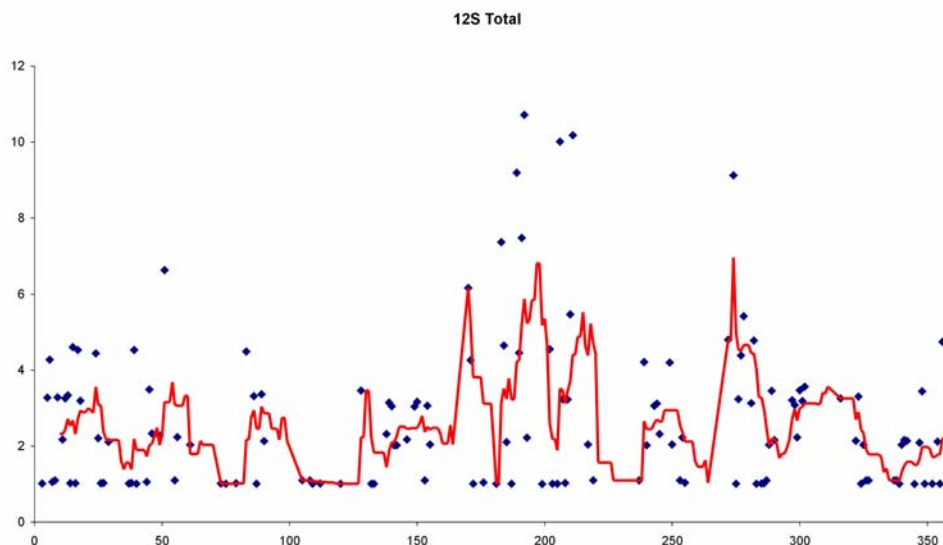


Figure 39 : Répartition des substitutions par site pour les séquences 12S rRNA (393 bp)



Groupe	Alpha 1 *	Alpha 2 **	Variance de la distribution gamma
EUL	0,068587	0,10 ± 0,05	9,55
HAP	0,064191	0,13 ± 0,08	7,78
IND	0,399521	0,32 ± 0,17	3,16
LEP	0,184190	0,57 ± 2,45	1,75
MGA	0,299264	0,55 ± 0,64	1,81
EHP	0,161236	0,34 ± 0,12	2,30
EHV	0,186832	0,31 ± 0,10	3,18
IMG	0,473505	0,58 ± 0,16	1,71
GRA	0,325995	0,48 ± 0,08	2,09
GRB	0,341132	0,44 ± 0,06	2,28
GRC	0,382829	0,47 ± 0,05	2,11

\* Estimations effectuées dans GZ-GAMMA

\*\* Estimations effectuées dans TREE-PUZZLE

Tableau 79 : Valeurs du paramètre alpha pour les séquences 12S rRNA (zones 'Loop')

Pour les séquences correspondant aux zones 'Loop' les variations observées dans les taux de substitution sont sensiblement identiques aux précédentes. La seule différence importante que l'on peut constater est le fait que l'inclusion de MED dans le groupe LEP entraîne ici une augmentation de plus de 60 % de  $\alpha$  (ce qui était déjà le cas avec le cytochrome b) pour l'estimation effectuée avec GZ-GAMMA. Par contre pour l'estimation à partir de TREE-PUZZLE les valeurs sont sensiblement identiques pour les groupes LEP et MGA. Pour les groupes EUL et HAP la variance de la distribution gamma reste élevée, même si elle est beaucoup plus faible que dans le cas des séquences 12S totales. Comme précédemment il existe des différences entre les deux types d'estimations qui sont généralement plus élevée avec TREE-PUZZLE qu'avec GZ-GAMMA (sauf pour le groupe IND). Avec TREE-PUZZLE ce sont les groupes LEP et MGA qui présentent les valeurs les plus élevées de  $\alpha$ . Pour l'ensemble des séquences (groupe GRC) le paramètre de forme à une valeur de 0,382829 avec GZ-GAMMA et de  $0,47 \pm 0,05$  avec TREE-PUZZLE. Il existe donc ici une hétérogénéité un peu plus faible du taux de substitution par site que dans le cas des séquences totales, ce qui se traduit dans le graphique de la **Figure 40**.

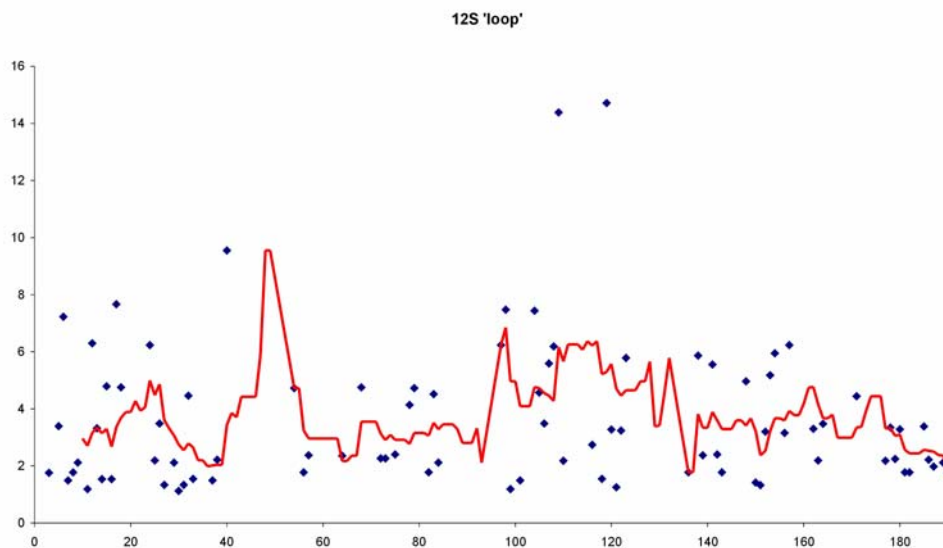


Figure 40 : Répartition des substitutions par site pour les séquences 12S rRNA (zones 'Loop')

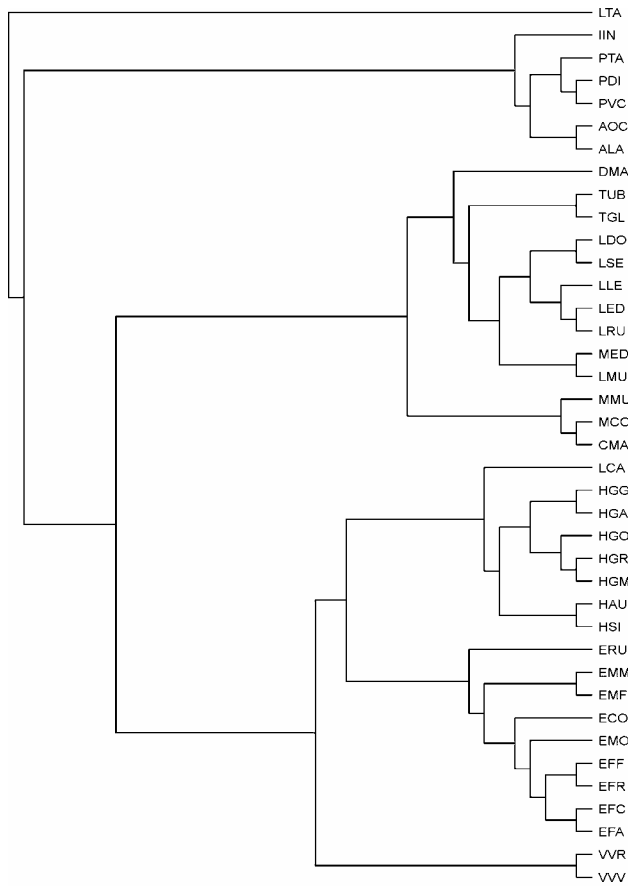
## Reconstructions phylogénétiques

Les reconstructions phylogénétiques utilisant les séquences 12S rRNA totales, ainsi que celles correspondant aux zones 'Loop', ont été effectuées à partir des quatre méthodes suivantes:

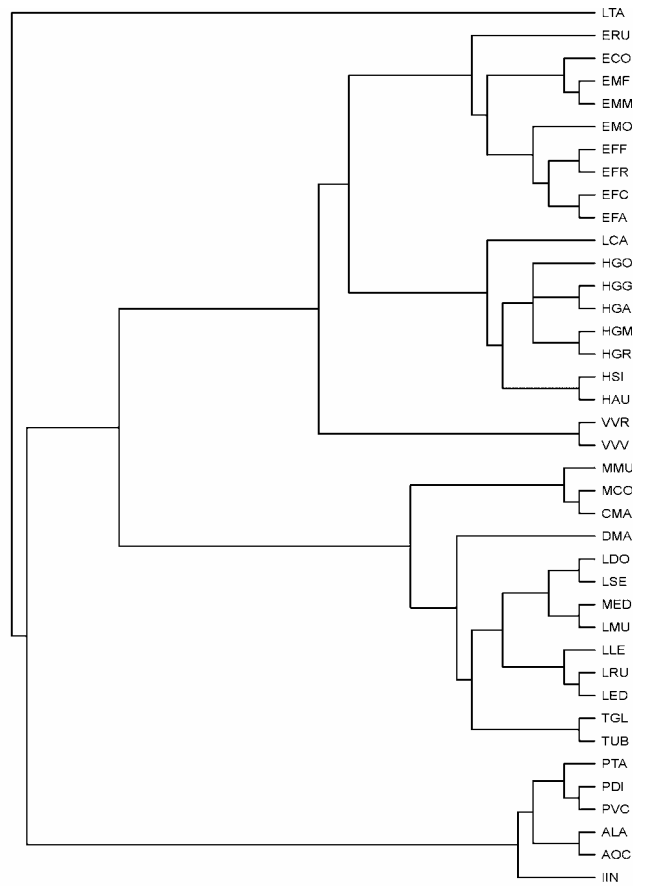
- **Maximum Likelihood (ML)**: les valeurs utilisées pour le rapport Transitions / Transversions sont de 2,1 pour les séquences totales et de 1,8 pour les zones 'Loop'. Chaque analyse est répétée 10 fois avec un ordre aléatoire d'introduction des séquences. L'arbre final est obtenu après réarrangements globaux.
- **Maximum Parsimony (MP)**: chaque analyse est effectuée 10 fois avec un ordre aléatoire d'introduction des séquences. Les arbres les plus parcimonieux sont obtenus après réarrangements globaux et l'arbre final est le consensus strict de ces arbres les plus parcimonieux. Les analyses utilisant uniquement les transversions pour les séquences totales ou les zones 'Loop'.
- **Parsimony Ratchet (PR)**: les analyses sont effectuées avec 500 répliques et le rééchantillonnage de 20 % des caractères. L'arbre final est le consensus strict des arbres les plus parcimonieux trouvés.
- **Neighbor-Joining**: les arbres sont obtenus avec un test de bootstrap de 5 000 répliques, en utilisant les distances de Tamura-Nei prenant en compte soit les transitions et les transversions (NJ), soit uniquement les transversions (NJV).

Les reconstructions effectuées à partir des séquences totales sont présentées à la **Figure 41**: (a) pour la reconstruction ML, (b) pour la reconstruction MP, (c) pour la reconstruction PR, (d) pour la reconstruction NJ et (e) pour la reconstruction NJV. La **Figure 42** rassemble les reconstructions effectuées à partir des zones 'Loop': (a) pour ML, (b) pour MP, (c) pour PR, (d) pour NJ et (e) pour NJV.

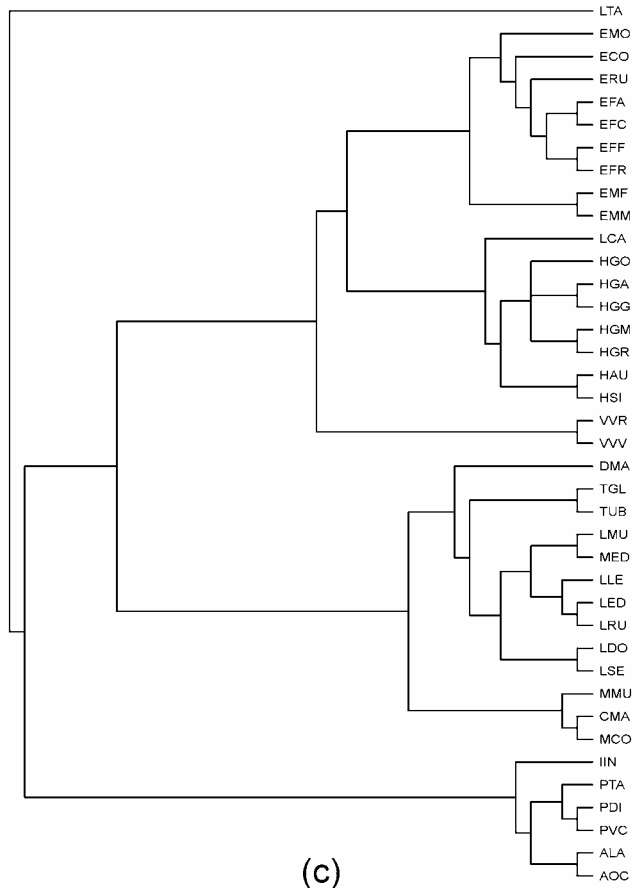
Dans le cas des séquences totales, l'arbre MP est le consensus de 6 arbres et celui obtenu à partir des zones 'Loop' est le consensus de 4 arbres. Pour la méthode PR l'arbre obtenu avec les séquences totales est le consensus de 3 arbres et celui obtenu avec les zones 'Loop' est le consensus de 31 arbres.



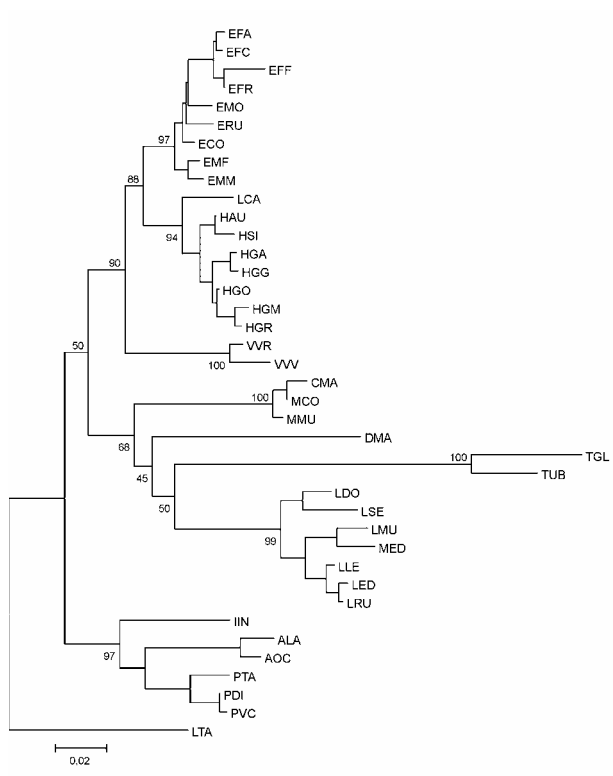
(a)



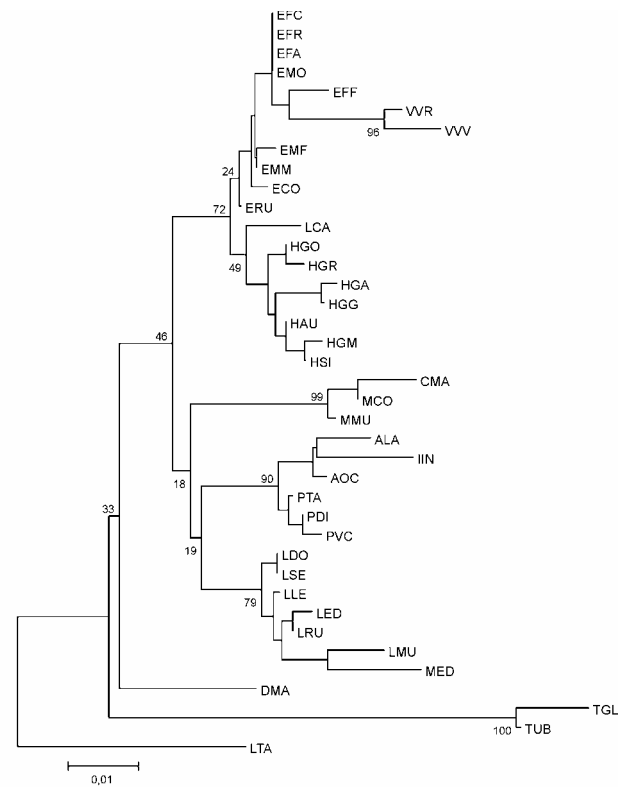
(b)



(c)

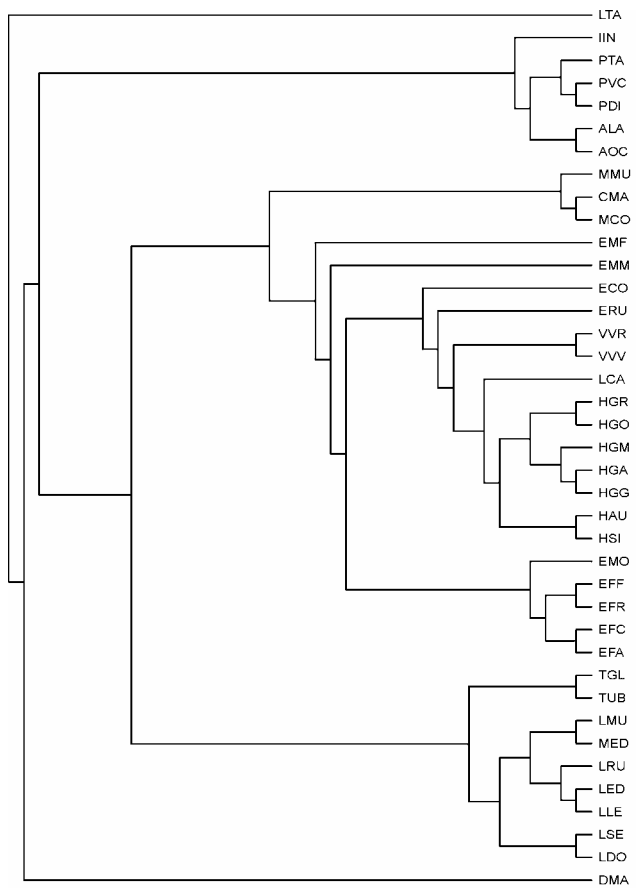


(d)

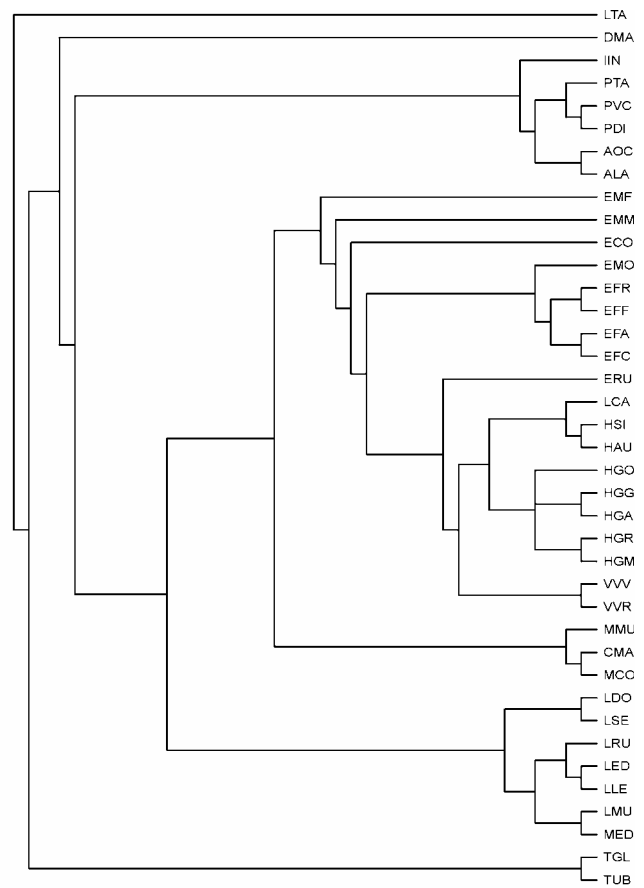


(e)

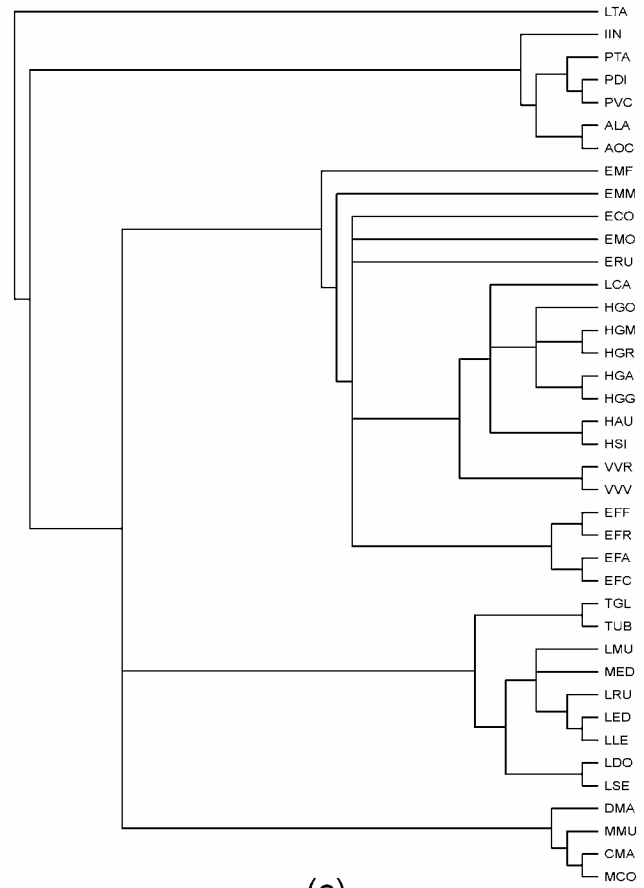
**Figure 41 : Reconstructions phylogénétiques pour les séquences 12S rRNA (393 bp)**



(a)



(b)



(c)

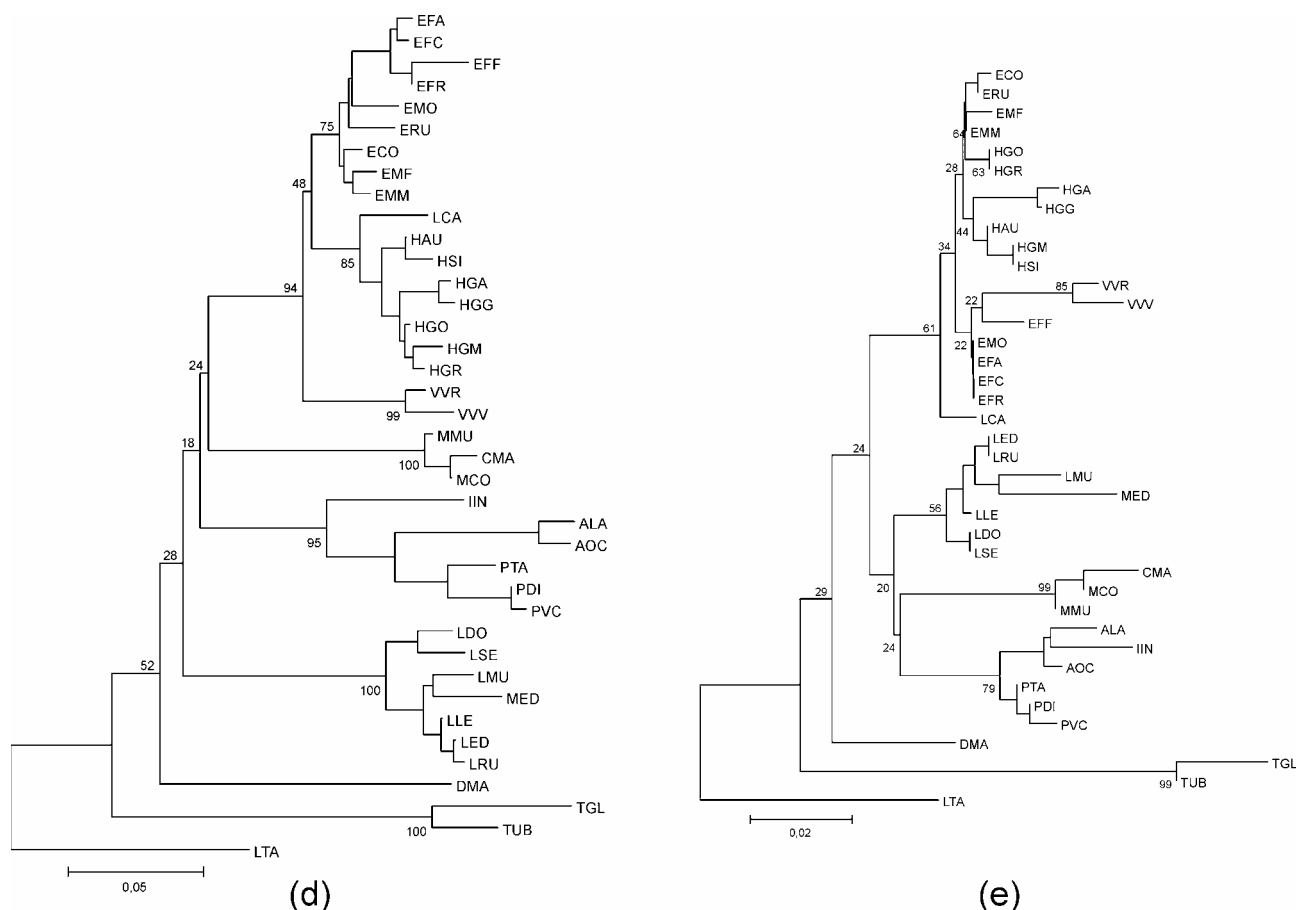


Figure 42 : Reconstructions phylogénétiques pour les séquences 12S rRNA (zones 'Loop')

Les différentes topologies ont été comparées dans le programme BASEML en considérant ou non l'existence d'une horloge moléculaire globale. Les résultats de ces tests sont résumés dans les **Tableaux 80** (séquences totales) et **81** (zones 'Loop').

Topologie	Sans horloge			Avec horloge		
	<i>LnL</i>	<i>pSH</i>	<i>pRELL</i>	<i>LnL</i>	<i>pSH</i>	<i>pRELL</i>
ML	- 2681,502	- 1,000	0,496	- 2742,884	0,794	0,195
MP	- 2683,285	0,726	0,026	- 2742,884	0,794	0,240
PR	- 2685,868	0,548	0,093	- 2745,558	0,625	0,161
NJ	- 2682,075	0,850	0,384	- 2742,611	- 1,000	0,001
NJV	- 2746,001	0,005	0,002	- 2824,232	0,002	0,001

Tableau 80 : Comparaison des topologies obtenues à partir des séquences 12S rRNA (393 bp)

Topologie	Sans horloge			Avec horloge		
	<i>LnL</i>	<i>pSH</i>	<i>pRELL</i>	<i>LnL</i>	<i>pSH</i>	<i>pRELL</i>
ML	- 1651,694	0,730	0,230	- 1686,019	0,362	0,034
MP	- 1655,735	0,508	0,096	- 1699,505	0,130	0,003
PR	- 1653,742	0,605	0,100	- 1686,940	0,338	0,046
NJ	- 1649,028	- 1,000	0,565	- 1673,876	- 1,000	0,917
NJV	- 1694,981	0,020	0,009	- 1739,051	0,004	0,001

Tableau 81 : Comparaison des topologies obtenues à partir des séquences 12S rRNA (zones 'Loop')

Dans le cas des séquences 12S totales, sans l'existence d'une horloge moléculaire globale, la topologie ML apparaît la plus crédible et on trouve ensuite la topologie NJ. Par contre si le test est effectué en considérant l'existence d'une horloge moléculaire globale, c'est la topologie NJ qui apparaît comme étant la plus crédible. Pour les zones 'Loop', avec ou sans horloge, la topologie la plus crédible est toujours la topologie NJ.

Dans le cas de la topologie ML utilisant les séquences totales (**Figure 41a**), on trouve un cluster formé des séquences DMA, CHE et MGA. Dans le cluster des séquences MGA, le sub-fossile MED est placé à la base des Lepilemuridae, associé à la séquence LMU. L'ensemble précédent est le cluster frère d'un autre formé des groupes EUL, HAP et VAR. Les séquences du groupe IND se trouvent à la base de l'arbre. La topologie MP (**Figure 41b**) utilisant les séquences totales, fournit la même disposition que dans le cas de la méthode ML. Ceci est également le cas de la topologie PR (**Figure 41c**). Dans le cas de la topologie NJ (**Figure 41d**) on trouve également la même disposition, mais on peut remarquer que les branches menant à DMA et au groupe TUP sont particulièrement longues. Dans cette topologie, l'ensemble EUL, HAP, VAR est supporté par une valeur de bootstrap de 68 % et celui formé de CHE, DMA, TUP et MGA est supporté par une valeur de bootstrap de 90 %. Par contre lorsque la reconstruction NJ ne prend pas en compte les transversions (**Figure 41e**), si on retrouve bien le cluster formé par les groupes EUL, HAP et VAR, la séquence DMA et celles du groupe TUP sont ici situées à la base de l'arbre. D'autre part les groupes MGA et IND forment deux clusters frères à la base desquels on trouve les séquences du groupe CHE, mais l'association MGA – IND est supportée par une faible valeur de bootstrap (19 %).

Si cependant les reconstructions sont effectuées en utilisant uniquement les zones 'Loop' des séquences 12S, la méthode ML (**Figure 42a**) fournit une topologie dans laquelle les clusters correspondant aux groupes EUL, HAP et VAR ne sont pas correctement individualisés. Les séquences du groupe TUP sont trouvées avec celles du groupe MGA. La séquence DMA est à la base de l'arbre et l'on trouve ensuite dans l'ordre le groupe IND, l'ensemble TUP – MGA, le groupe CHE et enfin l'ensemble EUL – HAP – VAR. Dans le cas de la reconstruction MP (**Figure 42b**) on retrouve à la base de l'arbre la séquence DMA et celles du groupe TUP, puis le groupe IND, le groupe MGA, le groupe CHE et enfin l'ensemble EUL – HAP – VAR. Dans ce dernier, ensemble les séquences des trois groupes apparaissent mieux réunies que dans le cas de la méthode ML. La méthode de Parsimony Ratchet (**Figure 42c**) montre l'existence de nombreux nœuds non résolus et c'est le groupe IND qui est en position basale de l'arbre. La reconstruction NJ (**Figure 42d**) montre que les séquences TUP et DMA sont situées à la base de l'arbre. Le cluster formé par les groupes EUL, HAP et VAR est supporté par une forte valeur de bootstrap (94 %). Le groupe CHE est situé à la base du cluster précédent, puis on trouve le groupe IND et enfin le groupe MGA. Dans ce dernier, le sub-fossile MED est associé à la séquence LMU, mais ces deux séquences ne se trouvent plus à la base des Lepilemuridae. Les valeurs de bootstrap pour la réunion des différents clusters sont faibles. Lorsque la méthode NJ prend en compte uniquement les transversions (**Figure 42e**) l'ensemble EUL – HAP – VAR présente un support de bootstrap inférieur à la valeur trouvée précédemment et il n'est plus possible d'individualiser les différents groupes.

En prenant en compte les résultats précédents concernant les séquences nucléotidiques et protéiques sur le cytochrome b, ce sont donc les topologies NJ qui sont retenues, tant pour les séquences 12S totales que pour celles regroupant uniquement les zones 'Loop'. Comme ces deux topologies ne sont pas identiques, les longueurs des différentes branches ont été estimées à partir des résultats fournis par l'analyse dans BASEML (**Tableau 82**) et la crédibilité de différents groupes de quatre clusters a été estimée dans PHYLTEST (**Tableau 83**).

Topologie	Arbre NJ		Arbre NJ	
Longueur totale de l'arbre	1,75009		2,27586	
	<i>Longueur de la branche</i>	<i>% *</i>	<i>Longueur de la branche</i>	<i>% *</i>
EUL – HAP	0,039335	2,25	0,033682	1,48
EHP – VAR	0,092225	5,27	0,044419	2,83
MGA – TUP	0,206358	11,79	-	-
MGA + TUP – DMA	0,167008	9,54	-	-
MGA + TUP + DMA – CHE	0,148637	8,49	-	-
MGA + TUP + DMA + CHE – EHV	0,107738	6,16	-	-
MGA + TUP + DMA + CHE + EHV – IND	0,100717	5,75	-	-
CHE- EHV	-	-	0,214495	9,42
CHE + EHV – IND	-	-	0,111221	4,89
CHE + EHV + IND – MGA	-	-	0,160568	7,05
CHE + EHV + IND + MGA – DMA	-	-	0,227515	9,99
CHE + EHV + IND + MGA + DMA – TUP	-	-	0,350566	15,40

\* Pourcentage calculé par rapport à la longueur totale de l'arbre

Tableau 82 : Longueurs des branches pour les topologies NJ (12S rRNA)

Topologie	Séquences	Branche interne	CP	Probabilité
(DMA , TUP) , (MGA + IND + CHE , HAP + EUL + VAR)	Totales	0,00961604	0,729	NS
	Loop	0,00509931	0,391	NS
(DAM, HAP + EUL + VAR) , (TUP , MGA + IND + CHE)	Totales	0,00804341	0,715	NS
	Loop	- 0,0064756	-	-
(MGA , VHE) , (IND , EUL + HAP + VAR)	Totales	0,00626373	0,605	NS
	Loop	0,00186081	0,289	NS
(MGA , IND) , (CHE , EUL + HAP + VAR)	Totales	0,0024957	0,497	NS
	Loop	0,00322814	0,236	NS
(EUL + VAR , HAP) , (DMA + TUP , CHE + IND + MGA)	Totales	0,0519637	1,000	< 0,1 %
	Loop	0,0529434	0,999	< 0,1 %
(EUL , HAP) , (VAR , CHE + DMA + TUP + MGA + IND)	Totales	0,012589	0,451	NS
	Loop	0,00480352	0,411	NS
(CHE , DMA) , (TUP , MGA)	Totales	0,00784997	0,451	NS
	Loop	0,0133847	0,484	NS
(DMA , IND) , (TUP , MGA)	Totales	0,0115966	0,621	NS
	Loop	0,000112095	0,000	NS
(DMA , TUP) , (MGA , IND)	Totales	0,00110206	0,103	NS
	Loop	0,00451775	0,182	NS

Tableau 83 : Test 'quatre clusters' pour les topologies NJ (12S rRNA)

L'analyse 'quatre clusters' montre que pour les deux types de séquences il n'existe qu'une seule topologie présentant une forte significabilité: celle dans laquelle un cluster est formé de EUL et HAP, celui-ci étant associé aux séquences du groupe VAR et d'autre part un cluster regroupant DMA et TUP, associé à un autre regroupant CHE, IND et MGA. Aucune des topologies reconstruites ne présente cette organisation précise. Dans les topologies NJ la plus proche est celle obtenue à partir des séquences totales. Cette analyse montre d'autre part que la topologie (DMA, TUP), (MGA, IND) est la moins probable, ce qui est en contradiction avec les résultats obtenus à partir des séquences de cytochrome b. Pour les groupes EUL, HAP et VAR l'analyse des longueurs des branches montre que celles-ci représentent un plus faible pourcentage de la longueur totale de



l'arbre dans le cas des zones 'Loop' que dans le cas des séquences totales. D'autre part, dans la topologie NJ avec les zones 'Loop', la branche séparant le groupe TUP des autres séquences représente un pourcentage sensiblement double de celui trouvé pour la branche séparant DMA des autres séquences (sauf celles du groupe TUP).

La topologie qu'il est possible de retenir à partir de l'analyse des séquences 12S rRNA peut donc être représentée de la manière suivante:

**((( EUL, HAP), VAR), ((((MGA, TUP), DMA), CHE)), IND)**

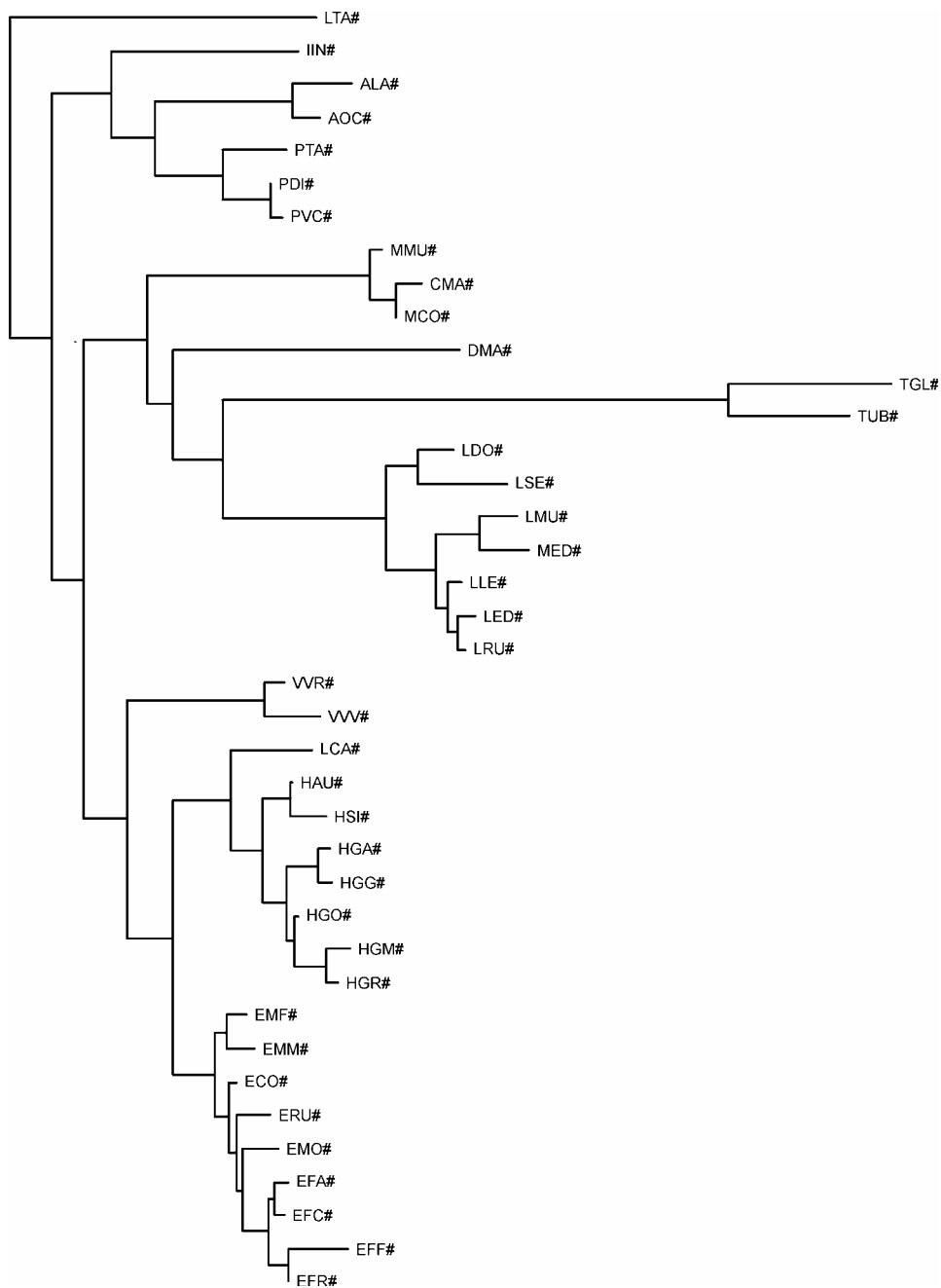
### Horloge moléculaire et arbres linéarisés

Dans un premier temps les séquences 12S totales sont utilisées pour la reconstruction d'un arbre par la méthode NJ en utilisant les distances de Tamura-Nei (**Figure 43**). Les valeurs de bootstrap, ainsi que les résultats du test 'deux clusters' effectué dans LINETREE sur cette topologie sont reportés dans le **Tableau 84**.

Cluster A	Cluster B	Bootstrap (%)	b <sub>A</sub>	b <sub>B</sub>	Z	CP
EUL	HAP	88	0,023188	0,033600	1,369947	83,24
VAR	EHP	90	0,037419	0,042306	0,516630	39,00
MGA	TUP	50	0,059997	0,146097	4,016263	99,96
MGA + TUP	DMA	45	0,091607	0,064337	1,690257	90,90
CHE	MGA + TUP + DMA	58	0,061993	0,098521	2,208738	97,22
CHE + MGA + TUP + DMA	EHV	50	0,099534	0,049918	3,697832	99,96
IND	EHV + CHE + DMA + TUP + MGA	-	0,056513	0,077596	1,653394	50,10

**Tableau 84 : Test 'deux clusters' sur la topologie NJ obtenue à partir des séquences 12S rRNA (393 bp)**

L'étude détaillée du test 'deux clusters' ne montre pas de séquences pour lesquelles l'évolution est significativement plus lente que la moyenne de celle calculée sur l'ensemble des séquences.



**Figure 43 : Topologie NJ obtenue dans LINETREE à partir des séquences 12S rRNA (393 bp)**

L'arbre linéarisé obtenu à partir de cette topologie est celui de la **Figure 44**. Si les divergences pour les différents groupes pris individuellement peuvent être définies, seules celles impliquant le regroupement de VAR, EUL et HAP sont calculables. Les séparations entre le groupe IND et les autres séquences, de même que CHE – DMA – TUP – MGA sont confondues dans une multifurcation.

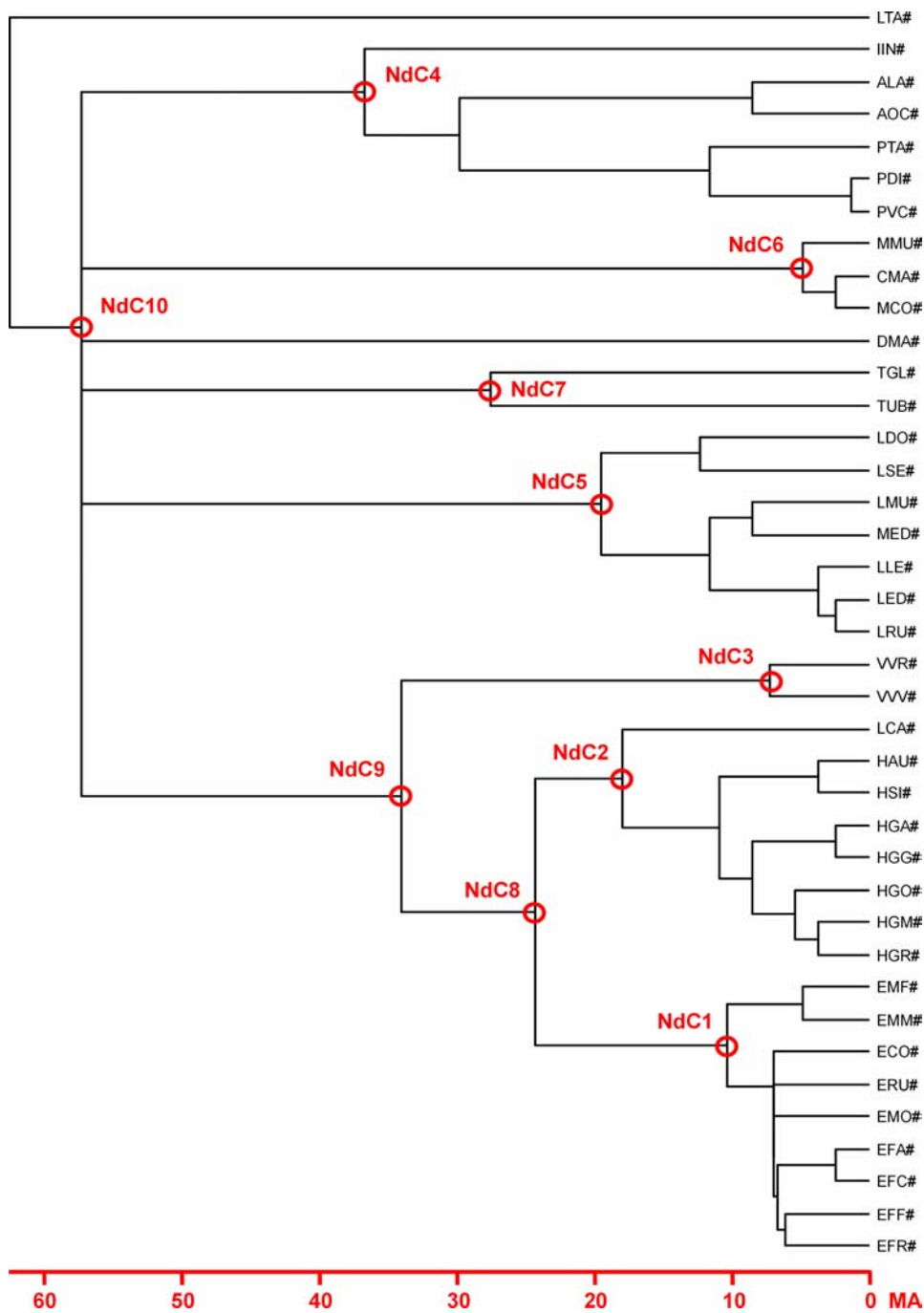


Figure 44 : Arbre linéarisé obtenu à partir des séquences 12S rRNA (393 bp)

Les estimations de dates de divergence à partir de cet arbre linéarisé sont reportées au **Tableau 85** (arbre linéarisé 1). D'autre part le test de la constance du taux d'évolution fournit une valeur de 60,898568, soit avec 36 ddl une significabilité inférieure à 0,1 % (valeur critique de 58,6192) faisant que l'hypothèse de la constance du taux ne peut être rejetée.

Nœud	Divergence	Arbre linéarisé 1	Arbre linéarisé 2	Arbre linéarisé 3	Arbre linéarisé 4	Moyenne ± SD
NdC1	EUL	11,13	-	7,66	-	9,39 ± 1,73
NdC2	HAP	19,47	-	11,85	-	15,66 ± 3,81
NdC3	VAR	7,75	-	6,20	-	6,97 ± 0,77
NdC4	IND	39,74	-	27,90	-	33,82 ± 5,92
NdC5	MGA	21,06	-	11,30	-	16,18 ± 4,88
NdC6	CHE	5,36	-	3,28	-	4,32 ± 1,04
NdC7	TUP	29,81	-	16,59	-	23,20 ± 6,61
NdC8	EUL – HAP	26,43	21,09	16,23	15,23	19,74 ± 4,45
NdC9	EHP – VAR	36,96	32,00	20,42	17,20	26,69 ± 8,05
NdC10	IND – CHE + EHV	-	-	40,48	37,89	39,18 ± 1,29
NdC11	LEP – MED	-	10,91	-	5,62	8,26 ± 2,64
NdC12	CHE – EHV	-	-	33,92	34,44	34,18 ± 0,26
NdC14	TUP – DMA + MGA + IND + CHE + EHV	-	-	59,08	58,92	59,00 ± 0,08

Tableau 85 : Estimations des dates de divergences à partir des séquences 12S rRNA (393 bp et zones 'Loop')

Afin de tenter de résoudre les multifurcations de l'arbre précédent, les séquences consensus des différents groupes sont utilisées. Cette étude permet d'obtenir l'arbre linéarisé de la **Figure 45**. Les multifurcations subsistent dans cet arbre et les quelques divergences qui peuvent être calculées sont résumées dans le **Tableau 85** (arbre linéarisé 2). On peut constater que les estimations pour les nœuds NdC8 et NdC9 sont plus faibles avec les séquences consensus qu'avec l'ensemble des séquences de chaque groupe. Le test sur la constance du taux d'évolution fournit ici une valeur de 34,492070, ce qui avec 7 ddl donne une significabilité inférieure à 0,1 % (valeur critique de 24,3219). Comme avec l'ensemble des séquences, l'hypothèse de la constance du taux d'évolution ne peut être rejetée.

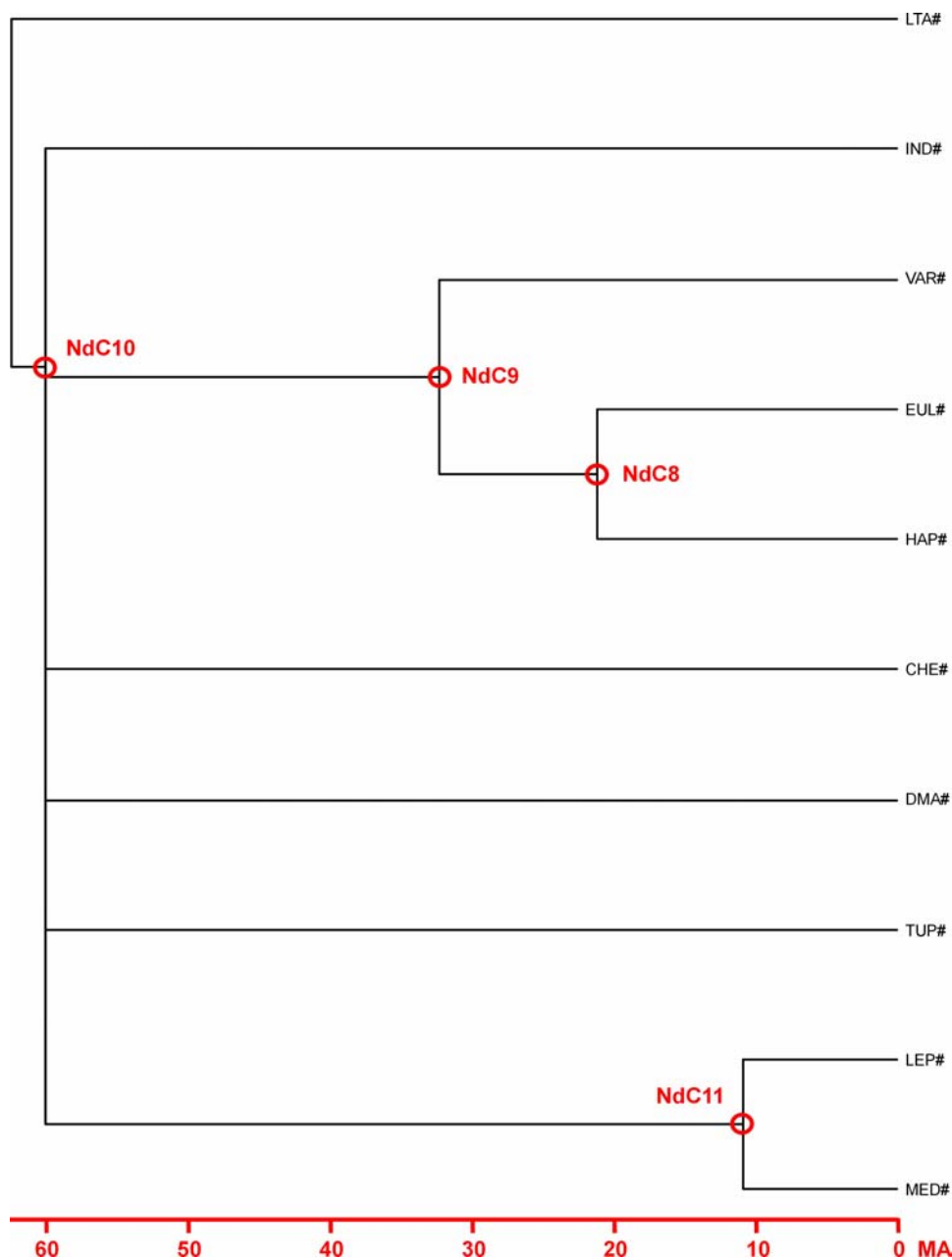
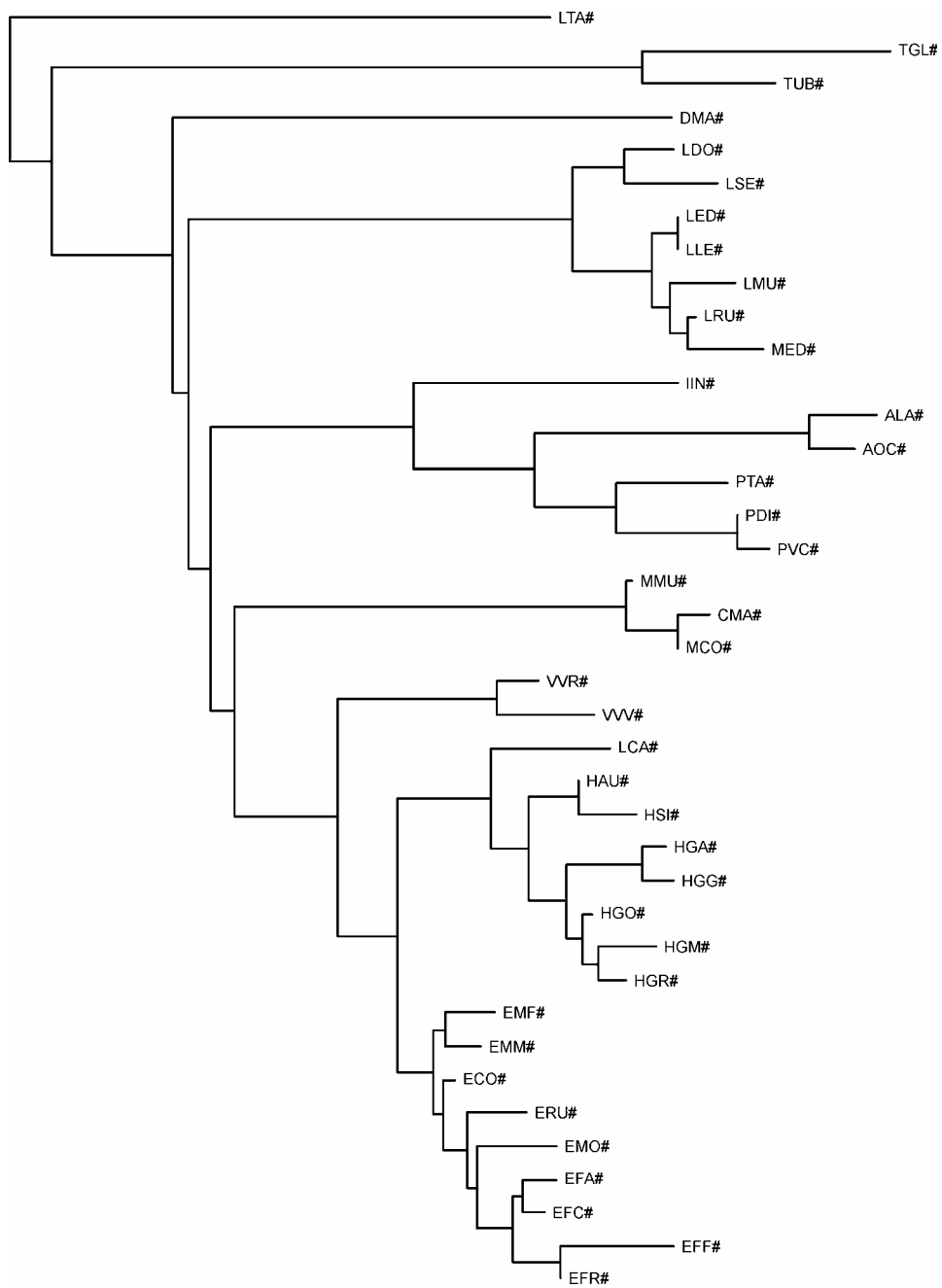


Figure 45 : Arbre linéarisé obtenu à partir des séquences consensus 12S rRNA (393 bp)

Une étude similaire a été effectuée en utilisant les séquences des zones 'Loop'. L'arbre NJ calculé dans LINETREE en utilisant les distances de Tamura-Nei est représenté à la **Figure 46** et l'arbre linéarisé qui en découle est celui de la **Figure 47**.



**Figure 46 : Topologie NJ obtenue dans LINETREE à partir des séquences 12S rRNA (zones 'Loop')**

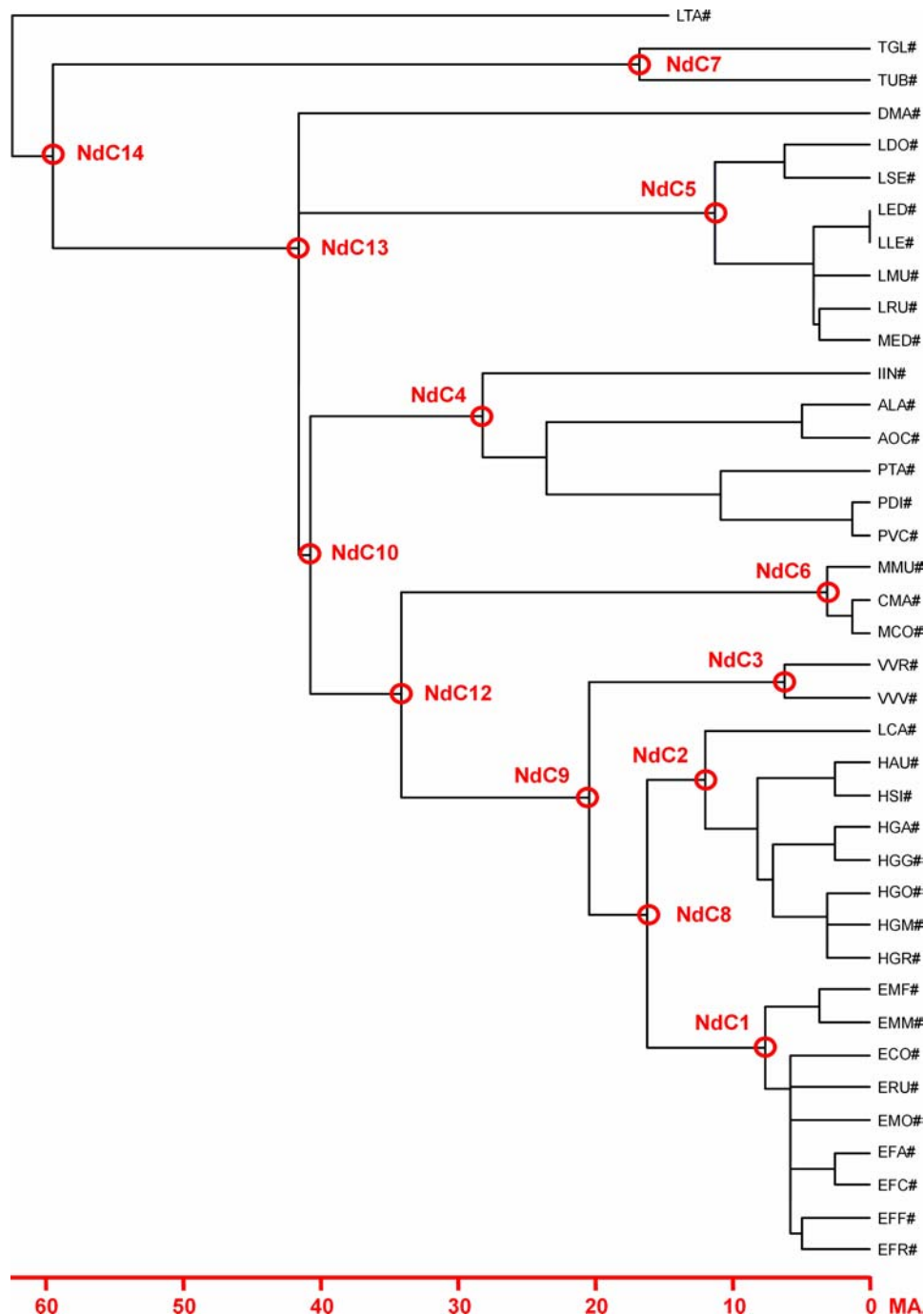
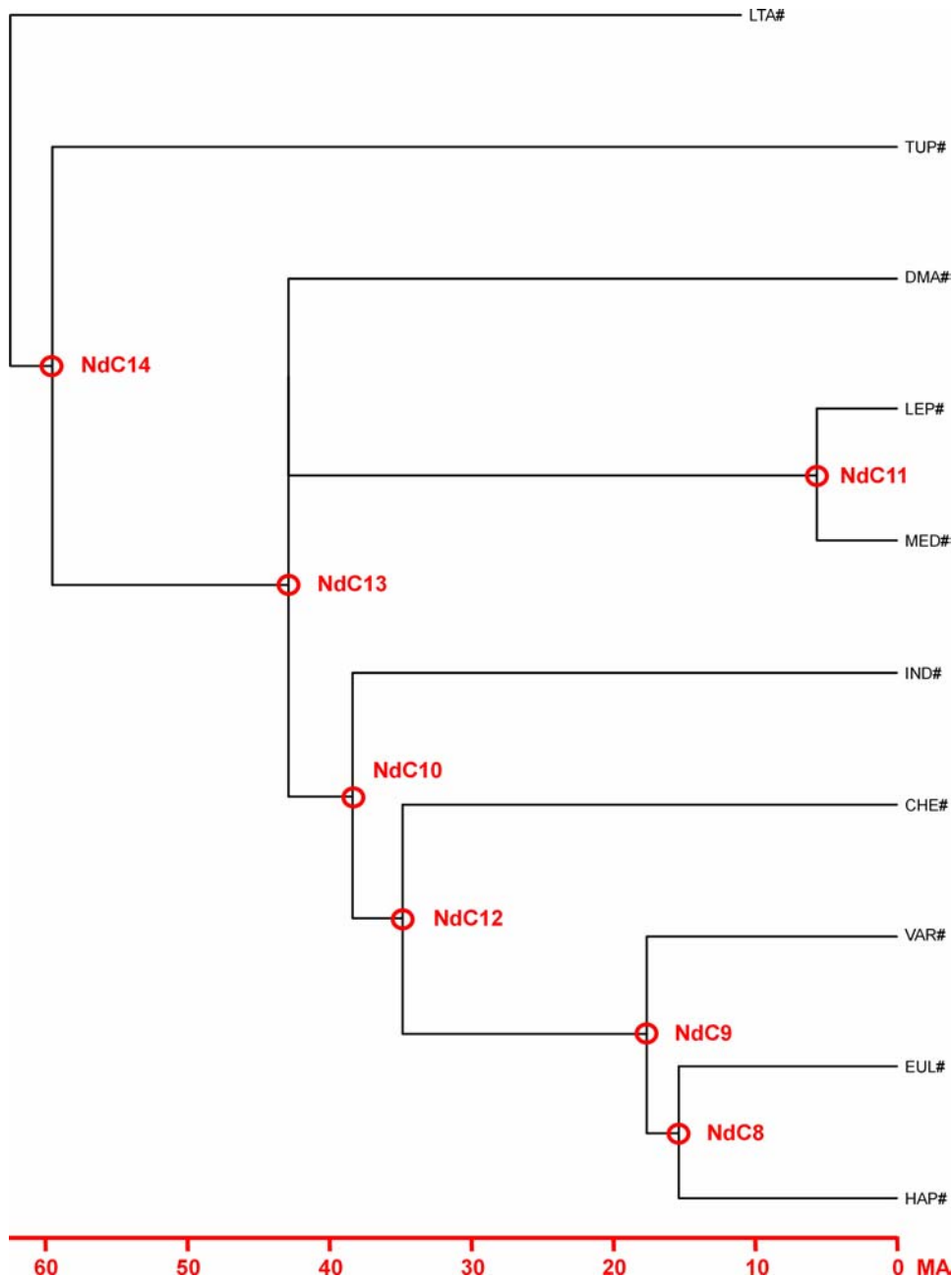


Figure 47 : Arbre linéarisé obtenu à partir des séquences 12S rRNA (zones 'Loop')

On peut constater que la plupart des nœuds sont ici résolus et les résultats des estimations des temps de divergence se trouvent dans le **Tableau 85** (arbre linéarisé 3). Toutes les estimations sont plus faibles que celles obtenues à partir des séquences totales. D'autre part le test du  $X^2$  sur la constance du taux d'évolution donne une valeur de 48,294017, soit une significabilité inférieure à 10 % avec 36 ddl (valeur critique de 47,2122) et ici, contrairement à ce qui avait été trouvé pour les séquences totales, il est possible d'envisager que le taux d'évolution ne soit pas constant au long de l'arbre.

L'utilisation des séquences consensus permet d'obtenir l'arbre linéarisé de la **Figure 48** pour lequel il existe toujours une multifurcation. Les dates de divergence calculées à partir de cet arbre

figurent dans le **Tableau 85** (arbre linéarisé 4). Le test de constance du taux d'évolution donne une valeur de 13,054187, soit une significabilité inférieure à 10 % (valeur critique de 12,0170 pour 7 ddl) ne permettant pas d'accepter la constance du taux d'évolution.



**Figure 48 : Arbre linéarisé obtenu à partir des séquences consensus 12S rRNA (zones 'Loop')**

Les temps de divergence ont également été calculés dans le logiciel BASEML en utilisant les topologies NJ pour les séquences totales et pour les zones 'Loop' avec le modèle de substitution de Tamura-Nei. Les résultats de cette analyse sont reportés dans le **Tableau 86**. Ces estimations sont toutes inférieures à celles obtenues à partir des arbres linéarisés. D'autre part on peut constater que toutes les dates de divergence fournies par l'étude des séquences 12S rRNA sont plus récentes que celles fournies par l'étude des séquences cytochrome b.



Nœud	Date moyenne de divergence ( $\pm$ SD)	
	Séquences totales	Zones 'Loop'
NdC1	4,09 $\pm$ 0,03	3,23 $\pm$ 0,87
NdC2	5,5 $\pm$ 0,12	5,07 $\pm$ 0,84
NdC3	2,83 $\pm$ 0,58	2,47 $\pm$ 0,41
NdC4	15,72 $\pm$ 1,01	16,21 $\pm$ 2,69
NdC5	8,07 $\pm$ 0,24	6,11 $\pm$ 1,01
NdC6	2,00 $\pm$ 0,05	1,86 $\pm$ 0,31
NdC7	6,75 $\pm$ 0,04	7,11 $\pm$ 0,31
NdC8	9,27 $\pm$ 0,01	8,24 $\pm$ 1,37
NdC9	16,52 $\pm$ 0,02	11,26 $\pm$ 1,87
NdC10	38,47 $\pm$ 0,01	31,41 $\pm$ 5,22
NdC12	-	26,23 $\pm$ 4,36
NdC14	-	37,67 $\pm$ 6,27

**Tableau 86 : Dates de divergences moyennes estimées à partir des séquences 12S rRNA**

Le logiciel PAML permet également l'estimation des taux de mutation pour l'ensemble des données et les valeurs trouvées sont de  $4,41 \cdot 10^{-3}$  pour les séquences 12S totales et  $5,45 \cdot 10^{-3}$  pour les zones 'Loop'.

Comme il a été possible de constater l'existence d'une hétérogénéité des taux d'évolution entre les différents groupes, le logiciel BASEML a été utilisé pour évaluer l'existence d'horloges moléculaires locales. Les résultats de cette étude, pour les deux types de séquences, sont regroupés dans le **Tableau 87**.

Branches du groupe	Taux de substitution	
	Séquences totales	Zones 'Loop'
EUL	0,15813	0,64878
HAP	0,34841	4,22703
VAR	0,55846	2,42633
IND	0,39494	3,25076
MGA	2,30193	1,30583
CHE	0,19003	0,53546
TUP	7,00428	1,22023
DMA	1,12064	0,61982

**Tableau 87 : Taux de substitutions pour les branches principales des topologies 12S rRNA**

L'utilisation des zones 'Loop' fournit des valeurs supérieures à celles obtenues avec les séquences totales, sauf pour les groupes DMA, TUP et MGA, avec une différence très marquée pour le groupe TUP (7,00428 avec les séquences totales et 1,22023 avec les zones 'Loop'). Ces constatations tendent à confirmer le fait que les zones 'Loop' sont les plus sujettes à des modifications. Pour ce qui est des trois groupes discordants il faut admettre que toutes les zones de la séquence 12S étudiée ici sont affectées par des mutations, ce qui pourrait indiquer que ces groupes sont encore dans une phase d'évolution rapide.

## Traitements sur les données mixtes

Les résultats obtenus à partir des séquences de cytochrome b et de celles de 12S rRNA ne fournissant pas des résultats entièrement concordants, une étude est réalisée en utilisant les zones 'Loop' des séquences 12S combinées avec les séquences de cytochrome b totales. Les séquences mixtes ainsi formées ont une taille de 570 bp.

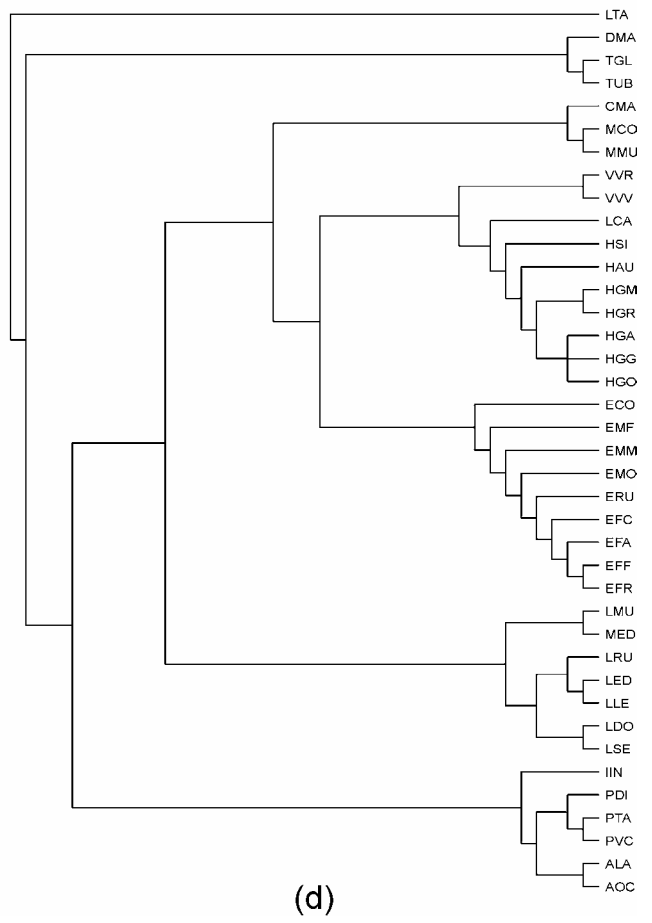
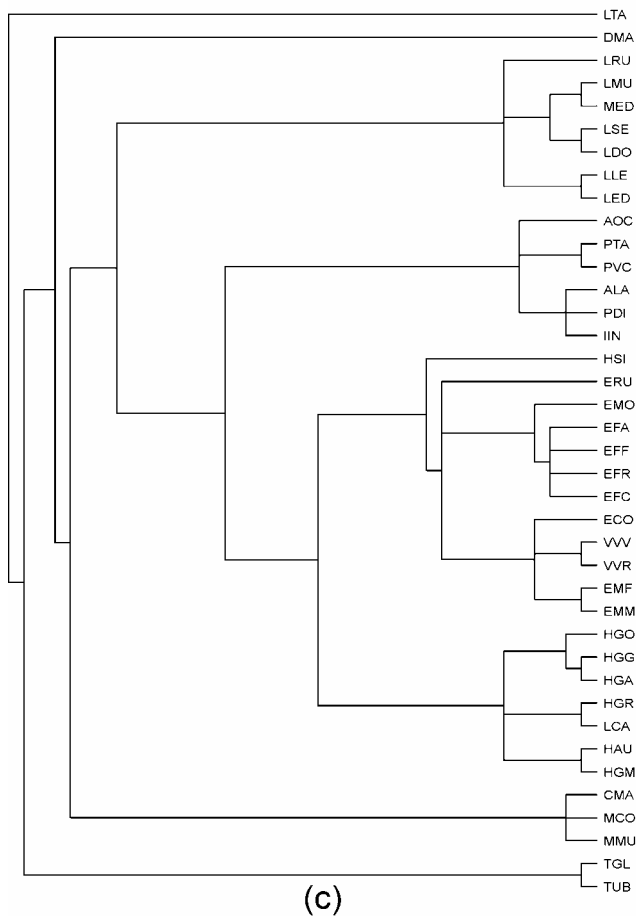
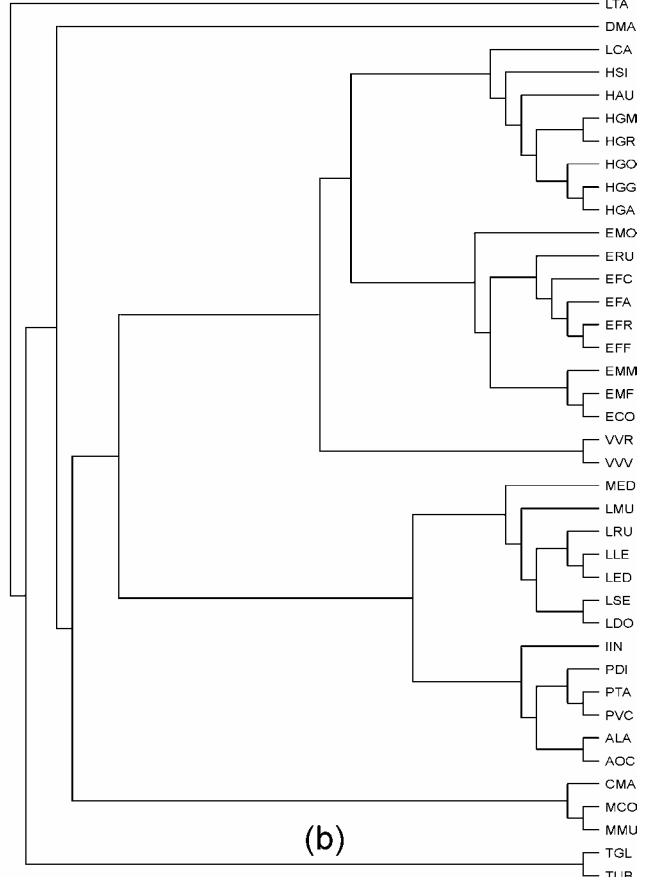
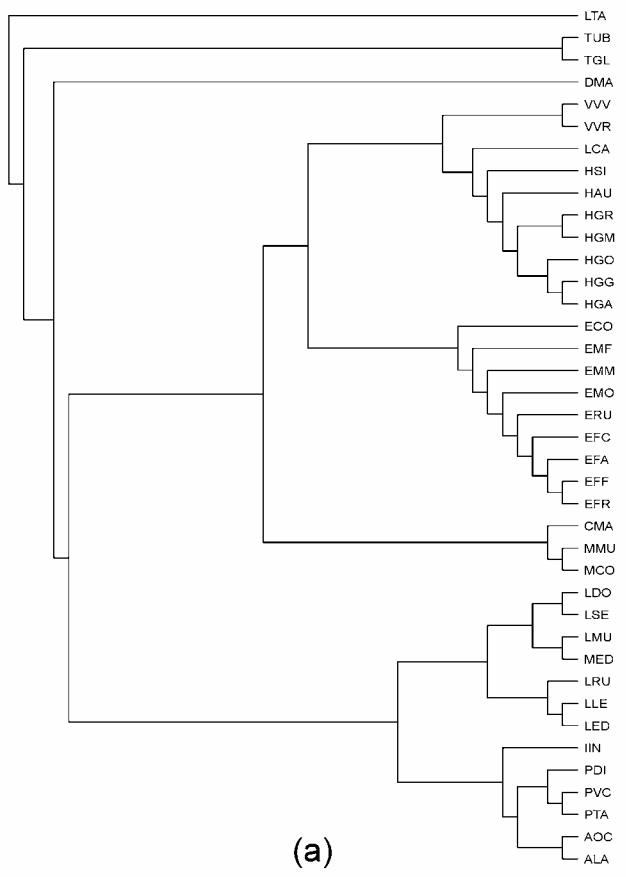
### Reconstructions phylogénétiques

Les méthodes utilisées pour ces reconstructions sont les suivantes:

- **Maximum Likelihood (ML)**: la valeur utilisée pour le rapport Transitions / Transversions est de 2,0. L'arbre final est obtenu après réarrangements globaux, l'analyse étant répétée 10 fois avec un ordre aléatoire d'introduction des séquences.
- **Maximum Parsimony**: une reconstruction est effectuée en utilisant les transitions et les transversions (MP) et une autre en utilisant uniquement les transversions (MPV). Les arbres les plus parcimonieux sont obtenus après réarrangements globaux, l'analyse étant répétée 10 fois avec un ordre aléatoire d'introduction des séquences. L'arbre final est obtenu par consensus strict des arbres les plus parcimonieux.
- **Parsimony Ratchet (PR)**: 500 répliquions sont effectuées avec le rééchantillonnage de 20 % des caractères et un arbre est conservé à chaque itération. L'arbre final est le consensus strict des arbres les plus parcimonieux trouvés.
- **Neighbor-Joining**: l'analyse utilise un test de bootstrap de 5 000 répliquions avec les distances de Tamura-Nei prenant en compte les transitions et les transversions (NJ), ou uniquement les transversions (NJV).

Ces reconstructions sont regroupées dans la **Figure 49**: (a) pour la méthode ML, (b) pour la méthode MP, (c) pour MPV, (d) pour PR, (e) pour NJ et (f) pour NJV.

L'arbre MP est le consensus de 4 arbres, les arbres MPV et PR sont le consensus de deux arbres.



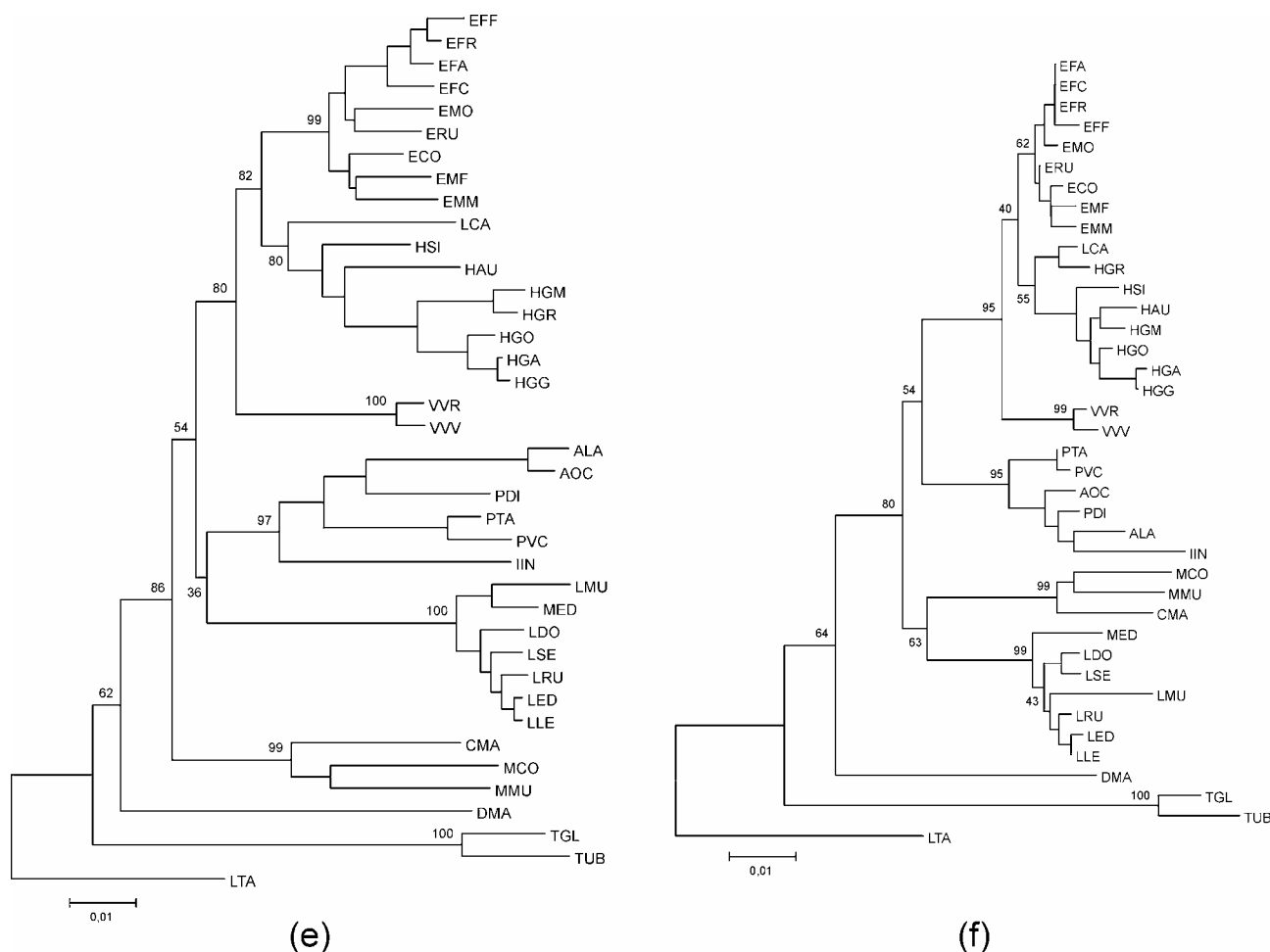


Figure 49 : Reconstructions phylogénétiques à partir des données mixtes (cytochrome b et 12S rRNA zones 'Loop')

Les résultats de la comparaison de ces différentes topologies dans BASEML sont résumés dans le **Tableau 88**.

Topologie	Sans horloge			Avec horloge		
	<i>LnL</i>	<i>pSH</i>	<i>pRELL</i>	<i>LnL</i>	<i>pSH</i>	<i>pRELL</i>
ML	- 5405,325	- 1,000	0,933	- 5451,238	0,826	0,244
MP	- 5415,519	0,642	0,047	- 5447,189	- 1,000	0,742
MPV	- 5559,611	0,000	0,000	- 5663,373	0,000	0,000
PR	- 5424,086	0,408	0,018	- 5473,392	0,301	0,005
NJ	- 5436,547	0,226	0,001	- 5469,984	0,352	0,008
NJV	- 5558,665	0,000	0,000	- 5615,935	0,000	0,000

Tableau 88 : Comparaison des topologies obtenues à partir des données mixtes

Sans horloge moléculaire la topologie la plus crédible est celle fournie par la méthode ML, alors qu'en prenant en compte l'existence d'une horloge moléculaire globale c'est la topologie MP qui est la plus crédible. On peut remarquer que dans les deux types de tests les topologies MPV et NJV apparaissent toujours les moins crédibles et que les topologies MPV, PR, NJ et NJV ont, avec le test pRELL, de très faibles significabilités.

La topologie ML (**Figure 49a**) montre un cluster formé des séquences VAR et HAP, associé au cluster des séquences EUL. A la base de ces deux clusters on trouve les séquences du groupe

CHE. L'ensemble HAP – VAR – EUL – CHE forme un cluster frère d'un autre regroupant les séquences des groupes MGA et IND, chacun de ces groupes étant bien individualisé. A la base de l'arbre on trouve la séquence DMA et celles du groupe TUP. Pour le groupe MGA le sub-fossile MED est associé à la séquence LMU, mais cet ensemble ne se trouve pas à la base des Lepilemuridae.

Dans le cas de la topologie MP (**Figure 49b**) on retrouve la disposition qui avait été définie à partir des séquences de cytochrome b, à l'exception près que les séquences du groupe TUP et la séquence DMA ne sont pas associées dans un cluster, bien qu'elles soient situées à la base de l'arbre. On observe donc un cluster formé de EUL – HAP et VAR, ce dernier groupe se trouvant dans la position la plus basale. Cet ensemble est le cluster frère d'un second formé des groupes MGA et IND. Le sub-fossile MED est ici en position basale de l'ensemble des séquences des Lepilemuridae. A la base des deux clusters précédents se trouve le groupe CHE, puis la séquence DMA et enfin le groupe TUP. DMA et TUP, comme dans le cas de la topologie ML ne sont pas réunis dans un cluster.

La topologie NJ (**Figure 49e**) présente une disposition identique à celle de la topologie MP. Dans celle-ci le groupement EUL – HAP est supporté par une valeur de bootstrap de 82 %, l'ensemble EHV par une valeur de 80 % et le groupement IND – MGA par une valeur de 36 %. D'autre part l'ensemble GRB est supporté par une forte valeur de bootstrap (86 %). La séquence MED est groupée avec la séquence LMU et les deux se trouvent à la base des autres séquences de Lepilemuridae.

L'analyse 'quatre clusters' menée dans PHYLTEST montre que les topologies (EHV, IND + MGA), (CHE, DMA + TUP), de même que (HAP + VAR, EUL), (IND + MGA, DTU) sont celles présentant les crédibilités les plus fortes. Par contre celles constituées de (EHV + CHE, DTU), (IND, MGA) et (MGA, CHE), (IND, DTU) sont les moins crédibles. La topologie (EHV, CHE), (IND + MGA, DTU), quant à elle, présente une crédibilité assez faible (avec une valeur de CP inférieure à 0,3).

C'est donc la topologie MP qui réunit les caractéristiques les meilleures, elle peut être représentée de la manière suivante:

**((((((EUL, HAP), VAR), (MGA, IND)), CHE), DMA), TUP)**

### **Horloge moléculaire et arbres linéarisés**

L'arbre NJ obtenu à partir des séquences mixtes, en utilisant les distances de Tamura-Nei est présenté à la **Figure 50**.

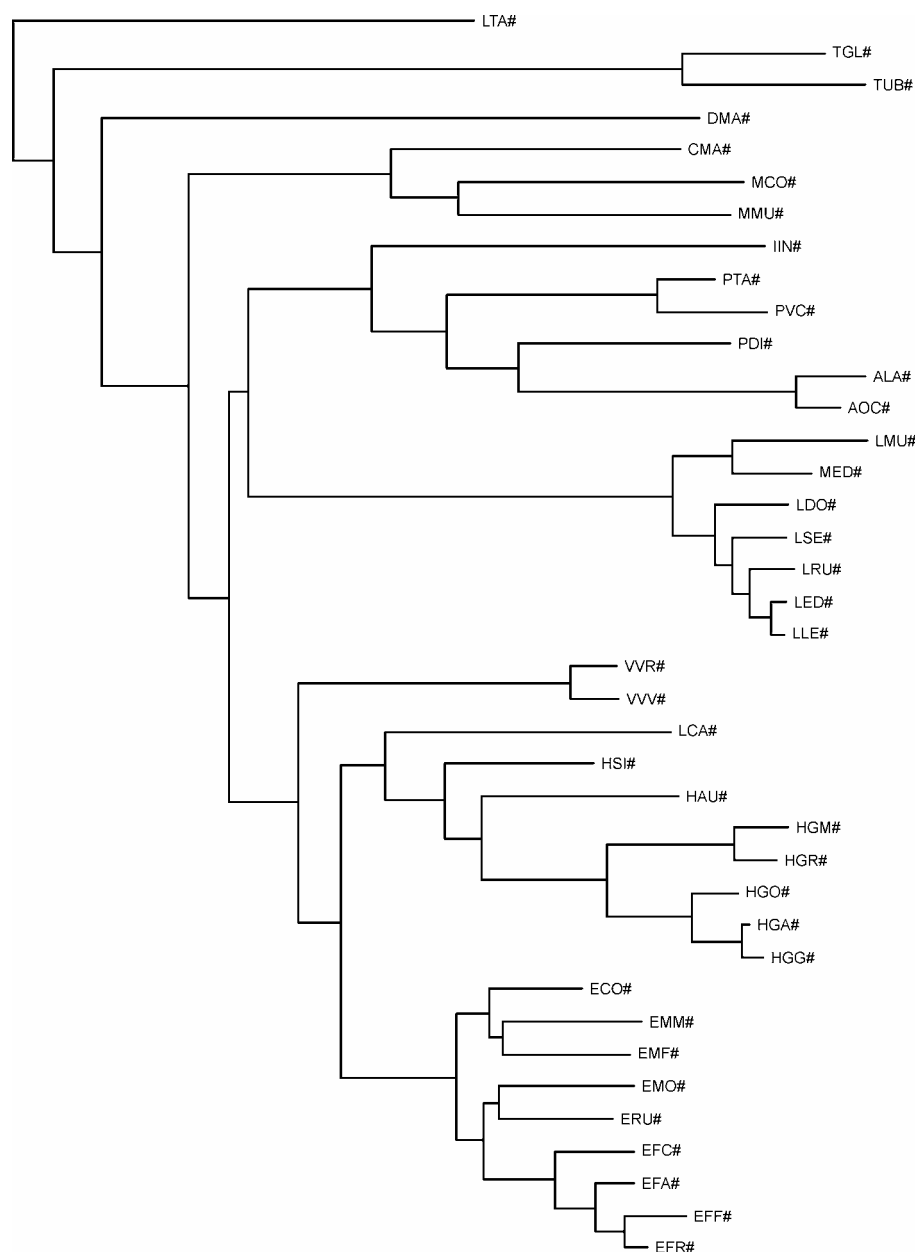


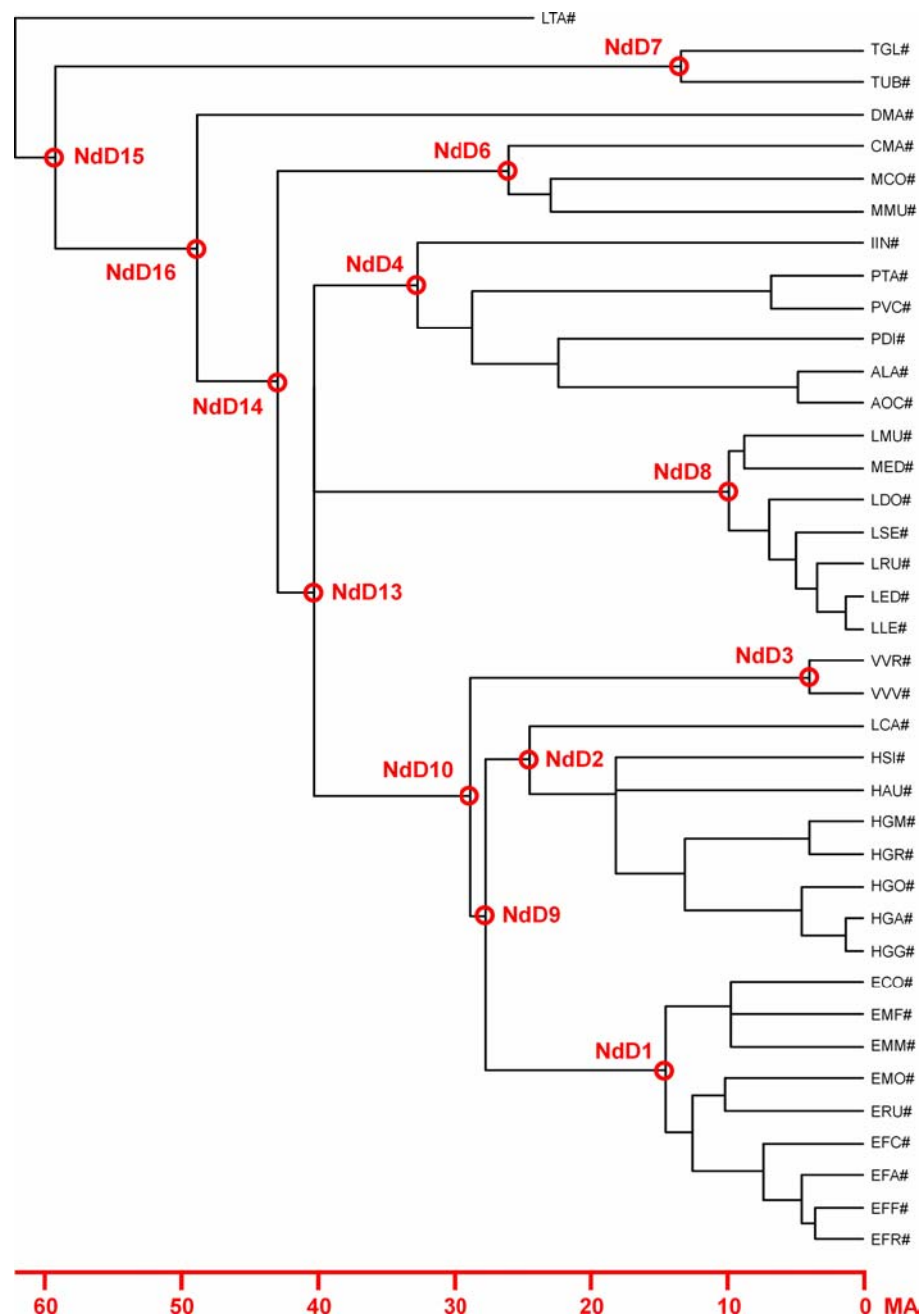
Figure 50 : Topologie NJ obtenue dans LINETREE à partir des données mixtes

Les résultats obtenus à partir de l'analyse 'deux clusters' effectuée dans LINETREE sont résumés, pour les clusters principaux, dans le **Tableau 89**. Dans ce test aucune des séquences utilisées ne semble avoir une évolution significativement plus lente que la moyenne calculée à partir de l'ensemble des séquences.

Cluster A	Cluster B	Bootstrap (%)	$b_A$	$b_B$	Z	CP
EUL	HAP	82	0,061524	0,070451	1,391653	83,54
VAR	EHP	80	0,073964	0,063353	1,565125	88,12
MGA	IND	36	0,110370	0,104809	0,429035	32,56
MGA + IND	EHV	54	0,107409	0,084718	2,493526	98,72
CHE	IMG + EHV	86	0,103693	0,101208	0,238740	18,20
DMA	CHE + IMG + EHV	62	0,114444	0,119049	0,298495	22,82
TUP	DMA + CHE + IMG + EHV	-	0,154884	0,127880	1,390488	83,54

Tableau 89 : Test 'deux cluster' pour la topologie NJ à partir des données mixtes

L'arbre linéarisé calculé à partir de ces données est représenté à la **Figure 51**.



**Figure 51 : Arbre linéarisé obtenu à partir des données mixtes**

Dans cet arbre il n'existe qu'une seule multifurcation faisant que les divergences entre IND, MGA et EHV ne peuvent être déterminées. Les estimations des dates de divergence effectuées à partir de cet arbre linéarisé sont indiquées dans le **Tableau 90** (arbre linéarisé 1). Le test du  $X^2$  pour la constance du taux d'évolution donne une valeur de 59,782885, soit une significabilité inférieure à 1 % (valeur critique de 58,6192 pour 36 ddl). On ne peut donc rejeter l'hypothèse de la constance du taux d'évolution au long de l'arbre étudié.

Nœud	Divergence	Arbre linéarisé 1	Arbre linéarisé 2
NdD1	EUL	14,44	-
NdD2	HAP	24,26	-
NdD3	VAR	4,04	-
NdD4	IND	32,54	-
NdD6	CHE	25,80	-
NdD7	TUP	13,28	-
NdD8	MGA	10,01	8,47
NdD9	EHP	27,53	-
NdD10	EHV	28,69	28,88
NdD13	IMG – EHV	40,05	37,74
NdD14	CHE – IMG + EHV	42,74	41,78
NdD15	TUP – DMA + CHE + IMG + EHV	58,92	59,11
NdD16	DMA – CHE + IMG + EHV	48,52	50,45
NdD17	EUL – VAR	-	23,49

**Tableau 90 : Estimations des dates de divergence pour les données mixtes**

L'utilisation des séquences consensus pour chaque groupe permet d'obtenir l'arbre linéarisé de la **Figure 52**. La multifurcation qui était présente avec l'ensemble des séquences de chaque groupe persiste ici. Les dates de divergence qui peuvent être calculées à partir de cet arbre apparaissent dans le **Tableau 90** (arbre linéarisé 2). Dans le cas du groupe MGA, la date obtenue est un peu plus faible que celle calculée précédemment. Pour les autres nœuds, les valeurs entre les deux estimations sont pratiquement identiques. Le test du  $X^2$  pratiqué à partir de cette topologie fournit une valeur de 21,71035, soit une significabilité inférieure à 1 % (valeur critique de 18,4753 avec 7 ddl) ne permettant pas, comme dans le cas de l'étude effectuée à partir de l'ensemble des séquences, de rejeter l'hypothèse d'une constance du taux d'évolution.



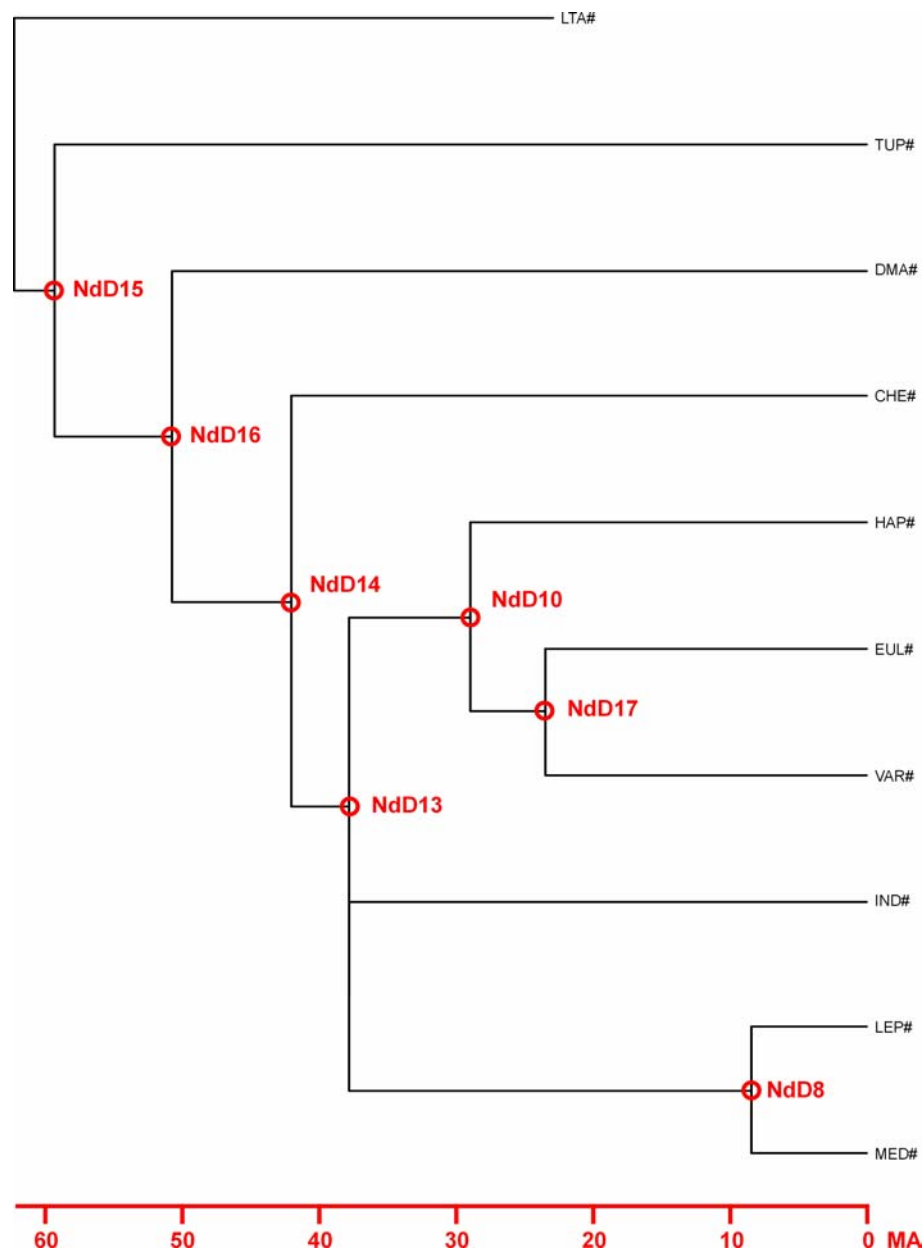


Figure 52 : Arbre linéarisé obtenu à partir des séquences consensus sur les données mixtes

Le logiciel BASEML a aussi été utilisé pour déterminer les dates de divergence (option 'TipDate') en utilisant les deux topologies les plus crédibles (ML et MP). Les résultats de cette analyse sont indiqués dans le **Tableau 91**. Les deux topologies fournissent des estimations très proches, mais qui sont toujours inférieures à celles obtenues par la méthode des arbres linéarisés. Les taux de mutation, estimés dans BASEML, sont de  $5,58 \cdot 10^{-3}$  à partir de la topologie ML et de  $5,54 \cdot 10^{-3}$  à partir de la topologie MP.

Nœud	Arbre ML	Arbre MP
NdD1	7,42 ± 1,22	7,54 ± 0,65
NdD2	12,60 ± 1,93	12,89 ± 1,12
NdD3	1,59 ± 0,54	1,59 ± 0,15
NdD4	19,00 ± 2,82	18,72 ± 1,62
NdD6	14,23 ± 2,31	14,27 ± 1,24
NdD7	6,52 ± 1,42	6,47 ± 0,57
NdD8	4,27 ± 0,76	5,05 ± 0,44
NdD10	16,87 ± 2,30	17,06 ± 1,47
NdD13	-	33,47 ± 2,89
NdD14	-	35,12 ± 3,03
NdD15	55,83 ± 6,65	55,43 ± 4,78
NdD16	44,26 ± 5,55	43,87 ± 3,78
NdD18	16,87 ± 2,30	-
NdD19	32,56 ± 4,20	31,02 ± 2,68

**Tableau 91 : Dates de divergence estimées à partir des topologies ML et MP avec les données mixtes**

Lorsque l'on teste l'existence d'horloges moléculaires locales en utilisant les topologies ML et MP on obtient, pour les branches menant aux différents groupes, des valeurs de taux de substitution indiqués dans le **Tableau 92**. Contrairement à ce qui avait été trouvé en utilisant, soit les séquences de cytochrome b, soit celles de 12S rRNA, le fait de combiner dans l'analyse les deux types de séquences montre une relative constance des taux à deux exceptions près. La première exception, trouvée avec la topologie ML, concerne le groupe MGA et la seconde, avec la topologie MP, concerne le groupe EUL. Dans les deux cas les valeurs du taux d'évolution sont très fortement augmentées.

Branches menant à	Arbre ML	Arbre MP
EUL	0,82258	48,83781
HAP	1,17337	2,29293
VAR	1,13156	2,30965
IND	1,26971	1,99509
MGA	18,76907	1,98534
CHE	1,28715	1,82668
TUP	1,20626	1,46540
DMA	1,44821	1,58743

**Tableau 92 : Taux de substitution par branche pour les topologies ML et MP avec les données mixtes**

L'ensemble des dates de divergences estimées à partir des séquences nucléotidiques et protéiques du cytochrome b, ainsi qu'à partir des séquences 12S rRNA sont résumées dans le **Tableau 93**. Elles permettent de reconstruire le schéma d'évolution de la **Figure 53**.

Divergence	Temps de divergence $\pm$ SD
EUL	10,39 $\pm$ 4,65
HAP	17,94 $\pm$ 9,59
VAR	4,02 $\pm$ 2,71
IND	25,78 $\pm$ 8,51
LEP	8,03 $\pm$ 2,27
CHE	17,56 $\pm$ 13,72
TUP	11,37 $\pm$ 6,71
MED – LEP	9,54 $\pm$ 3,73
EUL – HAP	23,49 $\pm$ 9,32
VAR – EHP	27,46 $\pm$ 9,93
IND – MGA	37,30 $\pm$ 4,87
DMA – TUP	55,39 $\pm$ 3,26
IMG – EHV	42,24 $\pm$ 3,47
CHE – GRA	46,62 $\pm$ 3,48
DTU – GRB	58,03 $\pm$ 2,03

Tableau 93 : Dates de divergence moyennes obtenues à partir de l'ensemble des données

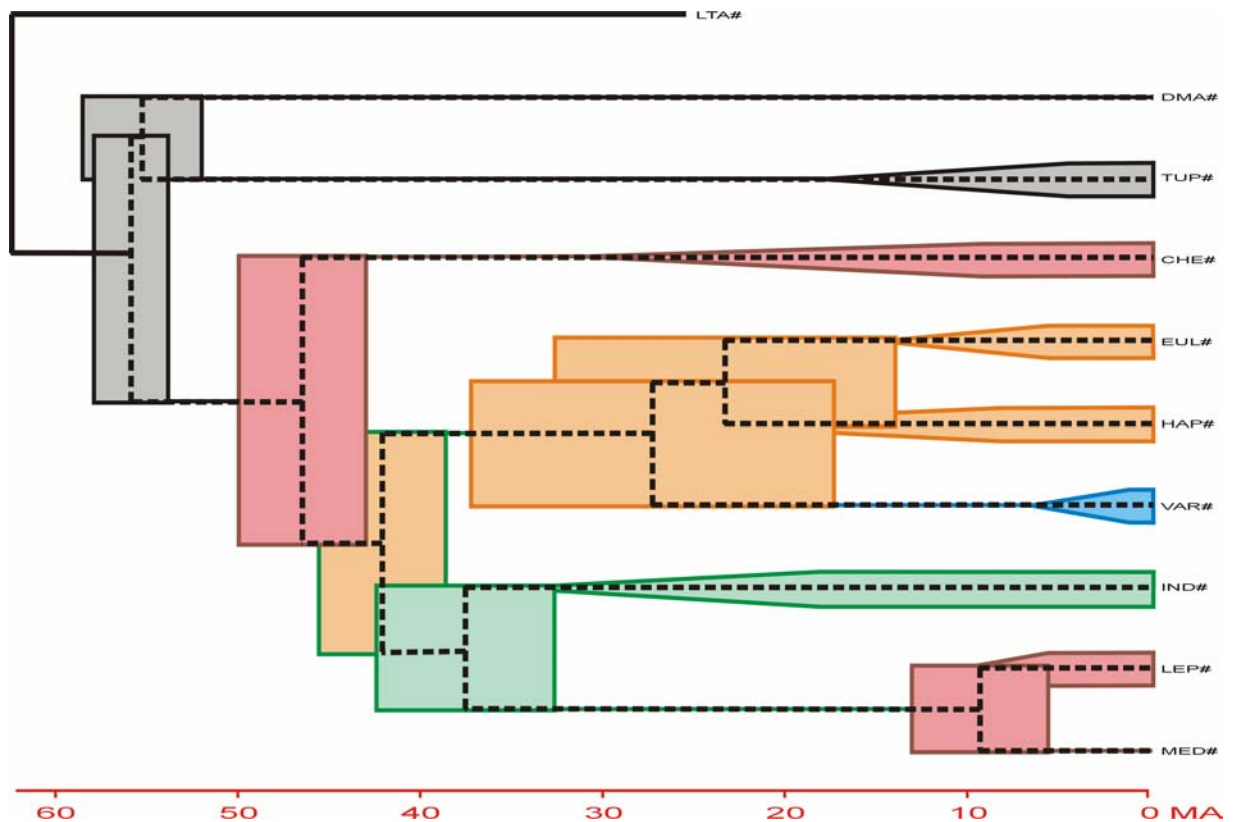


Figure 53 : Schéma d'évolution reconstruit à partir de l'ensemble des données

Si l'on compare ce schéma avec celui obtenu par l'utilisation des séquences de cytochrome b seules (Figure 28) on peut constater les points suivants:

- ✓ La divergence DTU – GRB (ensemble des Lemuriformes) est mieux individualisée avec l'ensemble des données qu'avec le cytochrome b seul. Cette divergence est relativement proche (58 MA environ) de celle fixée pour la séparation entre Lorisidae et Strepsirhini.
- ✓ Cette divergence DTU – GRB ne recoupe plus la divergence initiale du groupe des Lemuriformes (divergence CHE – GRA) qui se trouve fixée entre 46 et 47 MA.
- ✓ La divergence DMA – TUP (environ 55 MA) ne peut être différenciée de la divergence DTU – GRB, mais elle ne recoupe pas la divergence initiale du groupe des Lemuriformes.
- ✓ Les divergences pour les groupes EUL, HAP et VAR ne peuvent être différenciées.
- ✓ Les divergences IMG – HAP, IND – MGA et CHE – GRA se recoupent, mais seule la divergence IND – MGA recoupe celles de l'ensemble EUL, HAP et VAR.
- ✓ Les groupes HAP, IND et LEP apparaissent avoir divergé dès leur séparation d'avec VAR, MGA et MED, respectivement. Pour le groupe EUL la situation est moins nette, les domaines de divergence du groupe et de EUL – EHP ne se recoupant que très faiblement.

Une analyse de la variance a été effectuée sur l'ensemble des données utilisées pour établir le schéma d'évolution. Les résultats de cette analyse sont exploitables, le test de la normalité des résidus étant positif. Les moyennes pour les 15 nœuds étudiés ont été comparées en utilisant le test deux à deux de Tukey. Au seuil de 5 %, les dates de divergences statistiquement différentes permettent de supposer que:

- ✓ Les Tupaia ont débuté leur différenciation bien après qu'ils se soient séparés du Daubentonia.
- ✓ Le groupe des Eulemurs a débuté sa différenciation après un laps de temps de sensiblement 10 MA suivant la séparation de EUL et HAP. Par contre le groupe HAP aurait entamé sa différenciation dès sa séparation d'avec EUL.
- ✓ Le groupe des Varecia a débuté sa différenciation environ 20 MA après qu'il se soit séparé de l'ensemble Eulemurs – Hapalemurs.
- ✓ Les Cheirogaleidae ont entamé leur divergence environ 30 MA après qu'ils se soient séparés des Lemuriformes (GRA).
- ✓ Les Lemuriformes et Cheirogaleidae (GRB) ont divergé sensiblement 10 MA après qu'ils se soient séparés du groupe DTU.
- ✓ Le groupe EHV a débuté sa divergence sensiblement 15 MA après sa séparation d'avec le groupe IMG.

## Les relations DMA – TUP

Dans un certain nombre de reconstructions phylogénétiques la séquence *Daubentonia* et les séquences du groupe des *Tupaia* se trouvent regroupées dans un cluster qui est localisé à la base de l'arbre. Dans la majorité des autres cas ces séquences se trouvent également à la base de l'arbre avec le groupe TUP dans la position la plus basale. Il était alors intéressant de tester la validité du groupement DMA – TUP. Pour ce faire le test 'quatre clusters' implémenté dans PHYLTEST est utilisé.

Dans un premier temps les différentes associations des groupes GRA, CHE, DMA et TUP ont été testées avec les séquences de cytochrome b, 12S rRNA, ainsi que les séquences mixtes (**Tableau 94**). La topologie dans laquelle DMA est associé à TUP présente une forte crédibilité pour les séquences cytochrome b (séquences de 357 bp et ensemble des positions 1 et 3), ainsi que pour les données mixtes. Dans le cas des séquences protéiques la crédibilité est un peu plus faible que celle trouvée avec les positions 1 et 3 du cytochrome b. Pour ces mêmes séquences la topologie (GRA, TUP), (CHE, DMA) présente une branche interne nulle, ou négative, ce qui rend totalement improbables ces associations. La topologie (GRA, DMA), (CHE, TUP) ne présente aucune crédibilité pour les séquences de cytochrome b 357 bp et pour les données mixtes. Pour les positions 1 et 3 du cytochrome b, même si la branche interne de cette topologie présente une valeur positive, celle-ci est très faible, faisant que la topologie ne peut être retenue. Pour les séquences 12S rRNA, que ce soit celles de 393 bp ou celles correspondant aux zones 'Loop', aucune des trois topologies ne présente de branches internes significativement différentes de zéro; la valeur la plus élevée de CP (0,411) étant trouvée pour la topologie (GRA, CHE), (DMA, TUP) avec les séquences des zones 'Loop'.

Topologie testée		(GRA, CHE), (DMA, TUP)	(GRA, DMA), (CHE, TUP)	(GRA, TUP), (CHE, DMA)
Cytochrome b 357 bp	Branche interne	0,0276895	-	-
	CP	0,956	-	-
Cytochrome b positions 1 et 3	Branche interne	0,031438	0,00171955	-
	CP	0,835	0,0718	-
Cytochrome b protéines	Branche interne	0,0169631	0,00647373	-
	CP	0,785	0,497	-
12S rRNA 393 bp	Branche interne	0,00218297	0,00205578	-
	CP	0,174	0,166	-
12S rRNA zones 'Loop'	Branche interne	0,00775974	-	0,00428561
	CP	0,411	-	0,259
Données mixtes	Branche interne	0,0210218	-	-
	CP	0,963	-	-

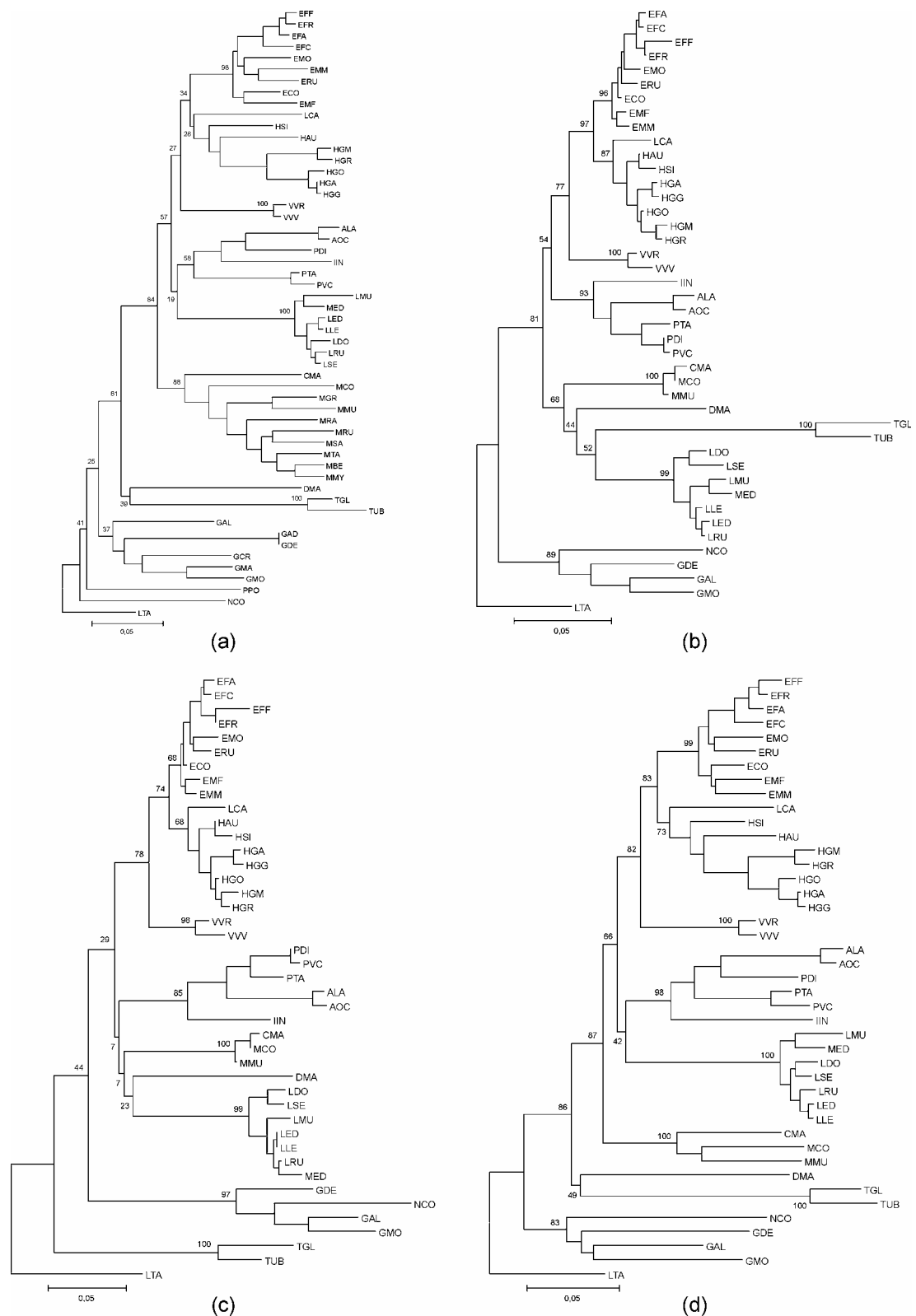
Tableau 94 : Test des topologies pour l'association DMA - TUP

Dans un second temps les relations du cluster DTU avec les groupes CHE, IMG et EHV ont été testées (**Tableau 95**). Pour ce test la topologie (EHV, DTU),(CHE, IMG) n'est crédible avec aucun des types de séquences étudiées. Pour les séquences cytochrome b 357 bp, les positions 1 et 3 du cytochrome b, les séquences 12S rRNA 393 bp, ainsi que les données mixtes, la topologie (DTU, CHE), (EHV, IMG) a une branche interne significativement différente de zéro avec des valeurs élevées de CP pour les deux types de séquences cytochrome b et plus faibles dans le cas des données mixtes et des séquences 12S de 393 bp. Cette topologie ne peut être retenue dans le cas des zones 'Loop' des séquences 12S. Enfin pour les séquences protéiques la crédibilité de cette topologie est très faible. Pour la troisième topologie: (EHV, CHE), (IMG, DTU), seules les séquences 12S rRNA 393 bp rendent cette topologie relativement crédible avec une valeur de CP de

0,715. Il est donc possible de considérer que l'association DTU – CHE est la plus crédible de toutes celles testées.

Topologie testée		(GRA, CHE), (DMA, TUP)	(GRA, DMA), (CHE, TUP)	(GRA, TUP), (CHE, DMA)
Cytochrome b 357 bp	Branche interne	0,0134725	0,000369764	-
	CP	0,936	0,0398	-
Cytochrome b positions 1 et 3	Branche interne	0,0253644	-	-
	CP	0,948	-	-
Cytochrome b protéines	Branche interne	0,00502329	0,00134503	-
	CP	0,491	0,160	-
12S rRNA 393 bp	Branche interne	0,00914839	0,00664452	-
	CP	0,770	0,715	-
12S rRNA zones 'Loop'	Branche interne	-	0,00539474	-
	CP	-	0,438	-
Données mixtes	Branche interne	0,00665478	0,0025705	-
	CP	0,729	0,376	-

**Tableau 95 : Test des topologies pour le cluster DTU**



**Figure 54 : Topologies NJ pour le test du cluster (DMA, TUP)**

La stabilité du cluster (DMA, TUP) a également été testée en augmentant le nombre de séquences du groupe des Lorisidae utilisées pour les reconstructions. Les numéros d'accès EMBL de ces séquences sont indiqués dans le **Tableau 11**. Les topologies présentées à la **Figure 54** ont été reconstruites en utilisant la méthode NJ avec un test de bootstrap de 5 000 répliquions et les

distances de Tamura-Nei prenant en compte les transitions et les transversions pour: (a) les séquences cytochrome b 357 bp, (b) les séquences 12S rRNA 393 bp, (c) les zones 'Loop' des séquences 12S et (d) les données mixtes. Dans le cas des séquences cytochrome b 357 bp il n'y a aucune modification par rapport à la topologie obtenue avec la séquence LTA seule (**Figure 18e**) et ici les nouvelles séquences de Lorisidae sont toutes placées à la base de l'arbre. Pour la reconstruction effectuée à partir des séquences 12S rRNA de 393 bp les positions relatives de DMA et TUP ne sont pas modifiées et les séquences de Lorisidae sont trouvées à la base de l'arbre. La seule modification constatée par rapport à l'analyse effectuée avec la séquence LTA seule (**Figure 41d**) concerne le groupe IND qui ne se trouve plus ici à la base de l'arbre. Dans le cas des zones 'Loop' des séquences 12S les modifications dues à l'addition des séquences de Lorisidae sont relativement importantes. Les séquences du groupe TUP se trouvent dans la position la plus basale, alors que la séquence DMA est regroupée avec celles du groupe MGA; de plus il existe un cluster (IND, CHE, DMA, MGA) qui apparaît comme étant le cluster frère du groupe EHV. Ces regroupements n'existaient pas dans l'analyse effectuée avec la séquence LTA seule (**Figure 42d**). Enfin avec les données mixtes, l'introduction de nouvelles séquences de Lorisidae provoque le regroupement de DMA et TUP dans un cluster que l'on trouve à la base d'un ensemble constitué par les Cheirogaleidae et les autres Lemuriformes, ce qui n'était pas le cas pour l'analyse avec la seule séquence LTA (**Figure 49e**). Les séquences de Lorisidae, quant à elles, occupent la position la plus basale de la topologie.

On peut donc constater qu'avec les séquences de cytochrome b et les données mixtes l'augmentation du nombre de séquences de Lorisidae utilisées, soit ne modifie pas les relations DMA – TUP (dans le cas des séquences cytochrome b 357 bp), soit permet de regrouper DMA et TUP dans un même cluster (dans le cas des données mixtes). Par contre avec les deux types de séquences 12S rRNA les perturbations des topologies sont importantes.

### **Les relations IND – MGA**

Les différentes topologies reconstruites dans cette étude font apparaître l'existence de relations particulières entre le cluster des séquences de Megaladapidae et le cluster des séquences des Indriidae.

Le test 'quatre clusters' est utilisé afin d'évaluer la crédibilité du groupement IND – MGA par rapport aux groupes EHV et CHE (**Tableau 96**).

Topologie testée		(EHV, IND), (MGA, CHE)	(EHV, MGA), (IND, CHE)	(EHV, CHE), (IND, MGA)
Cytochrome b 357 bp	Branche interne	-	0,00188306	0,00530282
	CP	-	0,213	0,528
Cytochrome b positions 1 et 3	Branche interne	-	0,00427105	0,0100121
	CP	-	0,266	0,548
Cytochrome b protéines	Branche interne	-	-	0,00693891
	CP	-	-	0,610
12S rRNA 393 bp	Branche interne	0,00626373	-	-
	CP	0,605	-	-
12S rRNA zones 'Loop'	Branche interne	-	-	0,00322814
	CP	-	-	0,236
Données mixtes	Branche interne	-	-	0,00278314
	CP	-	-	0,362

**Tableau 96 : Test des topologies pour le groupement IND – MGA**



Pour la topologie (EHV, IND), (MGA, CHE), seules les séquences 12S rRNA de 393 bp montrent une branche interne significativement différente de zéro, avec une valeur de CP de 0,605. Les deux types de séquences nucléotidiques de cytochrome b sont les seules pour lesquelles la branche interne de la topologie (EHV, MGA), (IND, CHE) n'est pas nulle, mais la crédibilité de cette topologie est faible, avec des valeurs de CP égales à 0,213 pour les séquences de 357 bp et 0,266 pour celles regroupant les positions 1 et 3. Par contre la branche interne de la topologie (EHV, CHE), (IND, MGA) présente des valeurs non nulles avec tous les types de séquences sauf celles correspondant au 12S rRNA 393 bp. Toutefois les crédibilités sont relativement peu élevées, les valeurs les plus fortes de CP étant trouvées pour les séquences de cytochrome b: 0,610 pour les séquences protéiques; 0,548 pour les positions 1 et 3, et 0,528 pour les séquences de 357 bp.

Une étude identique a été menée en considérant les groupes EHP, VAR, IND et MGA (**Tableau 97**).

Topologie testée		(EHP, VAR), (IND, MGA)	(EHP, IND), (VAR, MGA)	EHP, MGA), (VAR, IND)
Cytochrome b 357 bp	Branche interne	0,00487167	0,00108235	-
	CP	0,383	0,0718	-
Cytochrome b positions 1 et 3	Branche interne	0,0100033	-	-
	CP	0,451	-	-
Cytochrome b protéines	Branche interne	0,0105765	0,00232662	-
	CP	0,632	0,143	-
12S rRNA 393 bp	Branche interne	0,0163132	-	-
	CP	0,927	-	-
12S rRNA zones 'Loop'	Branche interne	0,0381088	-	-
	CP	0,971	-	-
Données mixtes	Branche interne	0,0175378	-	-
	CP	0,954	-	-

**Tableau 97 : Test des topologies pour les groupes EHP, VAR, IND et MGA**

La topologie (EHP, MGA), (VAR, IND) n'est crédible pour aucun des types de séquences étudiés. Par contre la topologie (EHP, VAR), (IND, MGA) présente une branche interne non nulle avec tous les types de séquences et elle apparaît la plus crédible pour les deux séquences 12S rRNA, ainsi que pour les données mixtes, avec des valeurs de CP supérieures à 0,9. Dans le cas des séquences cytochrome b la crédibilité est plus faible avec des valeurs de CP inférieures à 0,5.

On peut donc considérer que l'association, dans les reconstructions effectuées, des clusters formés par les séquences des Indriidae d'une part, et celles des Megaladapidae d'autre part, est crédible.

Un autre problème intéressant, dans le groupe des Megaladapidae concerne les relations entre le sub-fossile MED et les autres séquences de Lepilemuridae. Une étude par le test 'quatre clusters' a été effectuée pour ces différentes séquences. Les résultats les plus significatifs sont reportés dans le **Tableau 98**. On peut constater que ce sont, majoritairement, les topologies dans lesquelles MED est associé à LMU qui sont les plus crédibles pour les séquences cytochrome b et les données mixtes. Cette même association est trouvée pour les séquences 12S rRNA de 393 bp, mais dans ce cas l'association de MED avec LDO ou LED est également crédible. Dans le cas des zones 'Loop' du 12S, seule l'association de MED avec LED présente une valeur de CP supérieure à 0,9. Enfin pour les séquences protéiques du cytochrome b l'association MED – LRU donne une valeur de CP de 0,841.

Séquences	Topologie	Branche interne	CP
<b>Cytochrome b 357 bp</b>	(MED, LMU), (LDO, LED)	0,0128438	0,920
	(MED, LMU), (LED, LLE)	0,0205304	0,990
	(MED, LMU), (LLE, LRU)	0,015592	0,960
	(MED, LMU), (LLE, LSE)	0,0163356	0,969
	(MED, LMU), (LRU, LSE)	0,0198738	0,987
<b>Cytochrome b positions 1 et 3</b>	(MED, LMU), (LED, LLE)	0,026836	0,983
	(MED, LMU), (LLE, LRU)	0,0248448	0,968
	(MED, LMU), (LLE, LSE)	0,0250601	0,969
	(MED, LMU), (LRU, LSE)	0,0314325	0,990
<b>Cytochrome b protéines</b>	(MED, LRU), (LDO, LED)	0,00873401	0,841
<b>12S rRNA 393 bp</b>	(MED, LMU), (LDO, LED)	0,0110695	0,901
	(MED, LDO), (LED, LRU)	0,0106962	0,925
	(MED, LED), (LDO, LSE)	0,0213091	0,988
	(MED, LMU), (LLE, LSE)	0,0139145	0,946
	(MED, LMU), (LRU, LSE)	0,0110774	0,905
<b>12S rRNA zones 'Loop'</b>	(MED, LED), (LDO, LSE)	0,0261433	0,965
<b>Données mixtes</b>	(MED, LMU), (LDO, LED)	0,0106675	0,949
	(MED, LED), (LDO, LSE)	0,0152632	0,989
	(MED, LMU), (LLE, LRU)	0,0125894	0,977
	(MED, LMU), (LLE, LSE)	0,0109146	0,962
	(MED, LMU), (LRU, LSE)	0,0131764	0,984

**Tableau 98 : Tests des topologies pour les Megaladapidae**

Il est donc possible de supposer que le sub-fossile MED est plus proche de LMU que des autres Lepilemuridae actuels.

## DISCUSSION

Comme cela a été noté auparavant les études de biologie moléculaire effectuées à ce jour sur les Lemuriformes ont essentiellement porté sur tel ou tel groupe particulier dans le but d'évaluer les relations existant entre les différentes espèces. Par contre pour les quelques études dans lesquelles un certain nombre de groupes sont analysés, chaque groupe n'est généralement caractérisé que par une ou un petit nombre de séquences.

Dans notre étude nous avons pu rechercher les relations existant entre les différentes familles de Lemuriformes, en utilisant pour chacune d'elles la plupart des espèces actuellement décrites. Les différentes séquences concernant le fragment de 357 bp du cytochrome b, ainsi que le fragment de 393 bp du 12S rRNA ont été, soit extraites de la base de données EMBL, soit amplifiées au Laboratoire à partir des échantillons disponibles. De plus notre travail inclut les séquences amplifiées à partir d'un sub-fossile (*Megaladapis edwardsi*), ainsi que celles caractérisant deux représentants de la famille des Tupaia.

Comme nous avons pu le montrer précédemment (226), l'augmentation du nombre de taxa inclus dans une analyse phylogénétique est un facteur susceptible d'augmenter la robustesse et la crédibilité des reconstructions. Ce phénomène est particulièrement sensible avec les Lemuriformes.

### *Relations entre les principales familles*

Dans un travail portant sur le gène du cytochrome b complet (1140 bp), ainsi que sur les données du CO II déjà traitées par Adkins et Honeycutt (5), Yoder et coll. (390) ont donné une représentation des relations entre les différentes familles de Lemuriformes, analysées par la méthode du maximum de parcimonie (**Figure 55A**). Dans celle-ci le *Daubentonia* se trouve à la base des Lemuriformes, mais les relations entre les Indriidae, Cheirogaleidae et Lemuridae ne peuvent être précisées, bien que ces auteurs notent que le cytochrome b est le gène le plus informatif pour les clades récents et qu'il semble être peu utilisable dans le cas du *Daubentonia*. Par contre la combinaison des deux gènes est en mesure de fournir des reconstructions plus robustes, en particulier en ce qui concerne le placement du *Daubentonia*.

Par la suite des données morphologiques, portant sur des caractères d'anatomie craniale et post craniale ont été utilisées par Yoder et coll. (387), seules ou en combinaison avec les données moléculaires provenant du cytochrome b total. Dans ces analyses les données morphologiques utilisées incluent les sites précédemment utilisés pour la mise en évidence des affinités entre Cheirogaleidae et Lorisiformes. Lorsque la reconstruction est effectuée en ne tenant compte que des données moléculaires (**Figure 55C**), les différents modèles utilisés (attribuant des poids différents aux transitions et transversions) placent les Cheirogaleidae dans les Primates malgaches et à une position plus interne que celle occupée par le *Daubentonia*. De plus il existe un fort support pour ce placement des Cheirogaleidae (valeurs de bootstrap comprises entre 97 et 99 %). Par contre le support du clade complet, comprenant le *Daubentonia*, est plus faible (entre 50 et 61 %). Cette faible valeur pour le clade entier est également trouvée à partir des données morphologiques (**Figure 55B**), mais ici les Cheirogaleidae sont en position basale de la reconstruction, le *Daubentonia* se trouvant en position plus interne avec une disposition supportée par une valeur élevée de bootstrap (86 %). Enfin la combinaison des deux types de données (**Figure 55D**) fournit une topologie identique à celle obtenue avec les seules données moléculaires et dans ce cas la valeur

de bootstrap supportant le clade entier est comprise entre 61 et 70 %, et ce même si les caractères morphologiques ont un poids quatre fois supérieur à celui des caractères moléculaires.

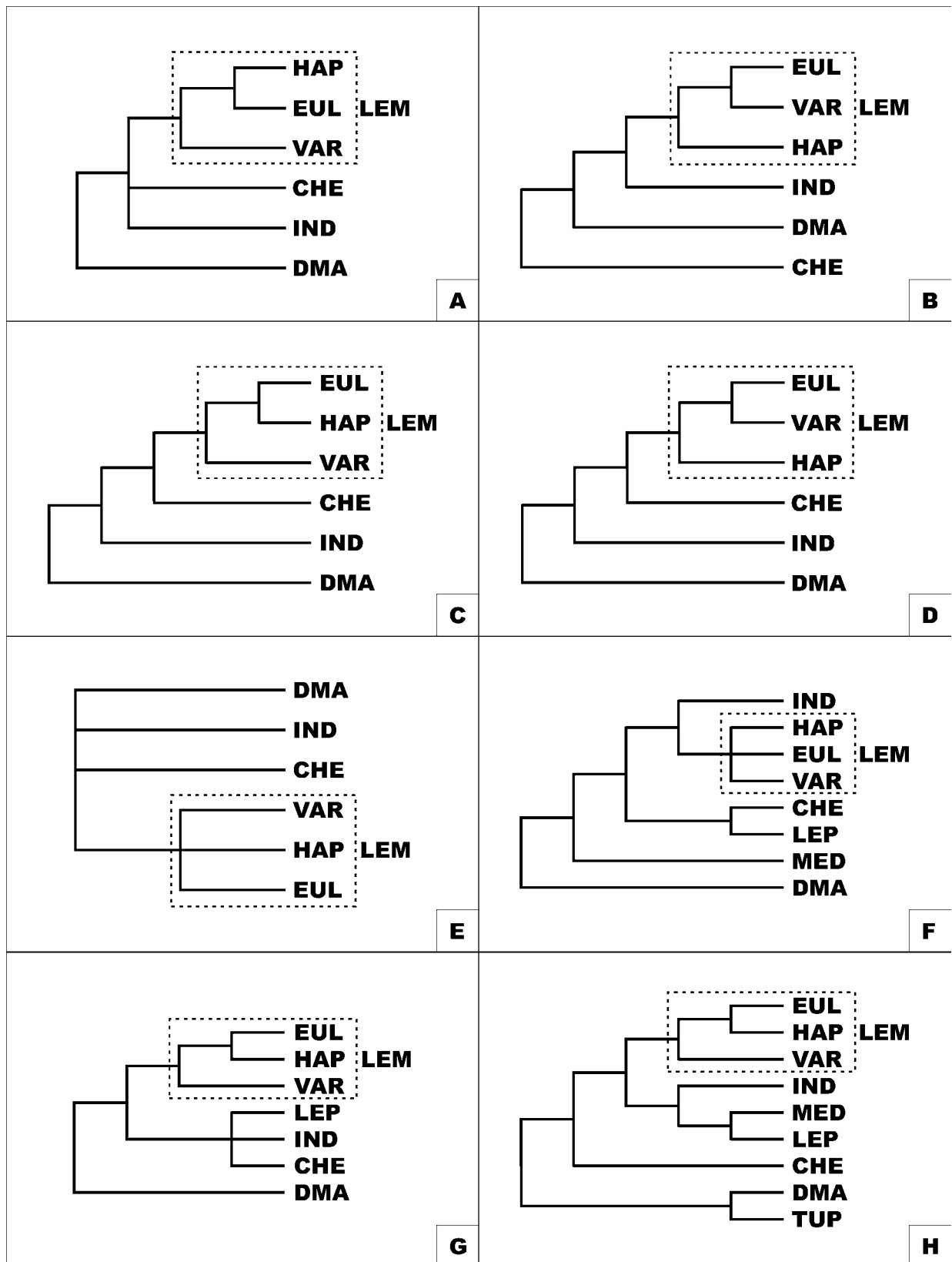


Figure 55 : Résumé des différentes topologies concernant la famille des Lemuriformes

On peut par ailleurs noter que les travaux de Adkins et Honneycutt (5) sur le CO II avaient permis de montrer que le DMA était une espèce sœur des autres Strepsirhini. Mais la valeur de bootstrap supportant cette distribution était faible.

Stanger-Hall et Cunningham (316) ont analysé des séquences partielles de 16S rRNA pour un grand nombre de Primates et les résultats concernant les Lemuriformes sont présentés à la **Figure 55E**. Il n'est pas ici possible d'établir une séquence évolutive pour *Daubentonia*, Indriidae, Cheirogaleidae et Lemuridae. Ces auteurs attribuent l'apparente incongruence entre reconstructions effectuées à partir de différents segments d'ADN à l'utilisation de modèles de reconstruction inappropriés. Ils montrent également que l'application d'un modèle de parcimonie à six paramètres (chaque type de transition et de transversion ayant un poids différent) augmente le support de la monophylie de certains groupes. L'application de modèles trop réalistes pourrait introduire les incongruences notées et il semblerait que le modèle MP soit le plus performant dans le cas de l'analyse de données combinées.

Dans une analyse portant sur un fragment de cytochrome b et incluant le sub-fossile *Megaladapis*, Yoder et coll. (389) placent le *Daubentonia* à la base des autres Lemuriformes (**Figure 55F**), alors que Cheirogaleidae et Lepilemuridae forment un cluster apparaissant comme cluster frère d'un autre regroupant Indriidae et Lemuridae. L'association CHE – LEP est supportée par une valeur de bootstrap de 64 %, celle des Indriidae – Lemuridae par une valeur inférieure à 50 %, et celle de l'ensemble des Lemuriformes (hors mis le *Daubentonia*) par une valeur de 89 %. Mais on peut noter que dans cette analyse une seule séquence de Lepilemuridae est utilisée, de même que les Indriidae ne sont caractérisés que par deux séquences de Propithecus. Si l'analyse n'est plus effectuée avec la totalité des sites, mais en prenant seulement en compte les troisièmes positions des codons, la séquence de Lepilemuridae apparaît groupée avec les Indriidae et les Cheirogaleidae, cet ensemble formant un cluster frère d'un autre constitué par les Lemuridae. Par contre *Megaladapis* et *Daubentonia* occupent les mêmes positions que précédemment.

Enfin dans un travail récent Pastorini (255) (**Figure 55G**), en utilisant les données issues du gène mitochondrial ND4, n'est pas en mesure de préciser la séquence évolutive pour les Lepilemuridae, Indriidae et Cheirogaleidae qui sont groupés dans un clade frère de celui renfermant les Lemuridae. Par contre le *Daubentonia* est toujours en position basale par rapport à tous les autres Lemuriformes.

Dans la reconstruction que nous considérons comme la plus crédible à partir de nos données (**Figure 55H**) les Cheirogaleidae se trouvent placés à la base d'un ensemble de deux clusters dont l'un regroupe les Lemuridae et l'autre les Indriidae et Megaladapidae (Lepilemurs et *Megaladapis*) et cette disposition est supportée par une valeur de bootstrap de 77 %. Le regroupement Lemuridae, Indriidae, Megaladapidae étant, pour sa part, supporté par une valeur plus faible de bootstrap (55 %). Si l'association *Megaladapis* – Lepilemurs est supportée par une valeur de bootstrap de 100 %, dans le cas des Indriidae – Megaladapidae la valeur est très faible (16 %). Toutefois les tests 'quatre clusters' qui ont été menés sont en faveur de cette association. On peut par ailleurs remarquer qu'un certain nombre d'arguments morphologiques sont également en faveur de cette relation; en effet les animaux des deux groupes sont tous des 'sauteurs', bien qu'ils ne présentent pas tous le même mode de vie, certains étant nocturnes et d'autres diurnes (251). D'autre part nous avons trouvé un groupement du *Daubentonia* avec les Tupaia, dont le support de bootstrap est de 49 %, et ici encore les tests 'quatre clusters' sont en faveur de ce type d'association.

### ***Position phylogénétique de Megaladapis edwardsi***

Le sub-fossile *Megaladapis edwardsi*, pour lequel nous avons pu amplifier le fragment de 357 bp du cytochrome b et celui de 393 bp du 12S rRNA occupe dans nos reconstructions (**Figure 55H**) une place très différente de celle qui avait été trouvée dans un travail de Yoder et coll. (1999) (Figure 55F).

Comme une étude antérieure incluant ce sub-fossile et de nombreuses séquences de Lemuriformes (225), les analyses effectuées dans ce travail avec les différents types de séquences et que ce soit avec les méthodes de maximum de parcimonie, de maximum de vraisemblance et de Neighbor-Joining, montrent toujours un regroupement du *Megaladapis* avec les Lepilemurs. De plus ce sub-fossile se trouve toujours en position basale du clade des Lepilemuridae. Dans le cas des analyses effectuées en utilisant la méthode de Neighbor-Joining cette association est supportée par une valeur de bootstrap de 100 %. Une telle relation est en accord avec la classification qui avait été précédemment établie par Schwartz et Tattersall (291) sur la base de caractères morphologiques.

Dans le travail de Yoder, portant sur un fragment de cytochrome b, le *Megaladapis* est trouvé en position basale de tous les Lemuriformes, cette association étant supportée par une valeur de bootstrap de 89 %. Il faut également remarquer que dans ce travail si la séquence d'un autre sub-fossile (*Palaeopropithecus*) est également incluse dans l'analyse, les deux séquences ancestrales se regroupent et occupent une position identique à celle définie avec le *Megaladapis* seulement. Dans ce dernier cas l'association séquences ancestrales – Lemuriformes est supportée par une valeur de bootstrap de 90 %.

Les différences entre les résultats de Yoder et ceux que nous avons obtenus peuvent avoir plusieurs explications. D'une part notre analyse porte sur l'ensemble des Lepilemuridae actuellement déterminés, ainsi que sur la presque totalité des autres Lemuriformes alors que dans le travail de Yoder une seule séquence de Lepilemuridae est utilisée, et les autres familles de Lemuriformes ne sont caractérisées que par une ou quelques séquences. D'autre part Yoder précise que la séquence ancestrale de cytochrome b, si elle provient de trois spécimens différents de *Megaladapis*, n'a pu être entièrement déterminée. Cette séquence ne comporte que 350 nucléotides utilisables, alors que pour les autres espèces 550 nucléotides sont traités. La séquence de cytochrome b que nous avons pu amplifier, quant à elle, représente un segment continu de 357 bp et même si elle ne provient de l'amplification que d'un seul échantillon, l'absence d'indéterminations doit être en mesure de fiabiliser les résultats. Il faut également remarquer qu'à partir du même sub-fossile nous avons été en mesure d'amplifier un fragment de 12S rRNA et que les reconstructions effectuées en utilisant celui-ci fournissent des résultats identiques à ceux trouvés dans le cas du cytochrome b.

On pourrait envisager l'hypothèse d'une contamination du *Megaladapis* par des séquences de Lepilemuridae actuels, favorisant le regroupement que nous avons observé. Cette hypothèse a été testée en comparant (avec le programme FASTA) la séquence de *Megaladapis* avec toutes celles amplifiées au Laboratoire. Les séquences les plus proches, si elles sont bien celles de Lepilemuridae, présentent une différence de 14 nucléotides (soit sensiblement 4 % des 357 bp) avec le sub-fossile, ce qui permet d'exclure un phénomène de contamination.

Le fait que Yoder, même en effectuant des amplifications partielles (portant sur de petites zones) n'ait été en mesure d'amplifier la totalité des 550 bp de la zone étudiée tend à montrer que l'ADN mitochondrial extrait était, au moins partiellement, dégradé. Une telle dégradation peut être

due à une mauvaise préservation des échantillons. Il faut également remarquer que les techniques d'extraction utilisées peuvent être, elles aussi, source de problèmes affectant la qualité de l'ADN et sa capacité à être amplifié, par exemple parce que des cofacteurs inhibant la Taq polymérase peuvent être extraits en même temps que l'ADN. Dans notre travail nous avons utilisé une technique spécifique que nous avons mise au point pour ce type d'étude (56) et qui est en mesure de préserver, au maximum, l'ADN mitochondrial, ainsi que de minimiser la présence de cofacteurs inhibants.

### *Variabilité de l'horloge moléculaire*

Sur un plan très général l'hypothèse d'une horloge moléculaire suppose que, pour un gène donné, les mutations se soient accumulées régulièrement au long des branches de l'arbre.

Dans le cas de l'analyse de caractères moléculaires l'utilisation d'une méthode cladistique permet d'établir une phylogénèse. Il est alors possible de dénombrer les mutations présentes dans les différentes branches; les valeurs trouvées permettant alors de vérifier si le taux d'évolution est constant ou non. Si de plus on dispose d'un point de repère (qui est généralement obtenu à partir d'un ou plusieurs fossiles) il devient possible de calibrer l'horloge moléculaire et d'estimer ainsi la vitesse d'évolution du gène considéré. Il sera enfin possible d'effectuer une datation des événements de spéciation et ainsi de passer d'une datation relative à une datation absolue.

Dans de très nombreux cas l'absence de fossiles vrais pose un problème pour la reconstruction de l'histoire phylogénétique. Il faut alors tenir compte de la distribution des espèces actuelles en admettant que la position géographique du nœud ancestral correspond à la localisation actuelle des espèces.

Si la séquence, ou le gène, étudiée a évolué sous un système prenant en compte l'existence d'une horloge moléculaire, les données peuvent être analysées par une méthode de distances afin de déterminer les temps relatifs des divergences.

Classiquement on considère que l'ADN mitochondrial provient uniquement de l'ovule et qu'ainsi il se comporte à la manière d'un locus non recombinant et à transmission matri-linéaire. Toutefois des études récentes ont montré que des mitochondries provenant du spermatozoïde pouvaient aussi pénétrer l'ovule. Ainsi, contrairement à ce que l'on pouvait penser, il pourrait y avoir des phénomènes de recombinaison entre les ADN mitochondriaux de l'ovule et du spermatozoïde. Si un tel phénomène est en mesure de se produire, même avec une faible fréquence, il peut, soit introduire des modifications multiples, soit supprimer des modifications préexistantes. Il y aurait alors un dérèglement de l'horloge moléculaire dont l'une des principales conséquences serait l'introduction d'erreurs sur les estimations des vitesses de changement des nucléotides. Il faut également noter qu'un autre phénomène, pouvant aboutir aux mêmes résultats sur l'horloge moléculaire, est la réversion des mutations. Mais il est très difficile d'évaluer la portée d'une telle réversion dans la mesure où l'on ne dispose pas généralement de l'ensemble des séquences allant de la forme la plus ancestrale à la forme actuelle.

Si l'on considère que les phénomènes décrits ci-dessus sont susceptibles de se produire chez les Lemuriformes, ceci pourrait apporter une explication, du moins partielle, aux différences observées entre les reconstructions effectuées à partir de différents gènes, certains pouvant être plus sensibles que d'autres à la recombinaison, à la réversion, ou aux deux à la fois.

Le concept d'une horloge moléculaire globale n'étant pas utilisable pour toutes les reconstructions, l'hypothèse d'horloges moléculaires locales a été formulée par Yang et Yoder

(382). Dans notre travail l'étude des taux de substitution par site montre l'existence de différences, parfois importantes, pour les différentes familles. Les tests effectués crédibilisent l'existence de tels processus pour les Lemuriformes (test de constance du taux d'évolution dans le logiciel LINETREE). Dans le cas du cytochrome b les taux de substitution pour les branches menant aux Tupaia et *Daubentonia* sont sensiblement deux fois plus élevés que ceux déterminés pour les branches menant aux Lemuridae ou Indriidae et cinq fois plus élevés que pour la branche menant aux Megaladapidae (Lepilemurs et *Megaladapis*). Par contre dans le cas des séquences 12S rRNA, si la branche menant aux Tupaia présente le taux de substitution le plus élevé, celle menant au Megaladapidae montre une valeur double de celle menant au *Daubentonia*; alors que pour les branches menant aux autres familles de Lemuriformes les taux de substitution sont faibles. Enfin lorsque les deux types de séquences sont combinés, dans l'analyse il existe une relative constance des taux de substitution par branche sauf en ce qui concerne les Megaladapidae (dans le cas d'une reconstruction ML) et des Lemuridae (dans le cas d'une reconstruction MP). Il apparaît donc que l'on puisse bien considérer qu'il existe, chez les Lemuriformes, des horloges moléculaires locales plutôt qu'une horloge globale.

### ***Datation des principales divergences***

Les sub-fossiles ne pouvant être considérés, sur le plan évolutif, comme de vrais fossiles, nous avons choisi d'effectuer la datation des divergences entre les différentes familles de Lemuriformes en prenant comme référence la divergence déterminée entre Lorisidae et Strepsirhini, telle qu'elle a été évaluée par Yoder et coll. (387), soit 62 millions d'années. Ces datations sont effectuées en utilisant la topologie apparaissant comme la plus crédible (obtenue par la méthode NJ avec le cytochrome b, ou l'association du cytochrome b et du 12S rRNA) et en utilisant la méthode des 'arbres linéarisés'.

Nous avons ainsi trouvé que l'ensemble formé des Tupaia et du *Daubentonia* avait divergé de l'ensemble des Lemuriformes entre 60 et 56 millions d'années, soit peu de temps après la séparation Lorisidae – Strepsirhini. Dans les Lemuriformes les Cheirogaleidae ont divergé il y a environ 47 millions d'années; puis on trouve la divergence entre d'une part les Indriidae et Megaladapidae et d'autre part les Lemuridae (il y a environ 42 millions d'années). La divergence entre *Megaladapis* et Lepilemurs est la plus récente de toutes celles déterminées, elle se situe il y a environ 10 millions d'années. Sur l'ensemble de notre travail deux divergences apparaissent comme bien individualisées, ce sont celles entre Tupaia et *Daubentonia* et entre *Megaladapis* et Lepilemurs. Pour les autres divergences les différentes estimations fournissent des écart types tels qu'il est beaucoup plus difficile de préciser la séquence exacte.

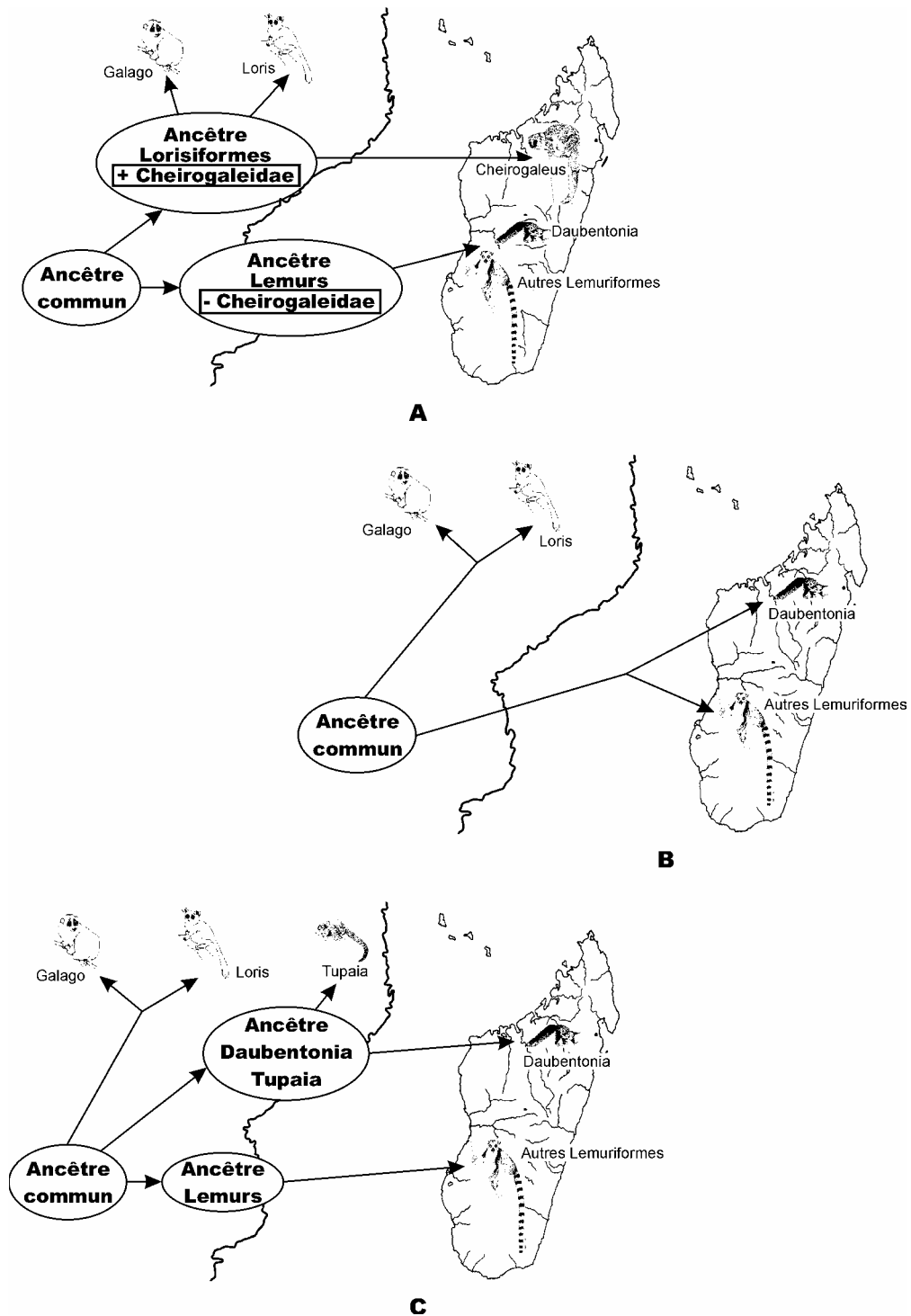
### ***Origine monophylétique des Lemuriformes***

Les Primates malgaches peuvent être considérés comme un assemblage taxonomique très diversifié et pour lequel un grand nombre de relations phylogénétiques ont été, et sont encore, sujettes à d'importantes controverses.

Pour un certain nombre d'auteurs ces Lemuriformes ne formeraient pas un groupe monophylétique et l'on pourrait alors considérer que leur présence à Madagascar serait le résultat d'un phénomène de migrations multiples. Entre autre pour Charles-Dominique et Martin (46) les Cheirogaleidae appartiendraient au même groupe que celui des Lorisidae, favorisant ainsi l'interprétation d'une double colonisation (**Figure 56A**). Par contre l'ensemble des données cytogénétiques disponibles, ainsi que de nombreuses données moléculaires, montrent clairement



une origine monophylétique des Lemuriformes avec l'existence d'un ancêtre commun aux Lorisidae et Lemuriformes (**Figure 56B**).



**Figure 56 : Hypothèses concernant la colonisation de Madagascar par les Lemuriformes**

Dans les travaux de Yoder (387) l'analyse de données morphologiques détaillées (portant sur ces caractéristiques craniales), ainsi que celles provenant du gène mitochondrial du cytochrome b,

qu'elle soit effectuée pour chaque type de donnée séparément, ou pour la combinaison des deux, montre que la totalité des Primates malgaches, y compris les Cheirogaleidae et le *Daubentonia*, doit être considéré comme un ensemble monophylétique. Lorsque les séquences du cytochrome b sont analysées les résultats suggèrent que les Lémuriens seraient arrivés à Madagascar et auraient commencé à se diversifier au début de l'Eocène. D'autre part la souche de ces Lemuriens aurait son origine en Afrique et ces animaux ont sans doute rejoint Madagascar en empruntant des radeaux de végétation. On admet généralement que ces migrants étaient de petite taille et que, par la suite, ils ont évolué sans aucun autre apport extérieur.

Les reconstructions obtenues dans notre travail, qu'elles utilisent soit le fragment de cytochrome b, soit la combinaison de ce fragment avec un autre provenant du 12S rRNA, montrent que les Primates malgaches forment un groupe monophylétique, mais qu'il existe à la base de ce groupe un autre dans lequel sont regroupés les Tupaia et le *Daubentonia*. L'utilisation de tests 'quatre clusters' nous a permis de montrer qu'avec les différents types de fragments d'ADN mitochondrial que nous avons utilisés, la crédibilité de l'association DMA – TUP était élevée. La stabilité de ce cluster a également été testée en augmentant le nombre de séquences utilisées pour le groupe des Lorisidae. Dans le cas des séquences de cytochrome b cette augmentation du nombre de taxa ne modifie en rien la topologie initiale (obtenue avec la seule séquence LTA). De même pour les séquences 12S rRNA les positions de DMA et TUP, à la base des Lemuriformes, n'est pas modifiée, seule la famille des Indriidae ne se retrouve plus dans la position qu'elle occupait avec la seule séquence LTA. Par contre dans l'analyse portant sur la combinaison des séquences de cytochrome b et de 12S rRNA, l'augmentation du nombre de séquences de Lorisidae utilisées permet la reconstruction du cluster (DMA – TUP).

Ces résultats permettent de définir une origine monophylétique des Lemuriformes dans lesquels sont inclus les Cheirogaleidae, mais ne prenant pas en compte le *Daubentonia*. Il est ainsi possible d'évoquer l'existence d'une colonisation double de Madagascar (**Figure 56C**). L'une de celles-ci aurait introduit l'ancêtre du *Daubentonia* (les Tupaia n'étant pas présents sur l'île), l'autre un ancêtre qui, par la suite, aurait donné naissance aux différents groupes sub-fossiles et actuels. Il n'est toutefois pas possible, avec l'ensemble des données que nous avons obtenues, de préciser si ces deux migrations se sont produites simultanément ou si l'une a été antérieure à l'autre.

L'hypothèse de cette double colonisation est en opposition avec les résultats de Charles-Dominique et Martin (46) (**Figure 56A**), ainsi que ceux obtenus par Yoder et coll. (387) (**Figure 56B**). Toutefois il faut remarquer que pour ces deux travaux aucune séquence de la famille des Tupaia n'a été utilisée. Comme le montre les estimations de dates de divergence que nous avons effectuées la séparation entre l'ensemble (DMA- TUP) et les autres Lemuriformes est ancienne (voisine de 58 millions d'années) et très proche de celle qui a pu être définie pour la séparation entre Lorisidae et Strepsirhini (62 millions d'années). Par contre la première divergence dans le groupe des Lemuriformes, permettant l'individualisation des Cheirogaleidae est beaucoup plus récente (voisine de 46 millions d'années). Elle est à rapprocher de celle déterminée par Yoder et coll. (387) pour le début de la radiation des Primates malgaches (environ 54 millions d'années). On peut alors supposer qu'il aurait existé, sur le continent Africain, à partir d'un ancêtre commun, un groupe ancestral aux Lemuriformes et un autre ancestral de l'ensemble (DMA – TUP). Les individus de ces deux groupes ont sans doute du présenter des caractéristiques morphologiques proches de celles des Microcèbes. La position taxonomique des Tupaia qui est fortement controversée à partir des études morphologiques et qui n'a pu être élucidée en quelque sens que ce soit par des études cytogénétiques (75) peut, au moins en partie, à partir de nos résultats, être précisée en évoquant l'existence d'un ancêtre commun entre *Daubentonia* et Tupaia. Cet ancêtre aurait donné, sur le continent Africain une lignée aboutissant aux Tupaia actuels et une seconde, ayant migré vers

Madagascar, qui serait à l'origine des *Daubentonia*. On pourrait alors inclure les *Tupaia* dans les Lemuriformes.

L'hypothèse proposée par Charles-Dominique et Martin (46) (**Figure 56A**) semble pouvoir être réfutée en ce qui concerne une colonisation séparée des Cheirogaleidae, toutes les reconstructions que nous avons effectuées montrant que les Cheirogaleidae font bien partie du groupe monophylétique des Lemuriformes.

## CONCLUSIONS

Les reconstructions phylogénétiques effectuées à partir des données issues des fragments de cytochrome b et de 12S rRNA amplifiés à partir des différents groupes de Lémuriformes ont été comparées en utilisant des tests LTR (Likelihood Test Ratio), ainsi que des tests 'quatre clusters'. Il apparaît ainsi que celle présentant le maximum de 'crédibilité' peut être représentée de la manière suivante:

((Cheirogaleidae, ((Indriidae, (Lepilemuridae, *Megaladapis*)),  
((*Eulemur*, *Hapalemur*), *Varecia*)), (Daubentoniidae, Tupaia))

On peut constater que le sub fossile *Megaladapis* se trouve associé au cluster de Lepilemuridae. D'autre part les Lémuriformes apparaissent groupés en deux clusters principaux: l'un composé des Lemuridae et l'autre des Indriidae, *Megaladapis* et Lepilemuridae. A la base du cluster des Lémuriformes se trouvent les Cheirogaleidae. Cette topologie est, dans ses grandes lignes, conforme à celles qui peuvent être obtenues en utilisant différents autres critères d'analyse, tels que des caractéristiques morphologiques ou caryologiques. Toutefois la répartition des séquences dans certains genres, tels que *Hapalemur* et *Lepilemur*, ne recoupe pas entièrement les phylogénèses caryologiques déjà publiées. Mais il faut remarquer que l'on dispose de peu de données provenant de l'utilisation de techniques, autres que la biologie moléculaire, et utilisant l'ensemble des Lémuriformes pour aider à la validation des reconstructions effectuées.

L'association, dans un même cluster, des Daubentoniidae et Tupaia est une donnée importante, même si la place des Tupaia a fait l'objet de multiples controverses: pour certains auteurs ce groupe se trouvant très séparé des Strepsirhini, voir même totalement en dehors du groupe des Primates. Comme les Tupaia ne sont pas représentés à Madagascar on peut supposer qu'il a existé deux colonisations distinctes de l'île, dont l'une a donné naissance à la lignée des Daubentoniidae, et l'autre à celle de l'ensemble des autres Strepsirhini malgaches.

La topologie définie comme étant le plus 'crédible' a été utilisée pour tester l'existence d'une horloge moléculaire globale, ou d'horloges moléculaires locales. L'étude des substitutions nucléotidiques par site montre en effet, sur l'ensemble des séquences étudiées, l'existence d'une hétérogénéité du taux de substitution par site. D'autre part pour les différents groupes de Lémuriformes ces taux sont également très variables. Les valeurs estimées du paramètre de forme (appelé alpha) de la distribution gamma varient de 0,091595 chez les Lepilemuridae à 0,504956 chez les Indriidae. En d'autres termes le groupe des Lepilemuridae présente un taux d'évolution beaucoup plus important que celui des Indriidae. Par contre les taux d'évolution trouvés pour les Cheirogaleidae et les *Eulemur* sont sensiblement identiques. Ces résultats permettent de suspecter l'existence d'horloges moléculaires locales, plutôt que d'une horloge moléculaire globale. Les études portant sur ces horloges, effectuées à partir de la topologie sélectionnée, permettent de confirmer les résultats obtenus à partir de l'étude des taux de substitution pour tous les groupes, sauf celui des Cheirogaleidae. En résumé les taux d'évolution pour les branches de la topologie menant aux *Hapalemur* et *Varecia* sont sensiblement identiques et ce taux est un peu plus élevé pour la branche menant aux *Eulemur*. Dans le cas des Cheirogaleidae, si la valeur du paramètre alpha est relativement élevée lorsqu'elle est calculée à partir des séquences des différents représentants du groupe, le taux d'évolution pour la branche menant à ce groupe est le plus bas de tous ceux trouvés. Une telle situation peut traduire l'existence d'une évolution lente jusqu'à la séparation des différentes espèces de Cheirogaleidae, suivie d'une évolution plus rapide de ces espèces. Dans les

autres groupes étudiés on constate généralement une constance, ou une légère diminution, du taux d'évolution avant et après la séparation des espèces.

A partir des différentes reconstructions effectuées il a été possible d'évaluer les dates de divergence des différents groupes étudiés, ainsi que d'évaluer les séquences ancestrales aux nœuds principaux de la topologie. L'utilisation de la méthode des 'arbres linéarisés' permet, en considérant que la séparation entre Lorisidae et Strepsirhini est intervenue il y a 62 millions d'années (c'est à dire bien après que Madagascar se soit séparé du continent africain, il y a environ 150 millions d'années), l'estimation des principales dates de divergence. Ainsi la séparation la plus ancienne (voisine de 60 millions d'années) est trouvée pour la divergence entre Daubentonia et Tupaia; alors que la plus récente (voisine de 11 millions d'années) est celle impliquant le *Megaladapis* et les Lepilemuridae. Pour les autres groupes les dates de divergence se répartissent entre 48 et 29 millions d'années.

Les données obtenues au cours de cette étude pourront être utilisées, par la suite, pour l'étude de l'ensemble des Lémuriformes sub fossiles en utilisant la topologie définie ici, ainsi que les séquences ancestrales estimées, comme guide pour préciser les relations pouvant exister entre ces sub fossiles et les Lémuriformes actuels.

## BIBLIOGRAPHIE

1. Adachi J., Cao Y., Hasegawa M. (1993). Tempo and mode of evolution of mitochondrial DNA evolution in Vertebrates at the amino acid sequence level: rapid evolution in warm-blooded Vertebrates. *J. Mol. Evol.* 36:270-281.
2. Adachi J., Hasegawa M. (1994). MOLPHY: programs for molecular phylogenetics. Tokyo, Institute of Statistical Mathematics.
3. Adachi J., Hasegawa M. (1996). Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J. Mol. Evol.* 42:459-468.
4. Adey N. B., Tollefsbol T. O., Sparks A. B., Edgell M. H., Hutchinson C. A. (1994). Molecular resurrection of an extinct ancestral promoter for mouse L1. *Proc. Natl. Acad. Sci. USA* 91:1569-1573.
5. Adkins R. M., Honeycutt R. L. (1994). Evolution of the Primate cytochrome c oxidase subunit II gene. *J. Mol. Evol.* 38:215-231.
6. Adkins R. M., Honeycutt R. L., Disotell T. R. (1996). Evolution of eutherian cytochrome c oxidase subunit II: heterogeneous rates of protein evolution and altered interaction with cytochrome c. *Mol. Biol. Evol.* 13:1393-1404.
7. Akashi H. (1995). Inferring weak selection from patterns of polymorphism and divergence at 'silent' sites in *Drosophila* DNA. *Genetics* 139:1067-1076.
8. Allard M. W., Carpenter J. M. (1996). On weighting and congruence. *Cladistics* 12:183-198.
9. Alvarez-Valin F., Jabbari K., Carels N., Bernardi G. E. (1999). Synonymous and nonsynonymous substitutions in genes from *Gramineae*: intragenic correlations. *J. Mol. Evol.* 49:330-342.
10. Andersson S. G. E., Stothard D. R., Fuerst P., Kubland C. G. (1999). Molecular phylogeny and rearrangement of rRNA genes in *Rickettsia* species. *Mol. Biol. Evol.* 16:987-995.
11. Archie J. (1989). Homoplasy excess ratio: new indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. *Syst. Zool.* 38:253-269.
12. Arnason U., Gullberg A. (1996). Cytochrome b nucleotide sequences and the identification of five primary lineages of extant cetaceans. *Mol. Biol. Evol.* 13:407-417.
13. Arnason U., Gullberg A., Janke A. (1998). Molecular timing of Primate divergences as estimated by two nonprimate calibration points. *J. Mol. Evol.* 47:718-727.
14. Arnason U., Xu X., Gullberg A., Graur D. (1996). The '*Phoca* standard': an external molecular reference for calibrating recent evolutionary divergences. *J. Mol. Evol.* 43:41-45.
15. Asenjo A. B., Rim J., Oprian D. D. (1994). Molecular determinants of human red/green colour discrimination. *Neuron* 12:1131-1138.
16. Atteson K. (1996). An analysis of the performance of the neighbor-joining method of phylogeny reconstruction. Workshop on Mathematical hierarchies and biology., Rutgers University,.
17. Avise J. C. (1992). Mitochondrial DNA evolution of a turtle pace: evidence for low genetic variability and reduced microevolutionary rate in the testudines. *Mol. Biol. Evol.* 9:457-473.
18. Avise J. C., Walker D., Johns J. C. (1998). Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proc. R. Soc. Lond. B* 265:1707-1712.
19. Bailey W. J., Fitch D. H. A., Tagle D. A., Gzelusniak J., Slightom J. L., Goodman M. (1991). Molecular evolution of the Yh-globin gene locus: gibbon phylogeny and the hominoid slowdown. *Mol. Biol. Evol.* 8:155-184.
20. Ballard J. W. O., Olsen G. J., Fauth D. P., Odgers W. A., Rowell D. M., Atkinson P. W. (1992). Evidence from 12S ribosomal RNA sequences that onychopterans are modified arthropods. *Science* 258:1345-1348.

21. Bandelt H. J., Dress A. (1986). Reconstructing the shape of a tree from observed dissimilarity data. *Adv. Appl. Math.* 7:309-343.
22. Beard K. C., Godinot M. (1988). Carpal anatomy of *Smilodectes gracilis* (Adapiformes, Notharctinae) and its significance for lemuriform phylogeny. *J. Hum. Evol.* 17:71-92.
23. Bender M. A., Chu E. H. Y. (1963). The chromosomes of Primates. In "Evolutionary and genetic biology of Primates." Buettner-Janush (Ed.). Academic Press. New York, I.
24. Bermingham E., Moritz C. (1998). Comparative phylogeography: concepts and applications. *Mol. Ecol.* 7:367-369.
25. 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. *Proc. Natl. Acad. Sci. USA* 89:6624-6628.
26. Bernardi G., Mouchiroud D., Gautier C. (1993). Silent substitutions in mammalian genomes and their evolutionary implications. *J. Mol. Evol.* 37:583-589.
27. Bishop M. J., Thompson E. A. (1986). Maximum likelihood alignment of DNA sequences. *J. Mol. Biol.* 190:159-165.
28. Bonner T. L., Heinemann R., Todaro G. J. (1980). Evolution of DNA sequences has been retarded in Malagasy primates. *Nature* 286:420-423.
29. Boore J. L., Collins T. M., Stanton D., Daehler L. L., Brown W. M. (1995). Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. *Nature* 376:163-165.
30. Bowen B. W., Nelson W. S., Avise J. C. (1993). A molecular phylogeny for marine turtles: trait mapping, rate assessment and conservation relevance. *Proc. Natl. Acad. Sci. USA* 90:5574-5577.
31. Bremer K. (1988). The limits of amino-acid sequences data in angiosperm phylogenetic reconstruction. *Evolution* 42:795-803.
32. Bremer K. (1994). Branch support and tree stability. *Cladistics* 10:295-304.
33. Britten R. J. (1986). Rates of DNA sequence evolution differ between taxonomic groups. *Science* 231:1393-1398.
34. Brown W. M., Prager E. M., Wang A., Wilson A. C. (1982). Mitochondrial DNA sequence of primates: tempo and mode of evolution. *J. Mol. Evol.* 18:225-239.
35. Bruno W. J., Halpern A. L. (1999). Topological bias and inconsistency of maximum likelihood using wrong models. *Mol. Biol. Evol.* 16:564-566.
36. Bryant H. N. (1995). Why autapomorphies should be removed. A reply to Yates. *Cladistics* 11:381-384.
37. Bryant H. N. (1996). Explicitness, stability and universality in the phylogenetic definition and usage of taxon names: a case study of the phylogenetic taxonomy of the Carnivora (Mammalia). *Syst. Biol.* 45:174-183.
38. Bull J. J., Huelsenbeck J. P., Cunningham C. W., Swofford D. L., Waddell P. (1993). Partitioning data, combining data in phylogenetic analysis. *Syst. Biol.* 42:384-397.
39. Burney D. A., MacPhee R. D. E. (1988). Mysterious island: what killed Madagascar's large native animals? *Natur. History* 97:46-55.
40. Cantatore P., Roberti M., Pesole G., Ludovico A., Mitella F., Gadaleta M. N., Saccone C. (1994). Evolutionary analysis of cytochrome b sequences in some perciformes: evidence for slower rate of evolution than in mammals. *J. Mol. Evol.* 39:589-597.
41. Cantion P. D., Olmstead R. G., Wagstaff S. J. (1997). A comparison of phylogenetic nomenclature with the current system: a botanical case study. *Syst. Biol.* 46:313-331.
42. Cao Y., Adachi J., Janke A., Pääbo S., Hasegawa M. (1994). Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of a tree based on a single gene. *J. Mol. Evol.* 39:519-527.

43. Cartmill M. (1975). Strepsirhine basicranial structures and the affinities of the Cheirogaleidae. In "Phylogeny of the Primates.". Lucken W. P., Szalay F. S. (Eds.). Plenum Press. New York, pp 313-354.
44. Catzefflis F. M., Aguilar J. P., Jaeger J. J. (1992). Muroid rodents: phylogeny and evolution. TREE 7:122-126.
45. Cavalli-Sforza L. L., Edwards A. W. F. (1967). Phylogenetic analysis: models and estimation procedures. Evolution 32:550-570.
46. Charles-Dominique P., Martin R. D. (1970). Evolution of lorises and lemurs. Nature 227:257-260.
47. Chippendale P. T., Wiens J. J. (1994). Weighting, partitioning and combining characters in phylogenetic analysis. Syst. Biol. 43:278-287.
48. Churchill G. A. (1989). Stochastic models for heterogeneous DNA sequences. Bull. Math. Biol. 51:79-94.
49. Cooper A., Fortey R. (1998). Evolutionary explosions and the phylogenetic fuse. Trends Ecol. Evol. 13:151-156.
50. Cooper A., Mourer-Chauviré C., Chambers G. K., Von Haeseler A., Wilson A. C., Pääbo S. (1992). Independent origin of New Zealand moas and kiwis. Proc. Natl. Acad. Sci. USA 89:8741-8744.
51. Corneli P. S., Ward R. H. (2000). Mitochondrial genes and Mammalian phylogenies: increasing the reliability of branch length estimation. Mol. Biol. Evol. 17(2): 224-234.
52. Cox D. R. (1961). Tests of separate families of hypotheses. 14 th. Berkeley Symposium, University of California Press.
53. Cox D. R. (1962). Further results on tests of separate families of hypotheses. J. R. Stat. Soc. B 24:406-424.
54. Cracraft J., Helm-Bychowski K. (1991). Parsimony and phylogenetic inference using DNA sequences: some methodological strategies. In "Phylogenetic analysis of DNA sequences". Miyamoto M. M., Cracraft J. (Eds.). Oxford University Press. Oxford.
55. Crandall K. A., Hillis D. M. (1997). Rhodopsin evolution in the duck. Nature 387:667-668.
56. Crovella S., Montagnon D., Rakotosamimanana B., Rumpler Y. (1994). Molecular biology and systematics of an extinct lemur: *Pachylemur insignis*. Primates 35:519-522.
57. Crovella S., Montagnon D., Rumpler Y. (1993). Highly repeated DNA analysis and systematics of the Lemuridae, a family of Malagasy prosimians. Primates 34:61-69.
58. Crovella S., Montagnon D., Rumpler Y. (1995). Highly repeated DNA sequences and systematics of malagasy Primates. Hum. Evol. 10(1): 35-44.
59. Cummings M. P., Otto S. P., Wakeley J. (1995). Sampling properties of DNA sequence data in phylogenetic analysis. Mol. Biol. Evol. 12(5): 814-822.
60. Da Silva M. N. F., Patton J. L. (1993). Amazonian phylogeography: mtDNA sequence variation in arborealechimyid rodents (Caviomorpha). Mol. Phylogenet. Evol. 2:243-255.
61. Darlu P., Tessy P. (1993). La reconstruction phylogénétique. Concepts et méthodes. Masson, Paris.
62. De Queiroz K., Gauthier J. (1992). Phylogenetic taxonomy. Annu. Rev. Ecol. Syst. 23:440-480.
63. De Queiroz K., Gauthier J. (1994). Toward a phylogenetic system of biological nomenclature. TREE 9(1): 27-31.
64. Debry R. W., Marzluff W. F. (1994). Selection on silent sites in the rodent H3 histone gene family. Genetics 138:191-202.
65. Dene H. T., Goodman M., Prychodko W., Moore G. W. (1976). Immunodiffusion systematics of the Primates: the Strepsirhini. Folia Primatol. 25:35-61.



66. Dene H. T., Goodman M., Prychodku W. (1976). Immunodiffusion evidence on the phylogeny of the Primates. In "Molecular anthropology". Goodman M. , Tashian R. E., Tashian J. K. (Eds.). Plenum Press. New York, pp 171-195.
67. Desjardin P., Morais M. (1990). Sequence and gene organization of the chicken mitochondrial genome. *J. Mol. Biol.* 212:243-256.
68. Dickerson R. E. (1971). The structure of cytochrome c and the rate of molecular evolution. *J. Mol. Evol.* 1:26-45.
69. Donoghue M. J., Cantino P. D. (1984). The logic and limitation of the outgroup substitution approach to cladistic analysis. *Syst. Bot.* 9:192-202.
70. Donoghue M. J., Olmstead R. G., Smith J. F., Palmer J. D. (1991). Phylogenetic relationships of *Dipsacales* based on *rboL* sequences. *Am. Missouri Bot. Gard.* 79:333-346.
71. Dress A., Von Haeseler A., Krüger M. (1986). Reconstructing phylogenetic trees using variants of the four-point condition. *Studien zur Klassifikation* 17:299-305.
72. Duret C., Mouchiroud D. (1999). Expression pattern, and surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila* and *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 96:4482-4487.
73. Dutrillaux B. (1979). Chromosomal evolution in Primates. Tentative phylogeny from *Microcebus murinus* (Prosimian, Primate) to man. *Hum. Genet.* 48:251-314.
74. Dutrillaux B. (1988). Chromosome evolution in Primates. *Folia Primatol.* 50:134-135.
75. Dutrillaux B., Couturier J., Muleris M., Rumpler Y., Viegas-Pequignot F. (1986). Relations chromosomiques entre sous-ordres et infra-ordres et schéma évolutif général des Primates. *Mammalia* 50:108-121.
76. Eaglen R. H. (1980). The systematics of living Strepsirhini, with special reference to the Lemuridae. Durham, NC, Duke University.
77. Eaglen R. H. (1982). Phenetic and cladistic approaches to strepsirhine systematics. *Am. J. Phys. Anthropol.* 71:185-201.
78. Eastal S., Collet C., Betty D. (1995). The mammalian molecular clock. Landes R.G., Austin, Tx.
79. Eck R. V., Dayhoff M. O. (1966). Atlas of protein sequence and structure. National Biomedical Research Foundation, Silver Springs, Md.
80. Edwards A. W. F. (1972). Likelihood. Cambridge University Press, Cambridge.
81. Edwards S. V., Arctander P., Wilson A. C. (1991). Mitochondrial resolution of a deep branch in the genealogical tree for perching birds. *Proc. R. Soc. Lond. B* 234:99-107.
82. Efron B. (1982). The jackknife, the bootstrap and other resampling plans. In "Regional Conference Series in Applied Mathematics.". CBMS-NSF (Ed.). SIAM. Philadelphia, Monography 38.
83. Elena S. F., Gonzalezcandelas F., Moya A. (1992). Does the Vp1 gene of foot-and-mouth-disease virus behave as a molecular clock. *J. Mol. Evol.* 35:223-229.
84. Farris J. S. (1982). Outgroups and parsimony. *Syst. Zool.* 33:83-103.
85. Farris J. S. (1983). The logical basis of phylogenetic analysis. In "Advances in Cladistics. Proceedings of the second meeting of the Willi Hennig Society". Platnick N. J., Funk V. A. (Eds.). Columbia University Press. New York, 2: pp 1-47.
86. Farris J. S. (1989). The retention index and rescaled consistency index. *Cladistics* 5:417-419.
87. Farris S. J. (1988). Hennig86: software and manual. Farris, S.J., Port Jefferson, N.Y.
88. Felsenstein J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27:401-410.
89. Felsenstein J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17:368-376.
90. Felsenstein J. (1984). Distance methods for inferring phylogenies: a justification. *Evolution* 38:16-24.

91. Felsenstein J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783-791.
92. Felsenstein J. (1988). Phylogenies from molecular sequences: inference and reliability. *Annu. Rev. Genet.* 22:521-565.
93. Felsenstein J., Churchill G. A. (1996). A hidden markov model approach to variation among sites in rate of evolution. *Mol. Biol. Evol.* 13(1): 93-104.
94. Field K. G., Olsen G. J., Lane D. J., Giovannoni S. J., Ghiselin M. T., Raff E. C., Pace N. R., Raff R. R. (1988). Molecular phylogeny of the animal kingdom. *Science* 239:748-753.
95. Fitch W. M. (1971). Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Zool.* 20:406-416.
96. Fitch W. M. (1976). Molecular evolutionary clocks. In "Molecular evolution.". Ayala F. J. (Ed.). Sinauer. Sunderland, Mass., pp 160-178.
97. Fitch W. M. (1981). A non-sequential method for constructing trees and hierarchical classifications. *J. Mol. Evol.* 18:30-37.
98. Fitch W. M., Margoliash E. (1967). A method for estimating the number of invariant amino acid coding positions in a gene using cytochrome c as a model case. *Biochem. Genet.* 1:65-71.
99. Fitch W. M., Markowitz E. (1970). An improved method for determining codon viability in a gene and its application to the rate of fixation of mutation in evolution. *Biochem. Genet.* 4:579-593.
100. Flacourt E. d. (1661). Histoire de la grande isle Madagascar. Avec une relation de ce qui s'est passé les années 1655, 1656 & 1657, non encor vué par la première impression. Oudot, N., L'Amy, P., Troyes, Paris.
101. Forsyth-Major C. I. (1894). On *Megaladapis madagascariensis*, an extinct gigantic Lemuroid from Ambolinatra. *Proc. R. Soc. Lond. B* 176-179.
102. Gascuel O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* 14(7): 685-695.
103. Gill T. (1872). Families of Mammals. Smithsonian Institute. Washington, pp 45.
104. Gillespie J. H. (1986). Rates of molecular evolution. *Annu. Rev. Ecol. Syst.* 17:637-665.
105. Gillespie J. H. (1989). Lineage effects and the index of dispersion of molecular evolution. *Mol. Biol. Evol.* 6:636-647.
106. Gillespie J. H. (1991). The course of molecular evolution. Oxford University Press, Oxford, England.
107. Gingerich P. D. (1990). African dawn for primates. *Nature* 346:411.
108. Gingerich P. D., Uhen M. D. (1994). Time of origin of primates. *J. Hum. Evol.* 27:443-445.
109. Gissi C., Reyes A., Pesole G., Saccone C. (2000). Lineage-specific evolutionary rate in Mammalian mtDNA. *Mol. Biol. Evol.* 17(7): 1022-1031.
110. Godfrey L. R., Jungers W. L., Reed K. E., Simons E. L., Chatrath P. S. (1997). Subfossil lemurs. In "Natural change and human impact in Madagascar". Goodman S. M., Patterson B. D. (Eds.). Smithsonian Institution Press. Washington, DC, pp 218-256.
111. Godfrey L. R., Simons E. L., Chatrath P., Rakotosamimanana B. (1990). A new fossil lemur (*Babakotia*, Primates) from northern Madagascar. *C.R. Acad. Sci. Paris* 310:81-87.
112. Gojobori T., Moriyama E. N., Kimura M. (1990). Statistical methods for estimating sequence divergence. In "Molecular evolution: computer analysis of protein and nucleic acid sequences.". Doolittle R. F. (Ed.). Academic Press. New York, pp 531-550.
113. Golding B., Felsenstein J. (1990). A maximum likelihood approach in the detection of selection from a phylogeny. *J. Mol. Evol.* 31:511-523.
114. Goldman N. (1990). Maximum likelihood influence on phylogenetic trees, with special reference to a Poisson process model of DNA substitution and to parsimony analyses. *Syst. Zool.* 39:345-361.

115. Goldman N., Yang Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* 11:725-736.
116. Goloboff P. A. (1991). Homoplasy and the choice among cladograms. *Cladistics* 7:215-232.
117. Goloboff P. A. (1997). Self-weighting optimization: tree searches and character state reconstruction under implied transformation costs. *Cladistics* 13:225-245.
118. Grandidier G. (1905). Recherche sur les lémuriens disparus et en particulier ceux qui vivaient à Madagascar. *Nouv. Arch. Mus. Natl. Hist. Nat. Paris* (4e. série) 8:1-140.
119. Grandidier G. (1929). Une variété du *Cheiromys madagascariensis* actuel et un nouveau *Cheiromys* subfossile. *Bull. Acad. malgache* 11:101-107.
120. Grantham R. (1974). Amino acid difference formula to help explain protein evolution. *Science* 185:862-864.
121. Graybeal A. (1993). The phylogenetic utility of cytochrome b: lesson from buffonid frog. *Mol. Phylogenet. Evol.* 2:256-269.
122. Graybeal A. (1998). Is it better to add taxa or characters to a difficult phylogenetic problem ? *Syst. Biol.* 47:9-17.
123. Groves C. P. (1974). Taxonomy and phylogeny of prosimians. In "Prosimian biology.". Martin R. D. , Doyle G. A., Walker A. C. (Eds.). Duckworth. London, pp 449-473.
124. Groves C. P., Eaglen R. H. (1988). Systematics of the Lemuridae (Primates, Strepsirhini). *J. Hum. Evol.* 17:513-538.
125. Groves C. P., Treuman J. W. H. (1995). Lemurid systematics revisited. *J. Hum. Evol.* 28:427-437.
126. Gu X. (1998). Early Metazoan divergence was about 830 million years ago. *J. Mol. Evol.* 47:369-371.
127. Harris S., Thackeray J. R., Jeffreys A. J., Weiss M. L. (1986). Nucleotide sequence analysis of the Lemur b-globin gene family: evidence for major rate fluctuations in globin polypeptide evolution. *Mol. Biol. Evol.* 3:465-484.
128. Hartigan J. A. (1973). Minimum evolution fits to a given tree. *Biometrics* 29:55-65.
129. Hasegawa M., Hashimoto T. (1993). Ribosomal RNA trees misleading ? *Nature* 361:23.
130. Hasegawa M., Hashimoto T., Otaha E., Adachi J., N. I., Miyata T. (1993). Early divergence in the evolution of eukaryotes: ancient divergence of *Eutamoeba* that lacks mitochondria revealed by protein sequence data. *J. Mol. Evol.* 36:380-388.
131. Hasegawa M., Kishino H., Yano T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22:160-174.
132. Hedges S. B., Parker P. H., Sibley C. G., Kumar S. (1996). Continental breakup and the ordinal diversification of birds and mammals. *Nature* 381:226-229.
133. Hendy M. D., Penny D. (1982). Branch and bound algorithms to determine minimal evolutionary trees. *Math. Biosci.* 59:1044-1053.
134. Hill W. C. O. (1953). *Primates: comparative anatomy and taxonomy. I- Strepsirhini.* Edinburgh University Press, Edinburgh.
135. Hillis D. M. (1996). Inferring complex phylogenies. *Nature* 383:130-131.
136. Hillis D. M. (1998). Taxonomic sampling, phylogenetic accuracy and investigator bias. *Syst. Biol.* 47(1): 3-8.
137. Hillis D. M., Allard M. W., Miyamoto M. M. (1993). Analysis of DNA sequence data: phylogenetic inference. *Methods Enzymol.* 224:456-487.
138. Hillis D. M., Mable B. K., Moritz C. (1996). Applications of molecular systematics: the state of the field and a look to the future. In "Molecular systematics.". Hillis D. M. , Moritz C., Mable B. K. (Eds.). Sinauer. Sunderland, MA, pp 515-543.
139. Holmes E. C., Worobey M., Rambaut A. (1999). Phylogenetic evidence for recombination in dengue virus. *Mol. Biol. Evol.* 16:405-409.

140. Honda D., Yokota A., Sogiyama J. (1999). Detection of seven major evolutionary lineages in Cyanobacteria based on 16S rRNA gene sequences analysis with new sequences of five marine *Synechococcus* strains. *J. Mol. Evol.* 48:723-739.
141. Honeycutt R. L., Nedbal M. A., Adkins R. M., Janecek L. L. (1995). Mammalian mitochondrial DNA evolution: a comparison of the cytochrome b and cytochrome oxidase II genes. *J. Mol. Evol.* 40:260-272.
142. Horai S., Hayasaka K., Kondo R., Tsugane K., Takahata N. (1995). Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc. Natl. Acad. Sci. USA* 92:532-536.
143. Horai S., Satta Y., Hayasaka K., Kondo R., Inoue T., Ishida T., Hayashi S., Takahata N. (1992). Man's place in hominoidea revealed by mitochondrial DNA genealogy. *J. Mol. Evol.* 35:32-43.
144. Huelsenbeck J. P. (1995). The performance of phylogenetic methods in simulation. *Syst. Biol.* 44:17-48.
145. Huelsenbeck J. P., Bull J. J., Cunningham C. W. (1996). Combining data in phylogenetic analysis. *Trends Ecol. Evol.* 11:152-158.
146. Huelsenbeck J. P., Large B., Swofford D. (2000). A compound Poisson process for relaxing the molecular clock. *Genetics* 154:1879-1892.
147. Huelsenbeck J. P., Rannala B. (1997). Phylogenetic methods come of age: Testing hypotheses in an evolutionary context. *Science* 276:227-232.
148. Hughes J. M., Baker A. J. (1999). Phylogenetic relationships of the enigmatic Hoatzin (*Ophistocomus hoazin*) resolved using mitochondrial and nuclear gene sequences. *Mol. Biol. Evol.* 16(9): 1300-1307.
149. Irwin D. M., Kocher T. D., Wilson A. C. (1991). Evolution of the cytochrome b gene of mammals. *J. Mol. Evol.* 32:128-144.
150. Jablonski N. G. (1986). A history of form and function in the primate masticatory apparatus from the ancestral primate through the strepsirrhines. In "Comparative Primate biology. Systematics, evolution and anatomy.". Swindler D. R., Erwin J. (Eds.). Alan R. Liss. New York, 1: pp 537-558.
151. Janczewski D. N., Yuhki N., Gilbert D. A., Jefferson G. T., O'Brien S. J. (1992). Molecular phylogenetic inference from saber-toothed cat fossils of Rancho La Brea. *Proc. Natl. Acad. Sci. USA* 89:9769-9773.
152. Janke A., Feldmaier-Fuchs G., Thomas W. K., Von Haeseler A., Pääbo S. (1994). The marsupial mitochondrial genome and the evolution of placental mammals. *Genetics* 137:243-256.
153. Jermann T. M., Opitz J. G., Stackhouse J., Benner S. A. (1995). Reconstructing the evolutionary history of the artyodactyl ribonuclease superfamily. *Nature* 374:57-59.
154. Jin L., Nei M. (1990). Limitations of the evolutionary parsimony method of phylogenetic analyses. *Mol. Biol. Evol.* 7:82-102.
155. Jouffroy F. K. (1962). La musculature des membres chez les Lémuriens de Madagascar: étude descriptive et comparative. *Mammalia* 26:1-326.
156. Jukes T. H. (1987). Transitions, transversions and the molecular clock. *J. Mol. Evol.* 26:87-98.
157. Jukes T. H., Cantor C. R. (1969). Evolution of protein molecules. In "Mammalian protein metabolism.". Munro H. N. (Ed.). Academic Press. New York, pp 21-132.
158. Jung K. Y., Crovella S., Rumpler Y. (1992). Phylogenetic relationships among lemuriform species determined from restriction genomic DNA banding patterns. *Folia Primatol.* 58:224-229.
159. Jungers W. L. (1980). Adaptive diversity in subfossil Malagasy prosimians. *Zeitschrift für Morphologie und Anthropologie* 71:177-186.

160. Jungers W. L., Godfrey L. R., Simons B. L., Chatrath P. S., Rakotosamimanana B. (1991). Phylogenetic and functional affinities of *Babakotia* (Primates), a fossil lemur from northern Madagascar. *Proc. Natl. Acad. Sci. USA* 88:9082-9086.
161. Källersjö M., Farris J. S., Kluge A. G., Bolt C. (1992). Skewness and permutation. *Cladistics* 8:275-287.
162. Kelsey C. R., Crandall K. A., Voevodin A. F. (1999). Different models, different trees: the geographic origin of PTLV-1. *Mol. Phylogenet. Evol.* 13:336-347.
163. Kendall M., Stuart A. (1979). *Advanced theory of statistics*. Charles Griffin, London.
164. Kim J. (1996). General inconsistency conditions for maximum parsimony: effects of branch lengths and increasing numbers of taxa. *Syst. Biol.* 45(3): 363-374.
165. Kim J. (1998). Large scale phylogenies and measuring performance of phylogenetic estimators. *Syst. Biol.* 47:43-60.
166. Kimura M. (1968). Evolutionary rate at the molecular level. *Nature* 217:624-626.
167. Kimura M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16:111-120.
168. Kimura M. (1983). *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge.
169. Kimura M. (1987). Molecular evolutionary clock and the neutral theory. *J. Mol. Evol.* 26:24-33.
170. Kishino H., Hasegawa M. (1989). Evaluation of maximum likelihood estimate of three evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* 29:170-179.
171. Kluge A. G., Farris J. S. (1969). Quantitative phyletics and the evolution of anurans. *Syst. Zool.* 18:1-32.
172. Kocher T. D., Thomas W. K., Meyer A., Edwards S. V., Pääbo S., Villablanca F. X., Wilson A. C. (1989). Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci. USA* 86:6196-6200.
173. Kocher T. D., Wilson A. C. (1991). Sequence evolution of mitochondrial DNA in humans and chimpanzees: control region and a protein-coding region. In "Evolution of life.". Osawa S., Honjo T. (Eds.). Springer Verlag, Berlin, pp 391-413.
174. Komiyama N. K., Miyazaki G., Tame J., Nagai K. (1995). Transplanting a unique allosteric effect from crocodile into human haemoglobin. *Nature* 373:244-246.
175. Konegay J. R., Schilling J. W., Wilson A. C. (1994). Molecular adaptation of a leaf eating bird: stomach lysozyme of the hoatzin. *Mol. Biol. Evol.* 11:921-928.
176. Koop B. F., Tagle D. A., Goodman M., Slightom J. L. (1989). A molecular view of primate phylogeny and important systematic and evolutionary questions. *Biol. Evol.* 6:580-612.
177. Kumada Y., Benson D. R., Hillemann D., Hosted T. J., Rochefort D. A., Thompson C. J., Wohlleben W., Tateno Y. (1993). Evolution of the glutamine synthetase gene, one of the oldest existing and functioning genes. *Proc. Natl. Acad. Sci. USA* 90:3009-3013.
178. Lake J. A. (1990). Origin of the Metazoa. *Proc. Natl. Acad. Sci. USA* 87:736-766.
179. Lambertson C. (1934). Contribution à la connaissance de la faune subfossile de Madagascar. Lémuriens et Ratites, *Archeoindris*, *Chiromys*, *Megaladapis*. *Mém. Acad. Malgache* XVII:1-168.
180. Lanave C., Preparata G., Saccone C., Serio G. (1984). A new method for calculating evolutionary substitution rates. *J. Mol. Evol.* 20:86-93.
181. Lee M. S. Y. (1998). Ancestors and taxonomy. *Trends Ecol. Evol.* 13:26.
182. Lee M. S. Y. (1999). Molecular clock calibrations and metazoan divergence dates. *J. Mol. Evol.* 49:385-391.
183. Lee Y. H., Vacquier V. D. (1992). The divergence of species-specific abalone sperm lysins is promoted by positive Darwinian selection. *Biol. Bull.* 182:97-104.

184. Leipe D. D., Gunderson J. H., Nerad T. A., Sogin M. L. (1993). Small subunit ribosomal RNAs of *Hexamita inflata* and the quest for the first branch in the eukaryotic tree. *Mol. Biochem. Parasitol.* 59:41-48.
185. Li P., Bousquet J. (1992). Relative-rate test for nucleotide substitution between two lineages. *Mol. Biol. Evol.* 9:1185-1189.
186. Li W. H., Gouy M., Sharp P. M., O'hUigin C., Yang Y. W. (1990). Molecular phylogeny of rodentia, lagomorpha, primates, artidactyla and carnivora and molecular clock. *Proc. Natl. Acad. Sci. USA* 87:6703-6707.
187. Li W. H., Graur D. (1991). *Fundamentals of molecular evolution.* Sinauer, Sunderland, MA.
188. Li W. H., Luo C. C., Wu C. I. (1985). Evolution of DNA sequences. In "Molecular evolutionary genetics.". MacIntyre R. J. (Ed.). Plenum Press. New York, pp 1.
189. Li W. H., Tanimura M. (1987). The molecular clock runs more slowly in man than in apes and monkeys. *Nature* 326:93-96.
190. Li W. H., Wu C. I., Luo C. C. (1985). A new method for estimating synonymous and nonsynonymous rates of nucleic substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* 2:150-174.
191. Lockhart P., Steel M., Hendy M., Penny D. (1994). Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* 11:605-612.
192. Lockhart P. J., Howe C. J., Bryant D. A., Beanland T. J., Larkum W. D. (1992). Substitutional bias confounds inference of organelle origins from sequence data. *J. Mol. Evol.* 34:153-162.
193. Loomis W. F., Smith D. W. (1990). Molecular phylogeny of *Dictyostelium discoideum* by protein sequence comparison. *Proc. Natl. Acad. Sci. USA* 87:9093-9097.
194. Lyons-Weiler J., Hoelzer G. A., Tausch R. J. (1998). Optimal outgroup analysis. *Biol. J. Linn. Soc.* 64:493-511.
195. Lyons-Weiler J., Takahashi K. (1999). Branch length heterogeneity lead to nonindependent branch length estimates and can decrease the efficiency of methods of phylogenetic inference. *J. Mol. Evol.* 49:392-405.
196. Ma D. P., Zharkikh A., Graur D., Vandenberg J. L., Li W. H. (1993). Structure and evolution of opossum, guinea pig and procupine cytochrome b genes. *J. Mol. Evol.* 36:327-334.
197. Mac Donald J., Kreitman M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652-654.
198. Macedonia J. M., Stanger K. F. (1994). Phylogeny of the Lemuridae revisited: evidence from communication signals. *Folia Primatol.* 63:1-43.
199. MacPhee R. D. E., Jacobs L. L. (1986). *Nycticeboides simpsoni* and the morphology, adaptation and relationships of Miocene Siwalik Lorisidae. *Contributions to Geology, University of Wyoming, Special Paper* 3:131-161.
200. MacPhee R. D. E., Simons E. L., Wells N. A., Vuillaume-Randriamanantena M. (1984). Team finds giant lemur skeleton. *Geotimes* 29:10-11.
201. MacPhee R. D. F., Raholimavo E. M. (1988). Modified subfossil aye-aye incisors from southwestern Madagascar. Species allocation and paleoecological significance. *Folia Primatol.* 51:126-142.
202. Maddison W. P. (1991). Squared-charge parsimony reconstruction of ancestral states for continuous valued characters on a phylogenetic tree. *Syst. Zool.* 40:304-314.
203. Maddison W. P., Donoghue M. J., Maddison D. R. (1984). Outgroup analysis and parsimony. *Syst. Zool.* 33:83-103.
204. Maddison W. P., Maddison D. R. (1992). *MacClade: Analysis of phylogeny and character evolution.* Sunderland, MA, Sinauer Associates.

205. Major C., Forsyth I. (1896). Preliminary notice on fossil monkeys from Madagascar. *Geological Magazine* 3:433-436.
206. Malcolm B. A., Wilson K. P., Matthews B. W., Kirsch J. F., Wilson A. C. (1990). Ancestral lysozymes reconstructed, neutrality tested, and thermostability linked to hydrocarbon packing. *Nature* 345:86-89.
207. Margoliash E. (1963). Primary structure and evolution of cytochrome c. *Proc. Natl. Acad. Sci. USA* 50:672-679.
208. Margush T., MacMorris F. R. (1981). Consensus n-trees. *Bull. Math. Biol.* 43:239-244.
209. Marko P. B., Vermeij C. J. (1999). Molecular phylogenetics and the evolution of labral spines among Easter Pacific ocenebrine gastropods. *Mol. Phylogenet. Evol.* 13:275-288.
210. Martin A. P., Naylor G. J. P., Palumbi S. R. (1992). Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature* 357:153-155.
211. Martin R. D. (1993). Primate origins: plugging the gaps. *Nature* 363:223-234.
212. Meireles C. M., Czelusniak J., Schneider M. P. C., Muniz J. A. P. C., Brigido M. C., Fesseira H. S., Goodman M. (1999). Molecular phylogeny of Ateline New World monkeys (*Platyrrhini Atelinae*) based on g-globin gene sequences: evidence that *Brachyteles* is the sister group of *Lagothrix*. *Mol. Phylogenet. Evol.* 12:10-30.
213. Messier W., Stewart C. B. (1997). Episodic adaptive evolution of primate lysozymes. *Nature* 385:151-154.
214. Meyer A. (1994). Shortcomings of the cytochrome b gene as a molecular marker. *TREE* 9:378-380.
215. Meyer A., Dolven S. I. (1992). Molecules, fossils and the origin of tetrapods. *J. Mol. Evol.* 35:102-113.
216. Meyer A., Wilson A. C. (1990). Origin of tetrapods inferred from their mitochondrial DNA affiliation to lungfish. *J. Mol. Evol.* 31:359-364.
217. Milinkovitch M. C., Orti G., Meyer A. (1993). Revised phylogeny of whales suggested by mitochondrial DNA sequences. *Nature* 361:346-348.
218. Mindell D. P., Dick C. W., Baker R. J. (1991). Phylogenetic relationships among megabats, microbats and Primates. *Proc. Natl. Acad. Sci. USA* 88:10322-10326.
219. Mindell D. P., Tacker C. E. (1996). Rates of molecular evolution: phylogenetic issues and applications. *Annu. Rev. Syst.* 27:279-303.
220. Miyamoto M. M., Allard M. W., Adkins R. M., Janecek L. L., Honeycutt R. L. (1994). A convergence test of reliability using linked mitochondrial DNA sequences. *Syst. Biol.* 43:236-249.
221. Miyamoto M. M., Cracraft J. (1991). *Phylogenetic analysis of DNA sequences*. Oxford University Press, New York.
222. Miyamoto M. M., Kraus F., Ryder G. A. (1990). Phylogeny and evolution of antlered deer determined from mitochondrial DNA sequences. *Proc. Natl. Acad. Sci. USA* 87:6127-6131.
223. Miyata T., Yasanaga T., Nishida T. (1980). Nucleotide sequence divergence and functional constraint in mRNA evolution. *Proc. Natl. Acad. Sci. USA* 77:7328-7332.
224. Miyata T., Yasunaga T. (1980). Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *J. Mol. Evol.* 16:23-36.
225. Montagnon D., Ravaoarimanana B., Rakotosamimanana B., Rumpler Y. (2001). Ancient DNA from *Megaladapis edwardsi* (Malagasy subfossil): preliminary results using partial cytochrome b sequence. *Folia Primatol.* 72:30-32.
226. Montagnon D., Ravaoarimanana B., Rumpler Y. (2001). Taxonomic relationships and sampling effects among Lepilemuridae and Lemuridae using partial cytochrome b gene. *C. R. Acad. Sci. Paris / Life Sci.* 334(7): 647-656.

227. Mooi R. (1989). The outgroup criterion revisited via naked zones and alleles. *Syst. Zool.* 38:283-290.
228. Moritz C., Dowling T. E., Brown W. M. (1987). Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu. Rev. Ecol. Syst.* 18:269-292.
229. Mouchiroud D., Gautier C., Bernardi G. (1995). Frequencies of synonymous substitutions in mammals are gene-specific and correlated with frequencies of non-synonymous substitutions. *J. Mol. Evol.* 40:107-113.
230. Muse S. V. (1995). Evolutionary analyses of DNA sequences subject to constraints of secondary structure. *Genetics* 139:1429-1439.
231. Muse S. V. (1996). Estimating synonymous and nonsynonymous substitution rates. *Mol. Biol. Evol.* 13(1): 105-114.
232. Muse S. V., Gaut B. S. (1994). A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to chloroplast genome. *Mol. Biol. Evol.* 11:715-724.
233. Muse S. V., Weir B. S. (1992). Testing for equality of evolutionary rates. *Genetics* 132:269-276.
234. Naylor G. J. P. (1997). Structural biology and phylogenetic estimation. *Nature* 388:527-528.
235. Naylor G. J. P., Collins T. M., Brown W. M. (1995). Hydrophobicity and phylogeny. *Nature* 373:565-566.
236. Needleman S. B., Wunsch C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48:443-453.
237. Nei M. (1986). Stochastic errors in DNA evolution and molecular phylogeny. In "Evolutionary perspectives and the new genetics.". Gershowitz H. , Rucknagel D. L., Tashian R. F. (Eds.). Alan R. Liss. New York, pp 133-147.
238. Nei M. (1987). *Molecular evolutionary genetics*. Columbia University Press, New York.
239. Nei m. (1991). Relative efficiencies of different tree making methods for molecular data. In "Phylogenetic studies of DNA sequences.". Miyamoto M. M., Cracraft J. L. (Eds.). Oxford University Press. Oxford, pp 90-128.
240. Nei M. (1996). Phylogenetic analysis in molecular evolutionary genetics. *Annu. Rev. Genet.* 30:371-403.
241. Nei M., Chakraborty R., Fuerst P. A. (1976). Infinite allele model with varying mutation rate. *Proc. Natl. Acad. Sci. USA* 73:4164-4168.
242. Nei M., Gojobori T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3:418-426.
243. Nei M., Tajima F. (1981). DNA polymorphism detectable by restriction endonucleases. *Genetics* 97:145-163.
244. Neigel J. E., Avise A. C. (1986). Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. In "Evolutionary processes and theory.". Warlin S., Nevo E. (Eds.). Academic Press. New York, pp 515-534.
245. Nielsen R., Yang Z. (1998). Likelihood models for detecting positively selected amino acid sites and application on the HIV-1 envelope gene. *Genetics* 148:929-936.
246. Nixon K. C. (1999). The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* 15:407-414.
247. Ohta T. (1995). Synonymous and non synonymous substitutions in mammalian genes and the nearly neutral theory. *J. Mol. Evol.* 40:56-63.
248. O'hUigin C., Li W. H. (1992). The molecular clock ticks regularly in muroid rodents and hamsters. *J. Mol. Evol.* 35:377-384.
249. Olsen G. (1987). Earliest phylogenetic branching: comparing rRNA-based evolutionary trees inferred with various techniques. *Cold Spring Harb. Symp. Quant. Biol.* LII:825-837.



250. Olsen G. J., Natsuda H., Hagström R., Overbeer R. (1994). FastDNAML: a tool for reconstruction of phylogenetic trees of DNA sequences using maximum likelihood. *CABIOS* 10:41-48.
251. Osman-Hill W. C. (1953). *Primates: comparative anatomy and taxonomy*. Edinburgh University Press, Edinburgh.
252. Pagel M. (1997). Inferring evolutionary processes from phylogenies. *Zool. Scripta* 26:331-348.
253. Palumbi S. R. (1989). Rates of molecular evolution and the fraction of nucleotide positions free to vary. *J. Mol. Evol.* 29:180-187.
254. Pamilo P., Nei M. (1988). Relationships between gene trees and species trees. *Mol. Biol. Evol.* 5:568-583.
255. Pastorini J. (2000). *Molecular systematics of Lemurs*. Zürich, University of Zürich: 182.
256. Patriat P., Achache J. (1984). India-Eurasia collision chronology has implications for crustal shortening and driving mechanisms of plates. *Nature* 311:615-621.
257. Patton J. L., Da Silva M. N. F., Malcolm J. R. (1994). Gene genealogy and differentiation among arboreal spiny rats (Rodentia: Echimyidae) of the Amazon Basin: a test of the riverine barrier hypothesis. *Evolution* 48:1314-1323.
258. Pauling L., Zuckerkandl E. (1963). Chemical paleogenetics: molecular 'restoration studies' of extinct form of life. *Acta Chem. Scand.* 17:S9-S16.
259. Penny D., Hendy M. D. (1985). The use of tree comparison metrics. *Syst. Zool.* 34:75-82.
260. Perna K., Kocher T. (1995). Unequal base frequencies and the estimation of substitution rates. *Mol. Biol. Evol.* 12:359-361.
261. Petter J. J., Albignac R., Rumpler Y. (1977). *Faune de Madagascar*. ORSTOM, Paris.
262. Pocock R. I. (1918). On the external characters of the lemurs and of *Tarsius*. *Proc. Zool. Soc. Lond.* 19-53.
263. Poe S. (1998). Sensitivity of phylogeny estimator to taxonomic sampling. *Syst. Biol.* 47:18-31.
264. Polloch D. D., Goldstein D. B. (1995). A comparison of two methods for constructing evolutionary distances from a weighted contribution of transition and transversion differences. *Mol. Biol. Evol.* 12:713-717.
265. Porter C. A., Sampaio I., Schneider H., Schneider M. P. C., Czelusniak J., Goodman M. (1995). Evidence on primate phylogeny from e-globin gene sequences and flanking regions. *J. Mol. Evol.* 40:30-55.
266. Purvis A., Bromham L. (1997). Estimating the transition / transversion ratio from independent pairwise comparison with an assumed phylogeny. *J. Mol. Evol.* 44:112-119.
267. Rabinowitz P. D., Coffin M. F., Falvey D. (1983). The separation of Madagascar and Africa. *Science* 220:67-69.
268. Rambaut A., Bromham L. (1998). Estimating divergence dates from molecular sequences. *Mol. Biol. Evol.* 15:442-448.
269. Ranala B., Huelsenbeck J., Yang Z., Nielsen R. (1998). Taxon sampling and the accuracy of large phylogenies. *Syst. Biol.* 47:702-710.
270. Rassmann K. (1997). Evolutionary age of the Galapagos iguanas predates the age of the present Galapagos islands. *Mol. Phylogenet. Evol.* 7:158-172.
271. Rice J. A. (1995). *Mathematical statistics and data analysis*. Duxbury, Belmont CA.
272. Rice K. A., Donoghue M. J., Olmstead R. G. (1997). Analyzing large data sets: *rbc* I 500 revisited. *Syst. Biol.* 46:554-563.
273. Rumpler Y., Albignac R. (1970). Evolution chromosomique des lémuriens Malgaches. *Annales de l'Université de Madagascar. Médecine et Biologie* 12-13.

274. Rumpler Y., Warter S., Petter J. J., Albignac R., Dutrillaux B. (1988). Chromosomal evolution of Malagasy lemurs. XI. Phylogenetic position of *Daubentonia*. *Folia Primatol.* 50:124.
275. Rzhetsky A., Kumar S., Nei M. (1995). Four-cluster analysis: a simple method to test phylogenetic hypotheses. *Mol. Biol. Evol.* 12:163-167.
276. Rzhetsky A., Nei M. (1993). Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Mol. Biol. Evol.* 10:1073-1095.
277. Saccone C., Lanave C., Pesole G., Preparata G. (1990). Influence of base composition on quantitative estimation of gene evolution. In "Methods in enzymology. Molecular evolution. Computer analysis of protein and nucleic acid sequences." Doolittle R. F. (Ed.). Academic Press. New York, 183: pp 570-598.
278. Saitou N., Imanishi M. (1989). Relative efficiency of the Fitch-Margoliash, maximum parsimony, maximum likelihood, minimum evolution and neighbor-joining methods of phylogenetic reconstructions in obtaining the correct tree. *Mol. Biol. Evol.* 6:514-525.
279. Saitou N., Nei M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406-425.
280. Sambrook J., Fritsch E., Maniatis T. (1990). *Molecular cloning: a laboratory manual* (2nd. edition). Cold Spring Harbor University Park Press. Cold Spring Harbor,.
281. Sanderson M. J. (1997). A nonparametric approach to estimate divergence times in the absence of rate constancy. *Mol. Biol. Evol.* 14:1218-1232.
282. Sanderson M. J., Donoghue M. J. (1989). Patterns and variation in levels of homoplasy. *Evolution* 43:1781-1795.
283. Sankoff D. (1975). Minimal mutation tree of sequences. *SIAM J. Appl. Math.* 28:35-42.
284. Sattath S., Tversky A. (1997). Additive similarity trees. *Psychometrika* 42:406-425.
285. Schandler C., Thollesson M. (1995). Phylogenetic taxonomy. Some comments. *Zool. Scripta* 24:263-267.
286. Schöninger M., Von Haeseler A. (1993). A simple method to improve the reliability of tree reconstruction. *Mol. Biol. Evol.* 10:471-483.
287. Schöninger M., Von Haeseler A. (1995). Performance of the maximum likelihood, neighbor-joining, and maximum parsimony methods when sequence sites are not independent. *Syst. Biol.* 44:533-547.
288. Schuler D. (1995). Uncertainty in ancient phylogenies. *Nature* 377:108-109.
289. Schuler D., Preece T., Mooers A., Ludwig D. (1997). Likelihood of ancestral states in adaptive radiation. *Evolution* 51:1699-1711.
290. Schwartz E. (1931). A revision of the genera and species of Madagascar Lemuridae. *Proc. Zool. Soc. Lond.* 399-428.
291. Schwartz J. H., Tattersall I. (1985). Evolutionary relationships of living lemurs and lorises (Mammalia, Primates) and their potential affinities with European Eocene Adapidae. *Anthrop. Papers Amer. Mus. Nat. Hist.* 60:1-100.
292. Schwartz J. H., Tattersall I., Eldredge N. (1978). Phylogeny and classification of the primates revisited. *Yearbook of Phys. Anthropol.* 21:95-113.
293. Seddon J. M., Baverstock P. R., Georges A. (1998). The rate of mitochondrial 12S rRNA gene evolution is similar in freshwater turtles and marsupials. *J. Mol. Evol.* 46:460-464.
294. Seligsohn D., Szalay F. S. (1974). Dental occlusion and the masticatory apparatus in *Lemur* and *Varecia*: their bearing on the systematics of living and fossil primates. In "Prosimian biology." Martin R. D., Doyle G. A., Walker A. C. (Eds.). Duckworth. London.
295. Sharkey M. J. (1993). Exact indices, criteria to select from minimum length tree. *Cladistics* 9:211-222.
296. Sharp P. M. (1997). In search of molecular darwinism. *Nature* 385:111-112.

297. Sharp P. M., Cowe E., Higgins D. G., Shields D. C., Wolfe K. H., Wright F. (1988). Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Saccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*: a review of the considerable within-species diversity. *Nucl. Acid Res.* 16:8207-8211.
298. Sheldon F. H., Jones C. E., Mac Cracken K. G. (2000). Relative patterns and rates of evolution in Heron nuclear and mitochondrial DNA. *Mol. Biol. Evol.* 17(3): 437-450.
299. Simons E. L. (1994). The giant aye-aye *Daubentonia robusta*. *Folia Primatol.* 62:14-21.
300. Simons E. L., Godfrey L. R., Jungers W. L., Chatrath P. S., Rakotosamimanana B. (1992). A new giant subfossil lemur *Babakotia*, and the evolution of the sloth lemurs. *Folia Primatol.* 58:197-203.
301. Simons E. L., Godfrey L. R., Jungers W. L., Chatrath P. S., Ravaoarisoa J. (1995). A new species of *Mesopropithecus* (Primates, Palaeopropithecus) from northern Madagascar. *Int. J. Primatol.* 16:653-682.
302. Simpson G. G. (1945). The principles of classification and classification of mammals. *Bull. Am. Mus. Nat. Hist.* 85:1-350.
303. Simpson G. G. (1978). Early mammals in South America: fact, controversy and mystery. *Proc. Am. Philos. Soc.* 122:318-328.
304. Sitnikova T. L., Rzhetsky A., Nei M. (1994). Interior-branch and the bootstrap test of phylogenetic trees. *Mol. Biol. Evol.* 12:319-333.
305. Smith A. B. (1994). Rooting molecular trees: problems and strategies. *Biol. J. Linn. Soc.* 51:279-292.
306. Smith M. F., Patton J. L. (1993). The diversification of South American murid rodents: evidence from mitochondrial DNA sequence data for the akodontine tribe. *Biol. J. Linn. Soc.* 50:149-177.
307. Smith M. J., Arndt A., Gorski S., Farber E. (1993). The phylogeny of echinoderm classes based on mitochondrial gene arrangements. *J. Mol. Evol.* 36:545-554.
308. Sober E. (1988). *Reconstructing the past*. MIT Press, Cambridge, Massachusetts.
309. Sogin M. L. (1991). Early evolution and the origin of the eukaryotes. *Curr. Opi. Genet. Dev.* 1:457-463.
310. Sogin M. L., Gunderson J. H., Elwood H. J., Alonso R. A., Peattie D. A. (1989). Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. *Science* 243:75-77.
311. Soltis D. E., Soltis P. S., Mort M. E., Chase M. W., Savolainen V., Hoot S. B., Morton C. M. (1998). Inferring complex phylogenies using parsimony: an empirical approach using three large DNA data sets from Angiosperms. *Syst. Biol.* 47(1): 32-42.
312. Sonnerat M. (1782). *Voyage aux Indes orientales et à la Chine*. Fraulé, Paris.
313. Sourdis J., Krimbas C. (1987). Accuracy of phylogenetic trees estimated from DNA sequences data. *Mol. Biol. Evol.* 4:159-166.
314. Standing H. F. (1908). On recently discovered subfossil primates from Madagascar. *Trans. Zool. Soc. Lond.* 18:69-162.
315. Stanger-Hall K. F. (1997). Phylogenetic affinities among the extant Malagasy lemurs (Lemuriformes) based on morphology and behaviour. *J. Mammal. Evol.* 4:63-194.
316. Stanger-Hall K. F., Cunningham C. W. (1998). Support for the monophyletic lemuriformes: overcoming incongruence between data partitions. *Mol. Biol. Evol.* 15:1572-1577.
317. Steel M. A., Lockhart P. J., Penny D. (1993). Confidence in evolutionary trees from biological sequence data. *Nature* 364:440-442.
318. Stewart C. B., Schilling J. W., Wilson A. C. (1987). Adaptive evolution in the stomach lysozymes of foregut fermenters. *Nature* 330:401-404.
319. Strimmer K., Von Haeseler A. (1996). Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* 13(7): 964-969.

320. Studier J. A., Keppler K. L. (1988). A note on the neighbor-joining algorithm of Saitou and Nei. *Mol. Biol. Evol.* 5:729-731.
321. Sullivan D. T. (1995). DNA excision repair and transcription implications for genome evolution. *Curr. Opin. Genet. Dev.* 5:786-791.
322. Swanson K. W., Irwin D. M., Wilson A. C. (1991). Stomach lysozyme gene and the langur monkey: tests for congruence and positive selection. *J. Mol. Evol.* 33:418-425.
323. Swofford D. L. (1991). When are phylogeny estimations from molecular and morphological data incongruent ? In "Phylogenetic analysis of DNA sequences.". Miyamoto M. M., Cracraft J. (Eds.). Oxford University Press. New York, pp 294-333.
324. Swofford D. L. (1993). *Phylogenetic Analysis Using Parsimony (PAUP)*. Champaign, University of Illinois.
325. Swofford D. L., Begle D. P. (1993). *PAUP: phylogenetic analyses using parsimony*. Washington DC, Laboratory of Molecular Systematics, Smithsonian Institution.
326. Swofford D. L., Olsen G. J., Waddell P. J., Hillis D. M. (1996). Phylogenetic inference. In "Molecular systematics.". Hillis D. M. , Moritz C., Mable B. K. (Eds.). Sinauer. Sunderland, Mass., pp 407-514.
327. Szalay F. S., Katz C. C. (1973). Phylogeny of lemurs, galagos and lorises. *Folia Primatol.* 19:88-103.
328. Tajima F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437-460.
329. Tajima F. (1993). Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* 135:599-607.
330. Tajima F. (1993). Unbiased estimation of evolutionary distance between nucleotide sequences. *Mol. Biol. Evol.* 10:677-688.
331. Tajima F., Nei M. (1984). Estimation of evolutionary distances between nucleotide sequences. *Mol. Biol. Evol.* 1:269-285.
332. Tajima F., Takezaki N. (1994). Estimation of evolutionary distances for reconstructing molecular phylogenetic trees. *Mol. Biol. Evol.* 11(2): 278-286.
333. Takahata N. (1991). Overdispersed molecular clock at the major histocompatibility complex loci. *Phil. Trans. Roy. Soc. Lond. (Serie B)* 243:13-18.
334. Takezaki N., Gojobori T. (1999). Correct and incorrect vertebrate phylogenies obtained by the entire mitochondrial DNA sequence. *Mol. Biol. Evol.* 16:590-601.
335. Takezaki N., Razhetsky A., Nei M. (1995). Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* 12:823-833.
336. Tamura K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases. *Mol. Biol. Evol.* 9:678-698.
337. Tamura K., Nei M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in human and chimpanzees. *Mol. Biol. Evol.* 10:512-528.
338. Tan J. H., Blomster J., Hansen G., Leskinen E., Meggs C. A., Mann D. E., Sluiman H. J., Stanhope M. J. (1999). Molecular phylogenetic evidence for a reversible morphogenetic switch controlling the gross morphology of the common genera of green seaweeds: *Ulva* and *Enteromorpha*. *Mol. Biol. Evol.* 16:1011-1018.
339. Tateno Y., Nei M., Tajima F. (1982). Accuracy of estimated phylogenetic trees from molecular clock. I. Distantly related species. *J. Mol. Evol.* 18:387-404.
340. Tateno Y., Takezaki N., Nei M. (1994). Relative efficiencies of maximum likelihood, neighbor-joining and maximum parsimony methods when substitution rate varies with site. *Mol. Biol. Evol.* 11:261-277.
341. Tattersall I. (1982). *The Primates of Madagascar*. Columbia University Press, New York.

342. Tattersall I., Schwartz J. (1991). Phylogeny and nomenclature in the 'Lemur-group' of Malagasy Strepsirrhine primates. *Anthrop. Papers Amer. Mus. Nat. Hist.* 69:3-18.
343. Tattersall I., Schwartz J. H. (1974). Craniodental morphology and the systematics of the Malagasy lemurs (Primates, Prosimi). *Anthrop. Papers Amer. Mus. Nat. Hist.* 52:141-192.
344. Tautz D., Renz M. (1983). An optimized freeze-squeeze method for the recovery of DNA fragments from agarose gels. *Analytical Biochemistry* 132:14-19.
345. Tavaré S. (1986). Some probabilistic and statistical problems on the analysis of DNA sequences. In "Lectures in Mathematics in the Life Sciences." 17: pp 57-86.
346. Thomas W. K., Martin S. L. (1993). A recent origin of marmots. *Mol. Phylogenet. Evol.* 2:330-336.
347. Thompson E. A. (1975). *Human evolutionary trees*. Cambridge University Press, Cambridge.
348. Thorne J. L., Goldman N., Jones D. T. (1996). Combining protein evolution and secondary structure. *Mol. Biol. Evol.* 13:666-673.
349. Thorne J. L., Kishino H., Felsenstein J. (1991). An evolutionary model for maximum likelihood alignment of DNA sequences. *J. Mol. Evol.* 33:114-124.
350. Thorne J. L., Kishino H., Felsenstein J. (1992). Finding toward reliability: an improved likelihood model of sequence evolution. *J. Mol. Evol.* 34:3-16.
351. Thorne J. L., Kishino H., Pointer I. S. (1998). Estimating the rate of molecular evolution. *Mol. Biol. Evol.* 15:1647-1657.
352. Uyenoyama M. K. (1995). A generalized least-squares estimate for the origin of sporophytic self-incompatibility. *Genetics* 139:975-992.
353. Uzzell T., Corbin K. W. (1971). Fitting discrete probability distribution to evolutionary events. *Science* 172:1089-1096.
354. Vallois H. (1955). *Ordre des Primates, sous ordre des Lemuroïdes*. In "Traité de zoologie, anatomie, systématique, biologie." Grassé P. P. (Ed.). Masson. Paris, pp 1860-1895.
355. Voelker G. (1999). Dispersal, vicariance, and clocks: historical biogeography and speciation in cosmopolitan passerine genus (*Anthus*: Motacillidae). *Evolution* 53:1536-1552.
356. Vuillaume-Randriamanantena M. (1988). The taxonomic attributions of giant sub-fossil bones from Ampasambazimba: *Archaeoindris* and *Lemuridoherium*. *J. Hum. Evol.* 17:379-381.
357. Wainright P. O., Hinkle G., Sogin M. L., Stickel S. K. (1993). Monophyletic origin of the metazoa: an evolutionary link with the fungi. *Science* 260:340-342.
358. Wakeley J. (1993). Substitution rate variation among sites in hypervariable region I of human mitochondrial DNA. *J. Mol. Evol.* 37:613-623.
359. Wakeley J. (1994). Substitution rate variation among sites and the estimation of transition bias. *Mol. Biol. Evol.* 11:436-442.
360. Wakeley J. (1996). The excess of transitions among nucleotide substitutions. New methods of estimating transition bias underscore its significance. *TREE* 11:158-163.
361. Walker A. C. (1974). Locomotor adaptations in past and present prosimian primates. In "Primate locomotion." Jenkins F. A. (Ed.). Academic Press. New York, pp 349-381.
362. Warren R. D., Crampton R. H. (1996). Lazy leapers: energetics, phylogenetic inertia and the locomotor differentiation of the Malagasy Primates. In "Biogeography of Madagascar". Lourenço W. R. (Ed.). ORSTOM. Paris, pp 259-265.
363. Waterman M. S., Eggert M., Lander E. (1992). Parametric sequence comparisons. *Proc. Natl. Acad. Sci. USA* 89:
364. Watrous L. E., Wheeler Q. D. (1981). The outgroup comparison method of character analysis. *Syst. Zool.* 30:1-11.

365. Williams P., Fitch W. (1990). Phylogeny determination using dynamically weighted parsimony method. In "Molecular evolution: computer analysis of protein and nucleic acid sequences". Doolittle R. F. (Ed.). 13: pp 615-626.
366. Willis C. (1995). When did Eve dive ? An evolutionary detective story. *Evolution* 49:593-607.
367. Wilson A. C., Cann R. L., Carr S. M., George M., Gyllenstein U. B., Helm-Bychowski K. M., Higuchi R. G., Palumbi S. R., Prager E. M., Sage R. D., Stoneling M. (1985). Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol. J. Linn. Soc.* 26:375-400.
368. Wilson S. J. (1999). Building phylogenetic trees from quartets by using local inconsistency measures. *Mol. Biol. Evol.* 16(5): 685-693.
369. Woese C., Achenbach L., Rouviere P., Mondelco L. (1991). Archaeal phylogeny: reexamination of the phylogenetic position of *Archaeoglobus fulgidis* in light of certain composition-induced artifacts. *Syst. Appl. Microbiol.* 14:364-371.
370. Wolfe K. H., Sharp P. M. (1993). Mammalian gene evolution. Nucleotide sequence divergence between mouse and rat. *J. Mol. Evol.* 37:441-456.
371. Wolfe K. H., Sharp P. M., Li W. H. (1989). Mutation rates differ among regions of the mammalian genome. *Nature* 337:283-288.
372. Wu C. I., Li W. H. (1985). Evidence for higher rate of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. USA* 82:1741-1745.
373. Wyss A. R., Meng J. (1996). Application of phylogenetic taxonomy to poorly resolved crown clades: a stem-modified node-based definition of Rodentia. *Syst. Biol.* 45:559-568.
374. Xu X., Arnasson U. (1996). A complete sequence of the mitochondrial genome of the western lowland gorilla. *Mol. Biol. Evol.* 13:691-698.
375. Yang Z. (1994). Estimating the pattern of nucleotide substitution. *J. Mol. Evol.* 39:105-111.
376. Yang Z. (1995). A space-time process model for the evolution of DNA sequences. *Genetics* 139:993-1005.
377. Yang Z. (1996). Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol. Evol.* 11:367-370.
378. Yang Z. (1997). How often do wrong models produce better phylogenies ? *Mol. Biol. Evol.* 14(1): 105-108.
379. Yang Z., Kumar S. (1996). Approximate methods for estimating the pattern of nucleotide substitution and the variation of substitution rates among sites. *Mol. Biol. Evol.* 13(5): 650-659.
380. Yang Z., Kumar S., Nei M. (1995). A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* 141:1641-1650.
381. Yang Z., Roberts D. (1995). On the use of nucleic acid sequences to infer early branching in the tree of life. *Mol. Biol. Evol.* 12(3): 451-458.
382. Yang Z., Yoder A. D. (1999). Estimation of the transition / transversion rate bias and species sampling. *J. Mol. Evol.* 48:274-283.
383. Yoder A. D. (1992). The applications and limitations of autogenic comparisons for phylogeny reconstruction: the case of the strepsirhine internal carotid artery. *J. Hum. Evol.* 23:183-195.
384. Yoder A. D. (1994). Relative position of the Cheirogaleidae in strepsirhine phylogeny: a comparison of morphological and molecular methods and results. *Am. J. Phys. Anthropol.* 94:25-46.
385. Yoder A. D. (1996). The use of phylogeny for reconstructing lemuriform biogeography. In "Biogeography of Madagascar". Laurenço W. R. (Ed.). ORSTOM. Paris, pp 245-258.
386. Yoder A. D. (1997). Back to the future: a synthesis of strepsirrhine systematics. *Evol. Anthropol.* 6:11-22.

387. Yoder A. D., Cartmill M., Ruvolo M., Smitt K., Vigalys R. (1996). Ancient single origin for Malagasy primates. *Proc. Natl. Acad. Sci. USA* 93:5122-5126.
388. Yoder A. D., Irwin J. A. (1999). Phylogeny of the Lemuridae: effects of character and taxon sampling on resolution of species relationships within Eulemur. *Cladistics* 15(3): 351-361.
389. Yoder A. D., R4akotosamimanana B., Parson T. J. (1999). Ancient DNA in subfossil Lemurs. Methodological challenges and their solutions. In "New directions in Lemur studies.". Ratotosamimanana B. , Rasamimanana H. , Ganzhorn J. U., Goodman S. M. (Eds.). Kluever Academic Press / Plenum Publishers. New York, Boston, Dordrecht, London, Moscow, pp 7-17.
390. Yoder A. D., Vilgalys R., Ruvolo M. (1996). Molecular evolutionary dynamics of cytochrome b in strepsirrhine primates. The phylogenetic significance of the third position transversion. *Mol. Biol. Evol.* 13:1339-1350.
391. Zang Y., Ryder E. R. (1995). Different rates of mitochondrial DNA sequence evolution in Kirk's dik-dik (*Madogna kikii*) population. *Mol. Phylogenet. Evol.* 4:291-297.
392. Zhang J., Nei M. (1997). Accuracies of ancestral amino acid sequences inferred by parsimony, likelihood and distance methods. *J. Mol. Evol.* 44:5139-5146.
393. Zuckerhandl E., Pauling L. (1965). Evolutionary divergence and convergence in proteins. In "Evolving genes and proteins.". Brysok V., Vogel H. J. (Eds.). Academic Press. New York, pp 97.

## TABLE DES ILLUSTRATIONS

Figure 1 : Relations entre transitions et transversions.....	4
Figure 2 : Relations entre les principaux modèles de substitutions .....	14
Figure 3 : Analyse 'likelihood mapping'.....	24
Figure 4 : Interior branch test.....	32
Figure 5 : Relative rate test .....	33
Figure 6 : Cartographie de l'ADN mitochondrial humain.....	37
Figure 7 : Représentation schématique du cytochrome b .....	39
Figure 8 : Classification des Lemuriformes de Osman-Hill (1953).....	44
Figure 9 : Classification des Lemuriformes de Petter et al. (1977).....	45
Figure 10 : Classification générale des Lemuriformes.....	46
Figure 11 : Relations entre formes sub-fossiles et actuelles (d'après Yoder et Irwin – 1999).....	50
Figure 12 : Structure secondaire du fragment de 12S rRNA étudié.....	59
Figure 13 : Test 'Likelihood mapping' pour les différents types de séquences.....	65
Figure 14 : Reconstruction phylogénétique à partir des séquences de cytochrome b (357 bp) pour les tests de monophylie.....	75
Figure 15 : Reconstruction phylogénétique à partir des séquences de 12S rRNA (393 bp) pour les tests de monophylie.....	76
Figure 16 : Reconstruction phylogénétique à partir des séquences mixtes (cytochrome b et 12S rRNA zones 'Loop') pour les tests de monophylie.....	77
Figure 17 : Répartition des substitutions par site (cytochrome b 357 bp).....	81
Figure 18 : Reconstructions phylogénétiques utilisant le cytochrome b (357 bp).....	83
Figure 19 : Reconstructions phylogénétiques utilisant le cytochrome b (positions 1 et 3).....	85
Figure 20 : Nœuds ancestraux étudiés pour le cytochrome b .....	88
Figure 21 : Codons variables pour les séquences ancestrales de cytochrome b .....	90
Figure 22 : Transitions et transversion pour les séquences ancestrales de cytochrome b.....	91
Figure 23 : Reconstruction NJ dans LINETREE pour le cytochrome b (357 bp) .....	94
Figure 24 : Arbre linéarisé obtenu à partir des séquences de cytochrome b (357 bp).....	96
Figure 25 : Arbre linéarisé obtenu à partir des séquences consensus de cytochrome b pour les différents groupes.....	98
Figure 26 : Arbre linéarisé obtenu à partir des séquences partielles de cytochrome b (positions 1 et 3).....	100
Figure 27 : Arbre linéarisé obtenu à partir des séquences consensus partielles de cytochrome b (positions 1 et 3).....	101
Figure 28 : Schéma d'évolution des différents groupes en fonction des données obtenues à partir des séquences de cytochrome b .....	103
Figure 29 : Acides aminés variables pour les séquences cytochrome b.....	107
Figure 30 : Alignement des acides aminés variables pour les différentes séquences (cytochrome b).....	108
Figure 31 : Reconstructions phylogénétiques à partir des séquences protéiques de cytochrome b.....	111
Figure 32 : Nœuds ancestraux analysés pour les séquences protéiques de cytochrome b.....	114
Figure 33 : Acides aminés variables pour les séquences protéiques ancestrales et actuelles de cytochrome b.....	115
Figure 34 : Modifications des acides aminés ancestraux à partir de la traduction des séquences nucléotidiques ancestrales .....	118
Figure 35 : Modifications des acides aminés ancestraux à partir de la traduction des séquences protéiques ancestrales.....	119
Figure 36 : Arbre NJ obtenu dans LINETREE à partir des séquences protéiques (cytochrome b).....	120
Figure 37 : Arbre linéarisé obtenu à partir des séquences protéiques de cytochrome b .....	121



Figure 38 : Arbre linéarisé obtenu à partir des séquences protéiques consensus (cytochrome b) ...	123
Figure 39 : Répartition des substitution par siteErreur ! Signet non défini. pour les séquences 12S rRNA (393 bp) .....	125
Figure 40 : Répartition des substitution par site pour les séquences 12S rRNA (zones 'Loop') .....	126
Figure 41 : Reconstructions phylogénétiques pour les séquences 12S rRNA (393 bp).....	129
Figure 42 : Reconstructions phylogénétiques pour les séquences 12S rRNA (zones 'Loop').....	131
Figure 43 : Topologie NJ obtenue dans LINETREE à partir des séquences 12S rRNA (393 bp)...	135
Figure 44 : Arbre linéarisé obtenu à partir des séquences 12S rRNA (393 bp).....	136
Figure 45 : Arbre linéarisé obtenu à partir des séquences consensus 12S rRNA (393 bp).....	138
Figure 46 : Topologie NJ obtenue dans LINETREE à partir des séquences 12S rRNA (zones 'Loop') .....	139
Figure 47 : Arbre linéarisé obtenu à partir des séquences 12S rRNA (zones 'Loop').....	140
Figure 48 : Arbre linéarisé obtenu à partir des séquences consensus 12S rRNA (zones 'Loop').....	141
Figure 49 : Reconstructions phylogénétiques à partir des données mixtes (cytochrome b et 12S rRNA zones 'Loop').....	145
Figure 50 : Topologie NJ obtenue dans LINETREE à partir des données mixtes.....	147
Figure 51 : Arbre linéarisé obtenu à partir des données mixtes .....	148
Figure 52 : Arbre linéarisé obtenu à partir des séquences consensus sour les données mixtes .....	150
Figure 53 : Schéma d'évolution reconstruit à partir de l'ensemble des données .....	152
Figure 54 : Topologies NJ pour le test du cluster (DMA, TUP).....	156
Figure 55 : Résumé des différentes topologies concernant la famille des Lemuriformes .....	161
Figure 56 : Hypothèses concernant la colonisation de Madagascar par les Lemuriformes .....	166

## TABLE DES TABLEAUX

Tableau 1 : Modèle JC69 .....	9
Tableau 2 : Modèle TN84 .....	9
Tableau 3 : Modèle K80 .....	10
Tableau 4 : Modèle F81 .....	11
Tableau 5 : Modèle T92 .....	12
Tableau 6 : Modèle HKY85 .....	12
Tableau 7 : Modèle TN93 .....	13
Tableau 8 : Matrice ML pour le modèle K80.....	23
Tableau 9 : Origine des séquences utilisées .....	56
Tableau 10 : Génomes mitochondriaux complets utilisés.....	56
Tableau 11 : Origine des séquences de Lorisidae .....	56
Tableau 12 : Définition des groupes taxonomiques .....	57
Tableau 13 : Composition des groupes d'étude.....	58
Tableau 14 : Compositions nucléotidiques du cytochrome b (séquences de 357 bp).....	60
Tableau 15 : Compositions nucléotidiques du cytochrome b (positions 1 et 3) .....	61
Tableau 16 : Caractéristiques des séquences cytochrome b (357 bp).....	61
Tableau 17 : Caractéristiques des séquences cytochrome b (Positions 1 et 3) .....	61
Tableau 18 : Transitions et transversions pour les séquences cytochrome b.....	62
Tableau 19 : Transitions observées et calculées pour le cytochrome b (357 bp).....	62
Tableau 20 : Transitions observées et calculées pour le cytochrome b (positions 1 et 3) .....	63
Tableau 21 : Transversions observées et calculées pour le cytochrome b (357 bp) .....	63
Tableau 22 : Transversions observées et calculées pour le cytochrome b (positions 1 et 3).....	63
Tableau 23 : Distances génétiques pour les séquences de cytochrome b (357 bp).....	64
Tableau 24 : Distances génétiques pour les séquences de cytochrome b (positions 1 et 3).....	65
Tableau 25 : Résultats du test 'Likelihood mapping' pour les séquences cytochrome b.....	66
Tableau 26 : Composition des séquences protéiques de cytochrome b .....	67
Tableau 27 : Distances génétiques pour les séquences protéiques de cytochrome b .....	67
Tableau 28 : Composition nucléotidique des séquences 12S rRNA (393 bp).....	68
Tableau 29 : Composition nucléotidique des séquences 12S rRNA (zones 'Loop') .....	69
Tableau 30 : Composition nucléotidique des séquences 12S rRNA (zones 'Steem').....	69
Tableau 31 : Caractéristiques des différents types de séquences 12S rRNA .....	70
Tableau 32 : Transitions et transversions pour les différents types de séquences 12S rRNA .....	70
Tableau 33 : Transitions observées et calculées pour les séquences 12S rRNA (393 bp).....	71
Tableau 34 : Transitions observées et calculées pour les séquences 12S rRNA (zones 'Loop').....	71
Tableau 35 : Transitions observées et calculées pour les séquences 12S rRNA (zones 'Steem') .....	71
Tableau 36 : Transversions observées et calculées pour les séquences 12S rRNA (393 bp) .....	71
Tableau 37 : Transversions observées et calculées pour les séquences 12S rRNA (zones 'Loop') .....	72
Tableau 38 : Transversions observées et calculées pour les séquences 12S rRNA (zones 'Steem') .....	72
Tableau 39 : Distances génétiques pour les séquences 12S rRNA(393 bp).....	73
Tableau 40 : Distances génétiques pour les séquences 12S rRNA (zones 'Loop').....	73
Tableau 41 : Résultats du test 'Likelihood mapping' pour les séquences 12S rRNA.....	73
Tableau 42 : Caractéristiques des séquences mixtes.....	74
Tableau 43 : Test pour la séquence outgroupe optimale (cytochrome b 357 bp) .....	78
Tableau 44 : Test pour la séquence outgroupe optimale (12S rRNA 393 bp) .....	79
Tableau 45 : Estimations du paramètre alpha pour les séquences cytochrome b .....	80
Tableau 46 : Comparaison des topologies (cytochrome b 357 bp).....	85
Tableau 47 : Comparaison des topologies (cytochrome b positions 1 et 3).....	85

Tableau 48 : Longueurs des branches principales des topologies ML et NJ (cytochrome b 357 bp)	87
Tableau 49 : Test des branches internes pour les topologies cytochrome b (357 bp)	87
Tableau 50 : Correspondances entre nœuds ancestraux et groupes de séquences pour le cytochrome b	89
Tableau 51 : Compositions nucléotidiques des séquences ancestrales de cytochrome b	89
Tableau 52 : Répartition des transitions et transversion pour les séquences ancestrales de cytochrome b	91
Tableau 53 : Substitutions observées et calculées pour les séquences ancestrales de cytochrome b	92
Tableau 54 : Modifications des sites sur les séquences ancestrales de cytochrome b en fonction des groupes	92
Tableau 55 : Modifications des sites sur les séquences ancestrales de cytochrome b entre nœud terminal et groupe actuel de séquences	93
Tableau 56 : Test 'deux clusters' pour la reconstruction NJ à partir du cytochrome b (357 bp)	95
Tableau 57 : Test 'deux clusters' pour la reconstruction NJ à partir du cytochrome b (357 bp) après élimination des séquences aberrantes	95
Tableau 58 : Temps de divergence estimés à partir des arbres linéarisés pour le cytochrome b	97
Tableau 59 : Test 'deux clusters' pour la topologie à partir des séquences consensus de cytochrome b	99
Tableau 60 : Temps de divergences calculés dans BASEML à partir des séquences cytochrome b (357 bp) avec différents modèles de substitution	102
Tableau 61 : Dates moyennes de divergence des nœuds ancestraux dans le cas des séquences de cytochrome b	103
Tableau 62 : Taux de substitution pour les différents groupes (cytochrome b)	104
Tableau 63 : Taux de mutation pour les différents groupes (cytochrome b)	104
Tableau 64 : Substitutions synonymes et non synonymes pour les différents groupes (cytochrome b)	105
Tableau 65 : Pourcentage des SSNCS observées et calculées pour les différents groupes (cytochrome b)	106
Tableau 66 : Distances de Grantham pour les différents groupes (cytochrome b)	106
Tableau 67 : Répartition des acides aminés variables en fonction des différents groupes (cytochrome b)	107
Tableau 68 : Acides aminés caractéristiques (cytochrome b)	109
Tableau 69 : Répartition des différents acides aminés (cytochrome b) dans les différents groupes	110
Tableau 70 : Comparaison des topologies obtenues à partir des séquences protéiques de cytochrome b	112
Tableau 71 : Longueurs des branches principales de la topologie NJ pour les séquences protéiques de cytochrome b	113
Tableau 72 : Test des branches internes pour différentes topologies à partir des séquences protéiques de cytochrome b	113
Tableau 73 : Identification des nœuds ancestraux pour les séquences protéiques de cytochrome b	115
Tableau 74 : Acides aminés ancestraux variables (cytochrome b) à partir de la traduction des séquences nucléotidiques ancestrales et à partir des séquences protéiques ancestrales	117
Tableau 75 : Test 'deux clusters' sur la topologie NJ (séquences protéiques de cytochrome b)	120
Tableau 76 : Temps de divergence estimés à partir des arbres linéarisés (séquences protéiques de cytochrome b)	122
Tableau 77 : Temps de divergence moyens estimés à partir des séquences protéiques de cytochrome b	124
Tableau 78 : Valeurs du paramètre alpha pour les séquences 12S rRNA (393 bp)	125
Tableau 79 : Valeurs du paramètre alpha pour les séquences 12S rRNA (zones 'Loop')	126

Tableau 80 : Comparaison des topologies obtenues à partir des séquences 12S rRNA (393 bp)....	131
Tableau 81 : Comparaison des topologies obtenues à partir des séquences 12S rRNA (zones 'Loop')	131
Tableau 82 : Longueurs des branches pour les topologies NJ (12S rRNA).....	133
Tableau 83 : Test 'quatre clusters' pour les topologies NJ (12S rRNA).....	133
Tableau 84 : Test 'deux clusters' sur la topologie NJ obtenue à partir des séquences 12S rRNA (393 bp).....	134
Tableau 85 : Estimations des dates de divergences à partir des séquences 12S rRNA (393 bp et zones 'Loop').....	137
Tableau 86 : Dates de divergences moyennes estimées à partir des séquences 12S rRNA.....	142
Tableau 87 : Taux de substitutions pour les branches principales des topologies 12S rRNA.....	142
Tableau 88 : Comparaison des topologies obtenues à partir des données mixtes.....	145
Tableau 89 : Test 'deux cluster' pour la topologie NJ à partir des données mixtes.....	147
Tableau 90 : Estimations des dates de divergence pour les données mixtes.....	149
Tableau 91 : Dates de divergence estimées à partir des topologies ML et MP avec les données mixtes.....	151
Tableau 92 : Taux de substitution par branche pour les topologies ML et MP avec les données mixtes.....	151
Tableau 93 : Dates de divergence moyennes obtenues à partir de l'ensemble des données.....	152
Tableau 94 : Test des topologies pour l'association DMA - TUP.....	154
Tableau 95 : Test des topologies pour le cluster DTU.....	155
Tableau 96 : Test des topologies pour le groupement IND – MGA.....	157
Tableau 97 : Test des topologies pour les groupes EHP, VAR, IND et MGA.....	158
Tableau 98 : Tests des topologies pour les Megaladapidae.....	159

# TABLE DES MATIERES

<b>SOMMAIRE</b> .....	<b>1</b>
<b>INTRODUCTION</b> .....	<b>2</b>
<b>TECHNIQUES D'ETUDE</b> .....	<b>3</b>
GENERALITES .....	3
<i>Etapes d'un traitement phylogénétique</i> .....	3
<i>Données utilisables</i> .....	3
<i>Alignement des séquences</i> .....	4
<i>Horloge moléculaire et temps de divergence</i> .....	5
RECONSTRUCTIONS PHYLOGENETIQUES .....	6
<i>Modèles de substitution et distances génétiques</i> .....	6
p-distances .....	8
Modèle JC69 .....	8
Modèle TN84 .....	9
Modèle K80 .....	10
Modèle F81 .....	11
Modèle T92 .....	11
Modèle HKY85 .....	12
Modèle TN93 .....	12
Modèle F84 .....	13
Modèle REV .....	13
Distances gamma .....	13
Séquences protéiques .....	15
<i>Substitutions synonymes et non synonymes</i> .....	15
<i>Méthodes de reconstruction</i> .....	18
Méthode UPGMA .....	19
Méthode ME .....	20
Méthode NJ .....	20
Méthode MP .....	20
Méthode ML .....	23
Quartet Puzzeling .....	24
Parsimony Ratchet .....	25
<i>Mesures de la validité des arbres</i> .....	25
Méthodes de rééchantillonnage .....	26
Tests de vraisemblance .....	26
<i>Méthodes de consensus</i> .....	27
<i>Méthodes de réarrangements</i> .....	28
Indices .....	28
MODALITES D'UTILISATION .....	29
<i>Analyses combinées</i> .....	30
<i>Effets d'échantillonnage</i> .....	30
PROBLEMES PRINCIPAUX .....	30
<i>Longueur des branches</i> .....	30
<i>Le phénomène LBA</i> .....	31
METHODES COMPLEMENTAIRES .....	32
<i>Interior branch test</i> .....	32

<i>Four cluster analysis</i> .....	32
<i>Relative rate test</i> .....	33
<i>Arbres linéarisés</i> .....	33
NIVEAU D'INFORMATION PHYLOGENETIQUE.....	33
DATATION DES SEQUENCES.....	34
RECONSTRUCTION DE SEQUENCES ANCESTRALES.....	35
EVOLUTION GENERALE DES SEQUENCES.....	36
<b>L'ADN MITOCHONDRIAL</b> .....	<b>37</b>
GENERALITES.....	37
CARACTERISTIQUES.....	38
GENES ETUDIES.....	38
<i>Cytochrome b</i> .....	38
<i>12S et 16S rRNA</i> .....	40
<b>LES LEMURIFORMES ACTUELS ET SUB-FOSSILES</b> .....	<b>41</b>
<b>MATERIEL ET METHODES</b> .....	<b>52</b>
EXTRACTION DE L'ADN.....	52
AMPLIFICATION DES SEQUENCES MITOCHONDRIALES.....	52
<i>Cytochrome b</i> .....	52
<i>12S rRNA</i> .....	53
ANALYSE DES PRODUITS D'AMPLIFICATION.....	53
SEQUENÇAGE DES PRODUITS D'AMPLIFICATION.....	53
PROGRAMMES INFORMATIQUES UTILISÉS.....	54
SEQUENCES UTILISEES.....	55
<i>Origine des séquences et groupes d'études</i> .....	55
<i>Prétraitement des séquences</i> .....	58
<i>Cytochrome b</i> .....	58
<i>12S rRNA</i> .....	58
<b>RESULTATS</b> .....	<b>60</b>
CARACTERISTIQUES GENERALES DES SEQUENCES.....	60
<i>Cytochrome b</i> .....	60
Séquences nucléotidiques.....	60
Composition.....	60
Transitions et transversions.....	62
Distances.....	64
Test 'Likelihood mapping'.....	66
Séquences protéiques.....	66
Composition.....	66
Distances.....	67
Test 'Likelihood mapping'.....	68
<i>12S rRNA</i> .....	68
Composition.....	68
Transitions et transversions.....	70
Distances.....	72
Test 'Likelihood mapping'.....	73
<i>Données mixtes</i> .....	74
Caractéristiques.....	74
Test 'Likelihood mapping'.....	74
TEST DE LA MONOPHYLIE DES LEMURIFORMES.....	74

DETERMINATION DE LA SEQUENCE OUTGROUPE OPTIMALE .....	78
TRAITEMENTS SUR LES SEQUENCES CYTOCHROME B .....	79
<i>Séquences nucléotidiques</i> .....	79
Substitutions nucléotidiques par site .....	79
Reconstructions phylogénétiques .....	81
Reconstruction des séquences ancestrales .....	88
Horloge moléculaire et arbres linéarisés .....	94
Substitutions synonymes et non synonymes .....	105
<i>Séquences protéiques</i> .....	107
Modifications des acides aminés .....	107
Reconstructions phylogénétiques .....	110
Reconstruction des séquences ancestrales .....	114
Horloge moléculaire et arbres linéarisés .....	119
TRAITEMENTS SUR LES SEQUENCES 12S RRNA .....	124
<i>Substitutions nucléotidiques par site</i> .....	124
<i>Reconstructions phylogénétiques</i> .....	127
<i>Horloge moléculaire et arbres linéarisés</i> .....	134
TRAITEMENTS SUR LES DONNEES MIXTES .....	143
<i>Reconstructions phylogénétiques</i> .....	143
<i>Horloge moléculaire et arbres linéarisés</i> .....	146
LES RELATIONS DMA – TUP .....	154
LES RELATIONS IND – MGA .....	157
<b>DISCUSSION .....</b>	<b>160</b>
<b>CONCLUSIONS .....</b>	<b>169</b>
<b>BIBLIOGRAPHIE .....</b>	<b>171</b>
<b>TABLE DES ILLUSTRATIONS .....</b>	<b>189</b>
<b>TABLE DES TABLEAUX .....</b>	<b>191</b>
<b>TABLE DES MATIERES .....</b>	<b>194</b>
<b>INDEX .....</b>	<b>197</b>

# INDEX

## A

Acide aminé 5, 6, 7, 15, 16, 19, 23, 35, 36, 38, 40, 41, 66, 67, 106, 107, 108, 109, 110, 115, 116, 117, 118, 124

### ADN

hybridation ..... 5  
mécanisme de réparation ..... 6, 7  
mitochondrial (mtDNA). 10, 11, 12, 16, 17, 37, 38, 39, 40, 49, 52, 163, 164, 167  
nucléaire ..... 38  
polymérase ..... 7

Afrique ..... 2, 42, 167, 170

Algorithme ..... 4, 5, 18, 20, 22, 23, 30, 35, 49

Allèles..... 6, 18

polymorphiques ..... 18

Alpha ..... 54, 80, 81, 125, 126, 169

Amphibiens ..... 17

Amplification ..... 51, 52, 53, 163

### Analyse

combinée ..... 30

Ancêtre ..... 2, 18, 33, 49, 166, 167

Anthropoïdes ..... 34, 49

### Arbres

branches.. 20, 21, 22, 23, 24, 30, 31, 32, 33, 34, 35, 49, 87, 92, 93, 103, 104, 105, 106, 112, 113, 117, 118, 132, 133, 134, 142, 151, 154, 155, 157, 158, 159, 164, 165, 169

courtes ..... 31, 49

internes courtes..... 30

Long Branch Attraction (LBA)..... 31, 35

Long Branch Distracts (LBD)..... 31

longues ..... 31, 35

longueur.... 18, 20, 21, 30, 33, 34, 35, 86, 95, 98, 105, 112, 133

réalité..... 28

correct..... 19, 20

également parcimonieux ..... 21, 27

enraciné ..... 18, 19, 20, 21, 35, 74, 78

évaluation statistique ..... 3

islands..... 28

le plus parcimonieux .... 21, 22, 28, 81, 110, 127, 143

linéarisé 6, 33, 95, 96, 97, 98, 99, 100, 121, 122, 135, 136, 137, 138, 140, 141, 148, 149, 150, 165, 170

moyen.....27

Nœuds .....22, 26, 29, 30, 34, 35, 42, 74, 79, 88, 89, 90, 92, 93, 96, 97, 102, 103, 114, 115, 117, 121, 122, 124, 132, 137, 140, 142, 149, 151, 153, 164, 170

internes.....29, 35

robustesse.....29

nombre de pas .....21

non enraciné .....18, 19, 22

non enracinés .....19, 20, 78

phylogénétique...4, 5, 18, 26, 28, 34, 38, 50

principes d'optimisation .....18

racine.....20, 99

### reconstructions

#### Méthodes

Quartet Puzzeling.....24

quartet .....24, 25, 26, 66, 68, 73, 74, 113

quartets non résolus.....24, 66

UPGMA .....19

recherche de type heuristique.....4, 22

### Reconstructions

#### méthodes

Maximum Likelihood (ML)23, 26, 29, 31, 81, 82, 83, 85, 86, 87, 110, 111, 112, 127, 131, 132, 143, 145, 146, 150, 151, 165

#### Méthodes

de comparaison des outgroupes .....23

dite exacte .....4

maximum de vraisemblance .....8, 79

Maximum Likelihood (ML)23, 24, 26, 29, 30, 31, 35, 49, 81, 82, 83, 85, 86, 87, 110, 111, 112, 127, 131, 132, 143, 145, 146, 150, 151, 165

Maximum Parsimony (MP) 20, 21, 26, 28, 29, 31, 49, 81, 82, 83, 85, 110, 111, 112, 127, 131, 132, 143, 145, 146, 150, 151, 162, 165

Minimum Evolution (ME) ..20, 21, 29, 30

Neighbor-Joining (NJ) ..20, 21, 24, 29, 31, 33, 74, 79, 82, 83, 85, 86, 87, 88, 94, 98, 101, 103, 104, 110, 111, 112, 113, 114, 119, 123, 124, 127, 131, 132, 133, 134, 138, 141, 143, 145, 146, 156, 163, 165



Parsimony Ratchet.....	25, 81, 127, 132, 143
tree species .....	6
réel.....	18
score .....	4, 5, 18
optimal.....	4, 5
sous arbre.....	24, 26, 27, 28
Species tree.....	18
sub optimal.....	25, 28
taxon sampling .....	31
topologie... 8, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 41, 42, 49, 77, 79, 85, 86, 87, 88, 94, 95, 96, 97, 99, 100, 101, 103, 104, 105, 111, 112, 113, 114, 116, 118, 119, 122, 123, 124, 131, 132, 133, 134, 135, 141, 145, 146, 149, 150, 151, 154, 155, 156, 157, 158, 159, 160, 165, 167, 169, 170	
Archaeoindri.....	49, 50
Archaeolemur .....	48, 49, 51
ARN	
de transfert.....	37
ribosomal.....	37, 40
rRNA.....	31, 40, 76, 154, 158, 167
Artiodactyles .....	5, 43
Asie.....	2
Autapomorphie.....	29

## B

Babakotia.....	50
Babouin .....	17
Bases..... 21, 22, 27, 35, 37, 41, 42, 49, 51, 86, 112, 116, 132, 145, 146, 154, 157, 160, 162, 163, 167, 169	
adénine ....	39, 60, 63, 68, 69, 70, 74, 89, 93
changements.....	4
cytosine... 11, 12, 21, 31, 60, 63, 68, 69, 70, 74, 89, 93	
fréquence .....	13, 27, 30, 35
guanine ... 11, 12, 21, 31, 39, 60, 63, 68, 69, 70, 74, 89, 92	
nucléotides.. 3, 6, 7, 8, 9, 11, 13, 19, 20, 21, 22, 23, 26, 35, 39, 43, 61, 62, 63, 69, 74, 89, 93, 163, 164	
positions .....	18
purines .....	12
pyrimidines.....	12
thymine.... 39, 60, 63, 68, 69, 70, 74, 89, 93	
Biais	
de topologie.....	31

des mutations .....	39
des transitions.....	7
du rapport Ti/Tv.....	21, 30
Biologie moléculaire.....	2, 41, 51, 160, 169

## C

Caractères...3, 6, 18, 19, 20, 22, 23, 25, 26, 28, 29, 30, 31, 35, 39, 40, 41, 42, 43, 46, 49, 50, 78, 81, 127, 143, 160, 163, 164	
discrets .....	18
Equivocal character state .....	35
Four fold degenerate .....	17
homoplasies.....	18, 28, 29, 40, 49
incongruence .....	29
indépendance.....	26
invariants.....	29
morphologiques.....	20, 41, 42, 161, 163
non informatifs.....	29
poids relatifs.....	4
quantitatifs.....	3
singletons .....	61, 69
troisième position des codons .....	39
validité	
de l'ensemble .....	28
Carnivores .....	17, 75, 76
Caryotypes .....	41
Changements	
évolutionnaires.....	20
mutations	
dites synonymes .....	5
neutres .....	5
non silencieux .....	15
substitutions	
modèle...7, 14, 15, 17, 23, 26, 29, 30, 31, 35, 64, 68, 72, 74, 101, 102, 104, 141	
réversion.....	22
transitions 3, 7, 8, 10, 11, 12, 17, 20, 23, 31, 34, 39, 40, 41, 62, 63, 64, 70, 72, 74, 79, 81, 82, 90, 91, 92, 124, 127, 143, 157, 160, 162	
fréquence.....	7
saturation.....	7
transitions-transversions	
rapport .....	7, 9
transitions-transversions-rapport.....	7
transversions .3, 7, 8, 10, 11, 17, 20, 23, 31, 34, 39, 40, 41, 62, 63, 64, 70, 72, 74, 79, 81, 82, 90, 91, 92, 124, 127, 132, 143, 157, 160, 162	

Cheirogaleidae.. 42, 43, 48, 55, 57, 58, 87, 89,  
118, 153, 157, 160, 162, 165, 167, 168, 169  
Cheiromys ..... 41  
Chimpanzés ..... 34  
Chloroplaste ..... 16  
Cladistique..... 164  
  approche ..... 29  
Cladogramme ..... 5  
Codons.. 16, 17, 39, 40, 41, 42, 43, 49, 58, 60,  
61, 64, 81, 89, 92, 104, 105, 106, 162  
  homologues ..... 16  
  saturation de la troisième position..... 34, 38  
  usage..... 17  
Colonisation unique..... 2  
Contraintes sélectives..... 6, 16, 38, 40

## D

Datation ..... 6, 34, 35, 43, 51, 164, 165  
  point de calibration..... 5  
Date géologique..... 6  
Daubentonia 41, 42, 43, 48, 49, 50, 51, 55, 57,  
58, 87, 104, 153, 154, 160, 162, 165, 167,  
169, 170  
  Aye-Aye ..... 41  
  *Daubentonia madagascariensis* ..... 57  
  *Daubentonia robusta* ..... 49, 50  
Délétion ..... 4, 15, 58  
Dérive génétique aléatoire..... 6  
Diphylétique ..... 42  
  diphylie..... 42  
Distances  
  avec correction de Poisson ..... 15  
  de Grantham ..... 16, 106, 117, 118  
  de Miyata..... 16, 117  
  Det Log..... 31  
  gamma ..... 15  
  génétique 6, 7, 10, 17, 26, 29, 33, 38, 43, 49  
  inter-taxa..... 34  
  Kimura-2-paramètres ..... 10, 64, 79, 82, 99  
  p-distances..... 8, 9, 15, 29, 31, 78  
Divergence ... 5, 6, 7, 8, 18, 20, 34, 49, 51, 96,  
97, 99, 100, 102, 103, 121, 122, 124, 135,  
137, 142, 148, 149, 152, 153, 164, 165,  
167, 170  
  datation ..... 34  
  date ..... 5, 6, 17, 33, 43, 102, 122, 136, 140,  
  141, 148, 149, 150, 151, 153, 167, 170  
  entre espèces..... 18  
  niveau ..... 7  
  taux ..... 7

temps ...5, 21, 43, 49, 96, 98, 100, 101, 102,  
121, 123, 124, 140, 141, 152

Données  
  génétiques ..... 6, 42  
  morphologiques..... 3, 42, 51, 160, 166

## E

Echantillonnage  
  erreur ..... 18  
Eocène..... 2, 167  
Eucaryotes..... 34  
Eulemur ..... 30, 42, 43, 169  
Euthériens ..... 7  
Evolution..... 2, 5, 6, 15, 16, 17, 19, 20, 22, 33,  
34, 35, 36, 38, 39, 40, 41, 43, 49, 51, 54,  
78, 92, 95, 98, 99, 100, 102, 117, 118, 119,  
134, 142, 147, 151, 153, 164, 169  
  adaptative ..... 15, 16  
  contraintes ..... 20  
  des Primates ..... 5  
  dynamique ..... 7, 30  
  mécanismes ..... 16  
  moléculaire..... 16, 117  
  processus ..... 27  
  relation ..... 18, 36  
  taux absolu ..... 17

## F

Fossile .5, 6, 34, 41, 42, 46, 49, 50, 51, 55, 64,  
80, 86, 87, 108, 112, 125, 132, 146, 158,  
159, 163, 164, 165, 170  
  sub fossile..... 169, 170  
Sub fossiles ...41, 46, 49, 50, 51, 55, 64, 80,  
86, 87, 108, 112, 125, 132, 146, 158,  
159, 160, 162, 163, 165, 167  
Fréquence..7, 9, 10, 11, 13, 17, 27, 30, 31, 35,  
39, 63, 93, 164

## G

Galagidae ..... 42  
Gamma  
  distribution ..... 13, 15, 31  
  modèle..... 34  
  paramètre alpha ( $\alpha$ )..... 169  
  paramètre de forme ..... 13, 79, 80, 126, 169  
Gènes...3, 5, 16, 17, 18, 19, 30, 31, 33, 34, 35,  
37, 38, 40, 49, 51, 160, 162, 164, 166  
  mitochondriaux 17, 31, 33, 38, 51, 162, 166

12S rRNA . 17, 34, 35, 51, 53, 55, 56, 68, 74, 75, 76, 78, 79, 127, 134, 141, 143, 151, 154, 155, 157, 158, 159, 163, 165, 167, 169	Decay Index (DC) .....29
zones loop. 58, 68, 69, 70, 72, 73, 124, 126, 127, 131, 132, 134, 138, 141, 142, 143, 154, 155, 157, 158, 159	Exact Indices .....29
zones steem ..... 58, 68, 69, 70	Proportional Support Index (PSI) .....29
16S rRNA ..... 40, 42, 162	Rescaled Consistency Index (RCI) .....29
CO II..... 41, 42, 160, 162	Retention Index (RI) .....29, 30
Control Region (CR) ..... 37	Support Index (SI).....28, 29
cytochrome b.. 17, 33, 38, 39, 40, 41, 42, 49, 51, 53, 55, 56, 58, 64, 66, 68, 74, 75, 76, 78, 79, 80, 81, 94, 99, 100, 124, 125, 126, 132, 133, 141, 143, 146, 151, 152, 153, 154, 155, 157, 158, 159, 160, 162, 163, 165, 166, 167, 169	Total Support Index (TSI).....29
D-loop..... 37	Indriidae .....41, 48, 50, 55, 57, 58, 80, 86, 87, 118, 157, 158, 160, 162, 165, 167, 169
NADH 4 ..... 38	Insertion .....4, 15, 58
modèle d'expression ..... 17	Invariant .....31
multiples ..... 30	
niveau d'expression ..... 17	<b>L</b>
pseudo- ..... 17	Lémuriens .....2, 41, 43, 49, 51, 167
saturation ..... 33	<i>Lemur</i> .....5, 34, 42, 43, 47
Génome ..... 16, 17, 35, 37, 38, 42, 55, 58	<i>Lemur catta</i> .....43
Gorille..... 34	Lemuriformes..2, 41, 42, 43, 47, 51, 55, 75, 76, 77, 78, 79, 80, 86, 99, 104, 116, 118, 153, 157, 160, 162, 163, 164, 165, 167, 168
Groupe monophylétique. 32, 41, 43, 165, 167, 168	Locus.....38, 164
	non recombinant.....38, 164
<b>H</b>	Logiciel ...3, 22, 32, 35, 55, 58, 62, 66, 79, 81, 82, 87, 88, 94, 101, 104, 105, 110, 113, 114, 124, 141, 142, 150, 165
Hadropithecus..... 48, 49	Lorisiformes.....2, 41, 160
Hapalemur ..... 42, 43, 47, 169	<i>Loris</i> .....5, 34
Hominoïdes	Lorisidae .....42, 55, 96, 121, 153, 156, 157, 165, 167, 170
<i>Homo</i> ..... 34	Lysozyme.....36
Homologies ..... 18	
Homologues ..... 3	<b>M</b>
Horloge moléculaire. 5, 17, 19, 22, 33, 34, 43, 85, 86, 104, 112, 131, 132, 142, 145, 151, 164, 169	Madagascar .2, 41, 42, 49, 165, 167, 168, 169, 170
calibration..... 34	Mammifères .5, 16, 17, 34, 38, 39, 42, 58, 104
globale ..... 5, 165	Markov chain .....23, 27
locale ..... 5	Marqueur génétique .....6, 38
	Maximum de parcimonie (MP).22, 26, 42, 51, 160, 163
<b>I</b>	Maximum de vraisemblance (ML)6, 8, 10, 11, 16, 17, 23, 26, 54, 81, 110, 127, 143, 163
Indépendants..... 3, 23	Megaladapidae ..57, 80, 86, 87, 118, 122, 157, 158, 162, 165
Indices ..... 28, 29, 31	Megaladapis 2, 47, 49, 50, 51, 55, 57, 58, 160, 162, 163, 165, 169, 170
Bremer ..... 28	<i>Megaladapis edwardsi</i> .....2, 51, 55, 57, 160
Consistency Index (CI)..... 28, 29	Mesopropithecus .....48, 50
de vraisemblance ..... 31	Métazoaires .....5, 34
	Méthodes
	consensus .....21, 26

Consensus	24, 27, 51, 55, 64, 67, 72, 73, 81, 85, 90, 93, 98, 99, 100, 105, 110, 116, 117, 118, 122, 127, 137, 140, 143, 149
Majority rule	24, 27
strict	27, 81, 110, 127, 143
High Order Parsimony (HOP)	31
Pondérations successives	29
Réarrangements	25
Branch breaking	25, 28
Branch swapping	25, 28
Branch-and-bound	21, 22
globaux	28, 81, 110, 127, 143
locaux	28
Nearest Neighbor Interchange (NNI)	28
Subtree Pruning and Regrafting (SPR)	28
Tree Bisection and reconnection (TBR)	25, 28
Tree Bisection and Reconnection (TBR)	25, 28
Reconstructions ancestrales	
Joint reconstruction	35
Marginal reconstruction	35, 36
Rééchantillonnage	26, 29
Bootstrap	26, 27, 28, 29, 41, 42, 74, 75, 76, 79, 85, 86, 95, 98, 110, 111, 112, 113, 119, 127, 132, 134, 143, 146, 156, 160, 162, 163
Half jackknife	26
Jackknife	26
Migration	42, 53, 165, 167
multiples	42, 165
Modèles	
F81	12, 17, 101, 102, 104
F84	101, 102, 104
HKY85	12, 13, 17, 29, 66, 101, 102, 104
JC699	10, 14, 16, 23, 29, 38, 101, 102, 104
K80	11, 14, 16, 23, 64, 66, 101, 102, 104
REV	13
T92	11
TN93	14, 64, 72, 74, 101, 102, 104
Modèles REV	7, 101, 102, 104
Modifications multiples	20, 164
Monophylétique	42, 50, 166, 167
Mutations	5, 6, 7, 16, 17, 38, 39, 49, 104, 142, 164
Théorie de la neutralité	5
<b>N</b>	
Non synonymes	15, 16, 17, 105
substitutions	16, 17, 105, 106

## O

Oiseaux	5, 17
Optimisations	
validité	29
Orangs-outangs	17
Ordre d'introduction	4, 28
Outgroupe	20, 33, 78, 79, 87

## P

Palaeopropithecus	48, 49, 50, 51, 163
Paléocène	2, 43
Pan	34, 56
<i>Pan paniscus</i>	34
<i>Pan troglodytes</i>	34, 56
Paraphylie	42
Parcimonieux	5, 23, 43, 61, 69, 74, 81, 110, 127, 143
Perissodactyles	17
Pertinent	29
<i>Phoca standard</i>	34
Phylogénèse	2, 3, 27, 29, 30, 31, 40, 41, 42, 50, 164, 169
moléculaire	2, 41
robustesse	27
Phylogénétique	
analyse	4, 26, 30, 33, 34, 38, 160
études	27, 49, 66, 78
information	31
reconstructions	3, 6, 8, 19, 24, 30, 31, 37, 39, 41, 68, 73, 74, 127, 154, 169
relations	2, 41, 49, 165
signal	29, 31, 33, 78
traitement	3
Poissons	
téléostéens	17
Population	
ancestrale	18
Populations	
taille effective	6
Position constante	19
Primates	2, 5, 17, 34, 39, 40, 41, 42, 49, 51, 54, 56, 75, 76, 160, 162, 165, 167, 169
ancestral	2, 34, 42
malgaches	2, 41, 42, 49, 51, 160, 165, 167
Primer	40, 52, 53
Processus multifactoriel	7
Programmes	
ANCESTOR	54
ANC-GENE	54, 88, 114

BASEML..	85, 88, 101, 103, 104, 131, 132, 141, 142, 145, 150
BIOEDIT .....	54, 55, 58
BN-BS .....	54
CLUSTAL X .....	54, 58
CODEML .....	111, 114, 123
CONSENSE .....	81, 110
DAMBE .....	54, 62, 105
DNAML .....	81
DNAPARS .....	81
DNASP .....	54
GZ-GAMMA	54, 79, 80, 81, 104, 124, 125, 126
LINETREE.	54, 94, 95, 119, 123, 134, 138, 147, 165
MEGA .....	22, 54, 58, 82, 94, 110, 119
MFOLD .....	54, 58
PAML....	35, 54, 85, 88, 101, 111, 114, 142
PAUP .....	25
PHYLIP .....	54, 81, 110
PHYLTEST.	32, 54, 87, 113, 132, 146, 154
PROTML .....	110
PROTPARS .....	110
TREE-PUZZLE..	54, 66, 68, 74, 79, 80, 81, 104, 124, 125, 126
TREEVIEW .....	54
WINCLADA .....	54, 81
Propithecus .....	42, 48, 51, 57, 162
Propriétés physico-chimiques .....	16
Protéines ...	7, 15, 17, 36, 37, 38, 40, 154, 155, 157, 158, 159

## R

Radeaux de végétation .....	2, 167
Réarrangements.	28, 37, 41, 81, 110, 127, 143
Réarrangements chromosomiques.....	41
Rééchantillonnages.....	26, 81, 143
Requins.....	17
Rongeurs.....	5, 43, 75, 76

## S

Sélection	
aléatoire .....	25
Darwinienne positive.....	16
forces .....	6
intensité .....	17
mécanisme .....	16
naturelle.....	16, 19
positive .....	16, 106
pression.....	16, 17, 39, 40

Séquences	
alignement.....	3, 92
paramètres .....	4
ancestrales .	6, 21, 35, 36, 79, 88, 89, 90, 92, 93, 105, 116, 163, 170
codantes.....	15
évolution .....	5, 20
initiales.....	6
mitochondriales.....	17, 60
mixtes .....	74, 143, 146, 154
non codantes.....	3
nucléotidiques .	3, 4, 6, 7, 15, 16, 18, 19, 23, 33, 35, 36, 40, 68, 88, 112, 113, 114, 116, 117, 121, 122, 132, 151, 158
protéiques	3, 4, 6, 15, 17, 18, 33, 35, 67, 68, 107, 109, 110, 112, 113, 114, 116, 117, 118, 122, 123, 124, 154, 158
Simiens.....	5
Sites...6, 7, 8, 9, 13, 15, 16, 17, 19, 20, 21, 22, 23, 26, 27, 31, 34, 35, 36, 38, 39, 40, 54, 61, 69, 74, 79, 80, 81, 89, 92, 93, 105, 108, 124, 125, 126, 160, 162, 165, 169	
de restriction.....	17
indépendance.....	27
Parsimony informative.....	21
Rate Among Sites (RAS).....	31
Spéciation	
date .....	6
Statistique.....	2, 13, 22, 27, 41, 43
analyse.....	6, 13
degrés de liberté (ddl) .	27, 97, 99, 100, 119, 122, 136, 137, 140, 141, 148, 149
distribution nulle .....	27
estimateurs .....	7, 20
forme de la distribution.....	13
hypothèse	
alternative.....	27
nulle.....	17, 27, 33
least squares .....	6
limites de confiance .....	26
loi de Poisson .....	15, 26
paramètres libres .....	27
Posterior probabilities .....	35
problème .....	18
processus de Markov.....	7
processus de Poisson.....	5, 17
Simulation .....	19, 20, 27, 30, 92
de type Monte Carlo.....	27
test .....	17, 33
Test two tailed normal deviate.....	32, 33

tirage aléatoire.....	26	d'évolution.....	5, 19, 20, 31, 33, 38, 40, 97, 99, 100, 103, 137, 140, 141, 148, 151, 164, 169
valeur de CP .	32, 87, 95, 99, 113, 120, 133, 134, 146, 147, 154, 155, 157, 158, 159	constance.....	6, 33, 97, 99, 100, 119, 122, 136, 137, 140, 141, 148, 149, 165
valeurs discrètes .....	3	hétérogénéité.....	13, 16, 17, 31, 142
variable aléatoire .....	35	matrice.....	7
Strepsirhini .	34, 41, 42, 49, 96, 121, 153, 162, 165, 167, 169, 170	Taxonomique .....	55, 165, 167
Structure secondaire .....	15, 40, 58	étude.....	19
Substitutions		Techniques cytogénétiques .....	41
matrice.....	22, 23	Test	
multiples.....	7, 16, 40	deux clusters.....	33, 87, 95, 112, 113, 119, 122, 132, 134, 145, 146, 147, 162, 169
par site .....	8, 15, 33, 79, 81	Interior branch.....	32
processus .....	7, 27	least squares .....	32
SSNCS.....	105, 106	Likelihood mapping .....	74, 113
taux .....	7	Likelihood Ratio Test (LTR) .....	27, 169
Synonymes .....	15, 16, 17, 38, 105	Tests	
différences .....	16	Interior branch.....	32
positions .....	17	Théorie de la neutralité .....	5, 15, 117, 118
substitutions.....	16, 38, 105	Tortues .....	17
<b>T</b>			
Tarsiers.....	5, 41	<b>U</b>	
Taux		Ungulates .....	17
de fixation.....	16, 38	<b>V</b>	
de mutation.....	17, 38, 104, 123, 142, 150	Vertébrés .....	2, 17, 37
de substitution .....	5, 7, 8, 9, 10, 13, 15, 16, 17, 19, 31, 34, 38, 39, 40, 79, 80, 81, 103, 104, 117, 124, 125, 126, 142, 151, 165, 169	<b>W</b>	
constant.....	6	Wagner.....	25
nucléotidique .....	5, 8, 34		

## RESUME

Les Lémuriformes utilisés dans cette étude comprennent des représentants de la plupart des genres de Lemuridae (*Eulemur*, *Hapalemur*, *Varecia*), d'Indriidae (*Indri*, *Propithecus*, *Avahi*), de Lepilemuridae, de Cheirogaleidae (*Cheirogaleus*, *Microcebus*, *Mirza*), de Daubentonia. Un sub fossile (*Megaladapis edwardsi*) est également inclus dans l'analyse, ainsi que deux espèces de Tupaia. Les reconstructions phylogénétiques sont effectuées à partir, soit d'un fragment de 357 bp issu du cytochrome b, soit à partir d'un fragment de 393 bp du 12S rRNA, soit à partir d'une combinaison des deux fragments précédents, en utilisant différentes méthodes (Neighbor-Joining, Maximum Likelihood, Maximum Parsimony, Parsimony Ratchet, Quarter Puzzling). Ces reconstructions sont comparées par des tests LTR (Likelihood Test Ratio) et des tests 'quatre clusters'. L'arbre phylogénétique le plus 'crédible' est alors:

((Cheirogaleidae, (Indriidae, (Lepilemuridae, *Megaladapis*)),  
((*Eulemur*, *Hapalemur*), *Varecia*)), (Daubentonia, Tupaia))

L'association trouvée entre Daubentonia et Tupaia peut indiquer l'existence de deux colonisations distinctes de Madagascar par les Lemuriformes, l'une ayant donné naissance à la lignée des Daubentonia et l'autre à l'ensemble des autres Strepsirhini malgaches. Le sub fossile *Megaladapis edwardsi* se trouve groupé avec les Lepilemuridae et cet ensemble apparaît comme le cluster frère de celui formé par les Indriidae. Il a également été possible de montrer que les différentes séquences étudiées évoluaient sous des horloges moléculaires locales. Enfin les dates de divergence des principaux groupes ont pu être déterminées. La séparation la plus ancienne (voisine de 60 millions d'années) est trouvée pour la divergence entre Daubentonia et Tupaia; alors que la plus récente (voisine de 11 millions d'années) est celle impliquant le *Megaladapis* et les Lepilemuridae. Pour les autres groupes les dates de divergence se répartissent entre 48 et 29 millions d'années.

### Molecular biology and Lemuriformes (Malagasy Primates) phylogeny

In this study most of the species included in the Lemuridae (*Eulemur*, *Hapalemur*, *Varecia*), Indriidae (*Indri*, *Propithecus*, *Avahi*), Lepilemuridae, Cheirogaleidae (*Cheirogaleus*, *Microcebus*, *Mirza*) and Daubentonia are used, as well as a sub-fossil (*Megaladapis edwardsi*) and two Tupaia species. Mitochondrial sequences from a cytochrome b fragment (357 bp) and a 12S rRNA fragment (393 bp) are used to infer phylogenetic relationships between Lemuriformes families. Reconstructions are done using different methods (Neighbor-joining, Maximum Likelihood, Maximum Parsimony, Parsimony Ratchet, Quartet Puzzling). The different trees are compared using LTR (Likelihood Test ratio) tests and 'four clusters' tests. The most 'credible' phylogenetic reconstruction is then:

((Cheirogaleidae, (Indriidae, (Lepilemuridae, *Megaladapis*)),  
((*Eulemur*, *Hapalemur*), *Varecia*)), (Daubentonia, Tupaia))

The association between Daubentonia and Tupaia may indicate the existence of two distinct Lemuriformes colonisations, one conducting to the Daubentonia and the other to all other Malagasy Strepsirhini. The *Megaladapis edwardsi* sub-fossil appear to be grouped with the Lepilemuridae and together formed a clade how is the sister clade of the one including Indriidae. It was also possible to show that the sequences studied here have evolved under local molecular clocks. Divergence dates are estimated for the main groups. The oldest separation (near 60 million years) is found for the split between Daubentonia and Tupaia and the most recent (near 11 million years) is the one between *Megaladapis* and Lepilemuridae. For the other groups estimated divergence dates are between 48 and 29 million years.

**Mots clés:** Lémuriformes, Madagascar, cytochrome b, 12S rRNA, phylogénèse, horloge moléculaire, datation

**Key words:** Lemuriformes, Madagascar, cytochrome b, 12S rRNA, phylogeny, molecular clock, datation

**Discipline:** Sciences de la Vie (Biologie Moléculaire)

Faculté de Médecine – Institut d'Embryologie – 11 rue Human – 67085 Strasbourg Cedex