

UNIVERSITE MONTPELLIER 2

SCIENCES ET TECHNIQUES DU LANGUEDOC

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE MONTPELLIER 2

Discipline : Biologie des Populations et Ecologie – CNU 67

Ecole Doctorale : SIBAGHE

Présentée et soutenue publiquement

Par

Cédric DELSOL

Le 14 Décembre 2010

LE GENOME MITOCHONDRIAL DES MAMMIFERES ET LES DONNEES MOLECULAIRES HYPERVARIABLES DANS LA DESCRIPTION DE L'HISTOIRE EVOLUTIVE DES ARVICOLINAE

JURY

Mr Douzery Emmanuel

Président

Mr Galtier Nicolas

Directeur de Thèse

Mme Chenuil Anne

Rapporteur

Mr Pompanon François

Rapporteur

Sommaire

Sommaire	5
Introduction	7
Position Systématique – Caractéristiques paléontologiques de la radiation	7
Phylogénie moléculaire et datation de la radiation	9
Hypermutable mitochondriale des Mammifères	15
Bilan-Objectifs de la thèse	18
1 ^e Partie : Acquisition des mitogénomes complets	21
Echantillonnage taxonomique	21
Définitions d’amorces	21
Extractions d’ADN mitochondrial	23
Mise au point des conditions PCR	23
Purification de l’ADN amplifié	23
Réaction de séquence et séquençage	24
Assemblage des séquences	24
Etude de la variabilité des séquences	27
2 ^e Partie : Phylogénie des Arvicolinae	29
3 ^e Partie : Séquençage de masse	54
Redéfinition des fragments d’amplification	54
Purification des produits d’amplification	54
Dosage des produits d’amplification	56
Comparaison des deux méthodes de séquençage	56
Perspectives : Génomique comparative	58
Constitution des bases de données : mitogénomes à évolution lente	58
Comparaison des distances génétiques - Observation Polymorphisme/Divergence	58
Détection des points chauds mutationnels	60
Détection des points chauds partagés et non-partagés	62
Conclusion et perspectives générales	64
BIBLIOGRAPHIE	65

Introduction

Position Systématique – Caractéristiques paléontologiques de la radiation.

Les Rongeurs représentent actuellement 40% de la diversité spécifique des Mammifères. Bien que très largement discutée au cours du siècle dernier, l'organisation systématique de la super-famille Rodentia est aujourd'hui l'une des plus étudiées et décrites. En effet, une grande richesse du registre fossile, qui a permis la description d'organismes modèles en paléontologie des Mammifères, ainsi qu'une grande diversité morphologique, autorisant la description de caractères exclusifs à certaines familles de Rongeurs, ont permis l'organisation des espèces en clades reconnus, mais dont les relations restent encore à définir, à différents degrés.

La meilleure représentation de ce problème reste la question de la séparation des 2 super-familles Muridae et Cricetidae, qui regrouperaient, à elles seules, la plupart des sous-familles de rongeurs Muroides. Réellement identifiés comme tels à la fin du 20^e siècle, les Cricetidae regroupent à présent 6 sous-familles de tailles inégales. Les Cricetinae (7 genres, 18 espèces), les Tylomyinae (4 genres, 10 espèces) et les Lophiomyinae (1 genre, 1 espèce) sont les 3 sous-familles les moins diversifiées. Les Sigmodontinae (74 genres, 377 espèces), les Neotominae (16 genres, 124 espèces) et les Arvicolinae (28 genres, 151 espèces) regroupent plus de 95% de la diversité spécifique des Cricetidae. Dans le cadre paléontologique, le registre fossile riche en taxons, profond dans le temps et de répartition large géographiquement, ne permet pas de description d'un progéniteur de la famille, mais plutôt de différentes lignées aujourd'hui éteintes – en fait des sous-familles de Cricetidae : eucricetodontinés, cricetodontinés, paracricetodontinés et cricetopinés. Ces descriptions restent utiles pour la mise en avant de caractères morphologiques dérivés partagés par tous les Cricetidae (synapomorphies non-exclusives) ; par exemple, la morphologie des dents jugales est caractéristique de la famille avec un arrangement bisérié des cuspidés (plan Cricetin) et rétention d'une connexion longitudinale, ainsi que la formation d'un anteroconide discret sur les premières molaires. Bien entendu, ces caractères sont à l'état dérivé chez les espèces actuelles, faisant apparaître des convergences entre sous-familles de Cricetidae, mais également avec d'autres groupes Muroides, comme les Muridae ou les Nesomyidae.

La sous-famille des Arvicolinae regroupe des clades (ou tribus) peu diversifiés, comme les Ondatrini, regroupant 2 genres monospécifiques (*Neofiber alleni* et *Ondatra zibethicus*). En revanche, le genre *Microtus* représente, avec ses 60 espèces actuelles, la radiation spécifique la plus rapide et la plus récente chez les Mammifères de l'Hémisphère Nord, tout en regroupant presque la moitié des espèces d'Arvicolinae. Depuis *Allophaiomys* du Pliocène, considéré comme la forme ancestrale commune des *Microtus*, le genre aurait donc connu une trentaine de spéciations au cours des 5 derniers Millions d'Années. Cette radiation apparaît d'autant plus explosive que la répartition géographique du genre est Holarctique, et que les 60 espèces sont parfaitement identifiées du point de vue morphologique comme du point de vue de leur répartition géographique. Dans le cas du genre *Microtus*, les spéciations récentes soumettent les caractères dentaires à des changements très rapides, et les premières définitions de sous-genres de *Microtus* sont principalement basées sur ces caractères. Les sous-genres *Stenocranius*, *Alexandromys*, *Pitimys*, *Terricola*, *Lasiopodomys*, *Blandfordimys*, *Chionomys*, *Neodon*, *Phaiomys* ou encore *Proedromys*, sont à présent considérés comme *Microtus s.l.*, la convergence morphologique de certaines espèces, et donc la paraphylie de certains de ces groupes, étant soulignée par diverses analyses moléculaires.



Figure 1 : Quelques exemples d'Arvicolinae. De haut en bas, de gauche à droite : *Microtus ochrogaster*, *Arvicola amphibius*, *Arborimus longicaudus*, *Alticola argentatus*, *Ondatra zibethicus*, *Myodes glareolus*.

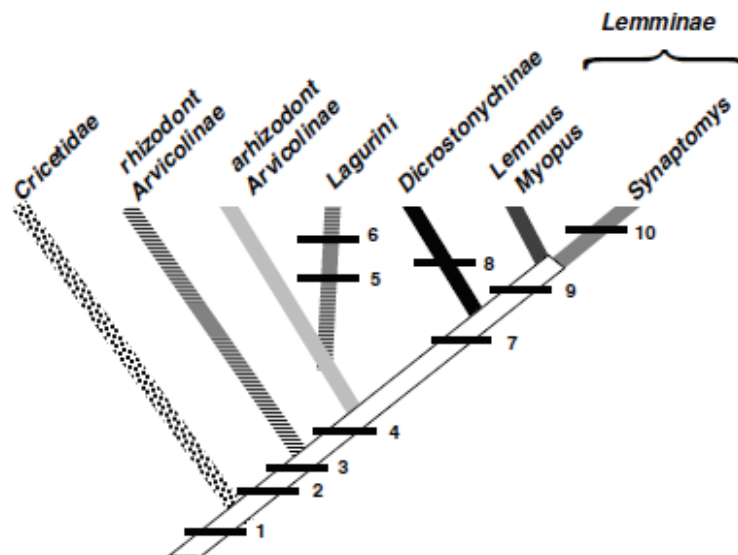


Figure 2 : (Courant F.,1997) Relations phylogénétiques des principaux groupes d'Arvicolinae d'après les caractères morphologiques et paléontologiques. Caractère 1 : Morphologie crânienne myomorphe ; Caractère 2 : Hypsodontie ; Caractère 3 : Nombre variable de triangles d'émail opposés ou alternés ; Caractère 4 : Absence de racines dentaires (apparition de l'arhizodontie) ; Caractère 5 : rupture du triangle d'émail 6 sur la M1 de *Lagurus* ; Caractère 6 : « Protubérance laguroïde » (un petit triangle lingual) sur les molaires supérieures ; Caractère 7 : Position des incisives inférieures ; Caractère 8 : polyisométrie sur la M1 ; Caractère 9 : M3 lophodontes ; Caractère 10 : Molaires à structure dissymétrique.

Phylogénie moléculaire et datation de la radiation

L'utilisation de données moléculaires pour la définition des relations phylogénétiques a permis la clarification de nombreuses situations problématiques, en particulier au sein des Mammifères, pour lesquels les radiations spécifiques sont nombreuses et concomitantes au cours du Quaternaire.

A un haut niveau taxonomique (supra-ordinal) chez les Mammifères, les données moléculaires nucléaires et mitochondriales ont corroboré des clades comme les Cetartiodactyles (De Jong, 1998) ou les Paenungulés (Lavergne et al., 1996), qui n'avaient pas été identifiés par la morphologie. Une comparaison de l'apport des données mitochondriales et des données nucléaires à un haut niveau taxonomique (Springer et al., 2001) chez les Mammifères souligne l'intérêt des informations phylogénétiques véhiculées par les séquences nucléaires pour définir les nœuds profonds. Les données mitochondriales sont considérées comme ayant moins de pouvoir résolutif à grande échelle, ce qui est une conséquence de la plus forte variabilité de ces séquences, et notamment de la saturation détectée sur la troisième position des codons (Lecompte E., 2008; Stepan S.J., 2005). Tout en recommandant l'utilisation de données nucléaires pour la résolution de nœuds profonds, les auteurs stipulent et illustrent le fait que la totalité du génome mitochondrial, incluant les ARN ribosomiques, possède plus de pouvoir résolutif que les 13 gènes codants mitochondriaux utilisés seuls. En revanche, les divergences plus récentes sont mieux résolues par les données mitochondriales que par les données nucléaires, du fait de leur niveau de variation plus élevé (Davis C.S., 2004; Hassanin A., 2009).

La complémentarité des données mitochondriales et nucléaires a permis plusieurs tentatives de résolution pour la phylogénie des Rongeurs, et ce en donnant la priorité soit à la résolution des relations basales au cœur des Muroides, soit à la résolution de radiations récentes, représentées par des sous-familles ou des genres de Rongeurs. Le problème majeur de ce type d'étude est d'obtenir à la fois une bonne représentativité taxonomique du groupe, et une quantité de marqueurs moléculaires suffisamment importante pour apporter les informations phylogénétiques nécessaires.

A l'échelle des Arvicolinae, les premières tentatives de résolution moléculaire sont exclusivement basées sur des données mitochondriales (Conroy et al., 1999). Ces premiers résultats rompaient la multifurcation basale causée par l'incertitude morphologique et paléontologique, en donnant une origine monophylétique aux Arvicolinae. C'est la première fois que deux phases de spéciations rapides sont décrites : une concernant les Arvicolinae, regroupant les lemmings, les rats musqués et les campagnols, et une concernant le genre *Microtus*, monophylétique également. L'utilisation de 28 taxons représentés chacun par environ 1100 bp du gène cytochrome b, permet de décrire précisément des radiations très récentes, même si l'introduction de sites saturés montre clairement une accumulation d'homoplasie, et par conséquent, une multifurcation basale des Arvicolinae.

Réalisée à une échelle taxonomique large, l'analyse de (Buzan E.V., 2008), avec 69 espèces (pour 18 genres et 10 tribus) au sein des Arvicolinae, permet de décrire l'histoire de la formation des différentes tribus (figure 4). L'utilisation du marqueur cytochrome b permet ici de donner une origine monophylétique à la plupart des tribus constituant la sous-famille, même si les positions d'*Arvicola terrestris* et d'*Ondatra zibethicus* ne sont pas résolues, ainsi que la polytomie basale des Arvicolinae.

L'apport d'informations phylogénétiques par les données mitochondriales au sein des radiations récentes est mis en avant au travers de deux études réalisées chez les Arvicolinae. Avec 48 espèces représentées par le marqueur cytochrome b dans leur analyse, (Jaarola M. et al., 2004) décrivent les monophylies de plusieurs clades du genre *Microtus*, réaffirmant l'existence de groupes morphologiques en tant que sous-genres (*Microtus*, *Terricola*, *Pallasinus*, *Alexandromys*, *Volemys*,...), et rapprochant de cette façon les clades biogéographiques européens (*Microtus*, *Terricola*) et américains (Espèces Néarctiques, ex : *Pedomys*), le clade asiatique étant leur groupe-frère (figure 5). Une étude moins complète au niveau taxonomique, avec 23 espèces représentées par le même marqueur mitochondrial (Bannikova A., 2010) souligne également la division du genre *Microtus* en sous-genres morphologiques et/ou biogéographiques. Les relations profondes entre ces sous-genres sont peu soutenues, exprimant la difficulté d'extraire une information phylogénétique consistante de données mitochondriales à haut niveau taxonomique. L'utilisation des gènes 12s et Cytochrome b dans la description de radiations à bas niveau taxonomique, par exemple les *Microtus* (*Terricola*) à distribution européenne, permet de refléter la distribution géographique actuelle (Brunet-Lecompte P. Tougard C., Fabre M., Montuire S., 2008; Ranvoisé E. Tougard C., Petitjean A., Quéré J.P., 2008), en accord avec les études réalisées à échelle plus large (Jaarola M. et al., 2004).

Ces constats sont renouvelés si l'on se penche sur l'apport des données nucléaires. (Lebedev V.S. Abramson N.I., Tesakov A.S., Bannikova A.A., 2009), avec l'utilisation des gènes nucléaires GHR et LCAT (introns et exons), retrouvent les principales monophylies décrites précédemment pour les tribus d'Arvicolinae (Arvicolini, Ellobiusini, Lagurini, Myodini) ainsi que celle du genre *Microtus*. Cependant les nœuds profonds, définissant les relations entre les différentes tribus, restent peu soutenus, et les positions d'*Arvicola*, en tant que groupe-frère des *Microtus*, ou encore d'*Ondatra*, ne sont pas clairement résolues.

La combinaison de marqueurs nucléaires et mitochondriaux au sein d'une analyse phylogénétique permet non seulement la prise en compte d'un plus grand nombre de caractères, mais également l'apport d'informations différentes et complémentaires, interprétées grâce aux modèles phylogénétiques adaptés. (Galewski T. and E., 2006) souligne l'utilité des séquences du gène mitochondrial cytochrome b quand il est combiné avec des séquences moins saturées. L'incorporation des données mitochondriales aux données nucléaires améliore la qualité du signal phylogénétique, permettant la description des tribus évoquées précédemment, avec un meilleur support (figure 6). En utilisant un échantillonnage taxonomique plus large chez les Arvicolinae (106 espèces, dont 95 Arvicolinae), les mêmes marqueurs nucléaires et mitochondriaux mais également 296 caractères morphologiques, (Robovsky J., 2008) établit une phylogénie très exhaustive de la sous-famille : 4 tribus monophylétiques (Dicrostonychini, Lemmini, Clethriomyini, Arvicolini) ainsi que 6 sous-genres de *Microtus* (*Neodon*, *Alexandromys*, *Mynomes*, *Lasiopodomys*, *Microtus s.str.*, *Terricola*). La quantité et l'incongruence des caractères pris en compte entraînent des incertitudes traduites par des multifurcations apparaissant dans les radiations récentes, soulignant les difficultés à combiner les informations morphologiques et moléculaires.

L'apport des données moléculaires permet, en outre, de réaliser des analyses en datation, afin de situer temporellement les divers événements de spéciation au cours de l'histoire évolutive des Arvicolinae. En fonction de la nature des marqueurs utilisés, mais également des points de

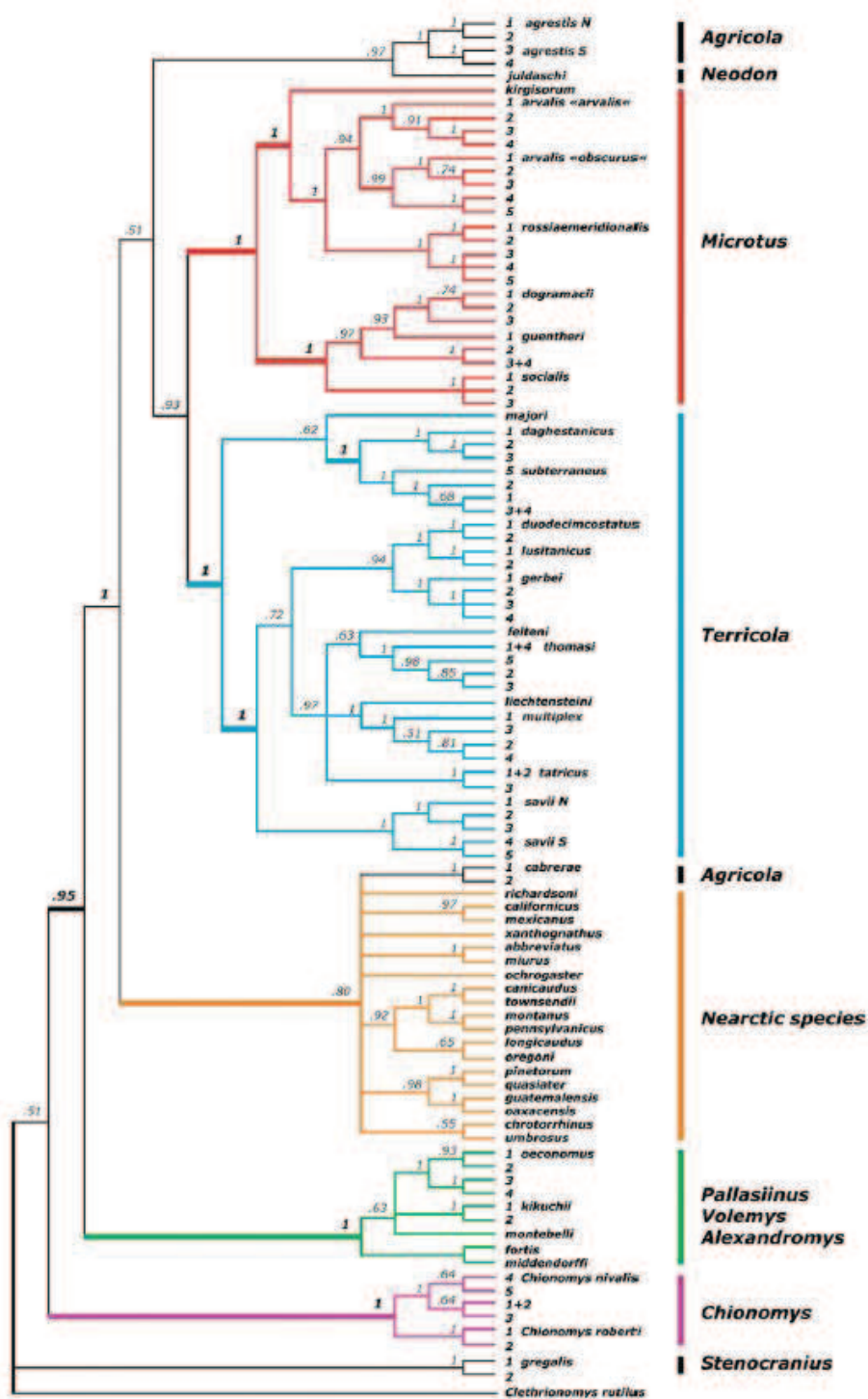


Figure 5 : (Jaarola M. *et al.*,2004) Arbre consensus (50% majoritaire) de 19000 arbres obtenus d'après une analyse bayésienne des haplotypes cytochromes b de 46 espèces de *Microtus* et 2 de *Chionomys*, raciné avec *Clethrionomys rutilus*. Les nombres au dessus des branches représentent les valeurs des probabilités postérieures (P>50).

calibration fossile définis dans l'analyse, la séparation Arvicolinae/Cricetinae est datée entre 13,5 et 18,8 Millions d'années (marqueurs nucléaires). La séparation *Chionomys/Microtus* est datée entre 2,3 et 6.5 Millions d'années si l'on considère les résultats obtenus à la fois avec des données mitochondriales et/ou nucléaires (Bannikova A.,2010; Buzan E.V.,2008; Steppan S.J.,2004). Ces résultats sont généralement concordants avec les données paléontologiques, mais les incertitudes morphologiques et sur les migrations réalisées par les espèces ancestrales, montrent la nécessité d'établir une analyse en datation moléculaire basée sur des calibrations multiples, à l'intérieur et à l'extérieur de la sous-famille Arvicolinae. Dans ce but, en plus d'un échantillonnage taxonomique complet, la constitution d'une base de données moléculaires plus riche en caractères se présente comme essentielle pour la définition de relations phylogénétiques plus solides chez les Arvicolinae, en particulier pour les nœuds apparaissant comme relativement peu soutenus dans les analyses précédentes : la position des Ondatrini en tant que groupe-frère des Arvicolini, ou encore la position du genre *Arvicola* en tant que groupe-frère des *Microtus*.

L'amélioration de la description de l'histoire évolutive des Arvicolinae avec le génome mitochondrial complet peut être réalisée à la condition d'utiliser des modèles d'évolution des séquences adaptés à l'hétérogénéité du taux d'évolution entre sites et entre lignées, qui exclut l'utilisation de méthodes appliquant l'horloge moléculaire aux données. La précédente étude comparative concernant marqueurs nucléaires et mitochondriaux des Arvicolinae (Galewski T. and E.,2006) met en évidence une évolution du gène Cytochrome b (mitochondrial) 11,6 fois plus rapide que celle du gène GHR (nucléaire), et cette vitesse d'évolution est plus visible sur les troisièmes positions des codons, respectivement 24 et 166 plus rapides que celle des premières et deuxièmes positions. Jusqu'à présent, peu d'études font intervenir les génomes mitochondriaux complets dans la définition des relations entre Rongeurs, mais les différentes études basées sur le marqueur cytochrome b font apparaître une évolution mitochondriale très rapide, en tout cas plus rapide par rapport aux génomes mitochondriaux représentés dans d'autres radiations de Mammifères. Ce phénomène est à la base de l'attraction des longues branches visibles sur des phylogénies réalisées sur le gène cytochrome b, liée à une perte de signal phylogénétique avec la saturation des données. Cette évolution rapide est plus particulièrement soulignée sur l'intégralité du génome mitochondrial au sein du genre *Microtus* dans des études récentes (Triant D.A. ,2006), où le taux de substitution nucléotidique entre les génomes entiers de deux *Microtus* (*M. kikuchii* et *M. rossiaemeridionalis*) apparaît comme plus élevé en comparaison avec d'autres couples d'espèces Mammifères congénériques (figure 7).

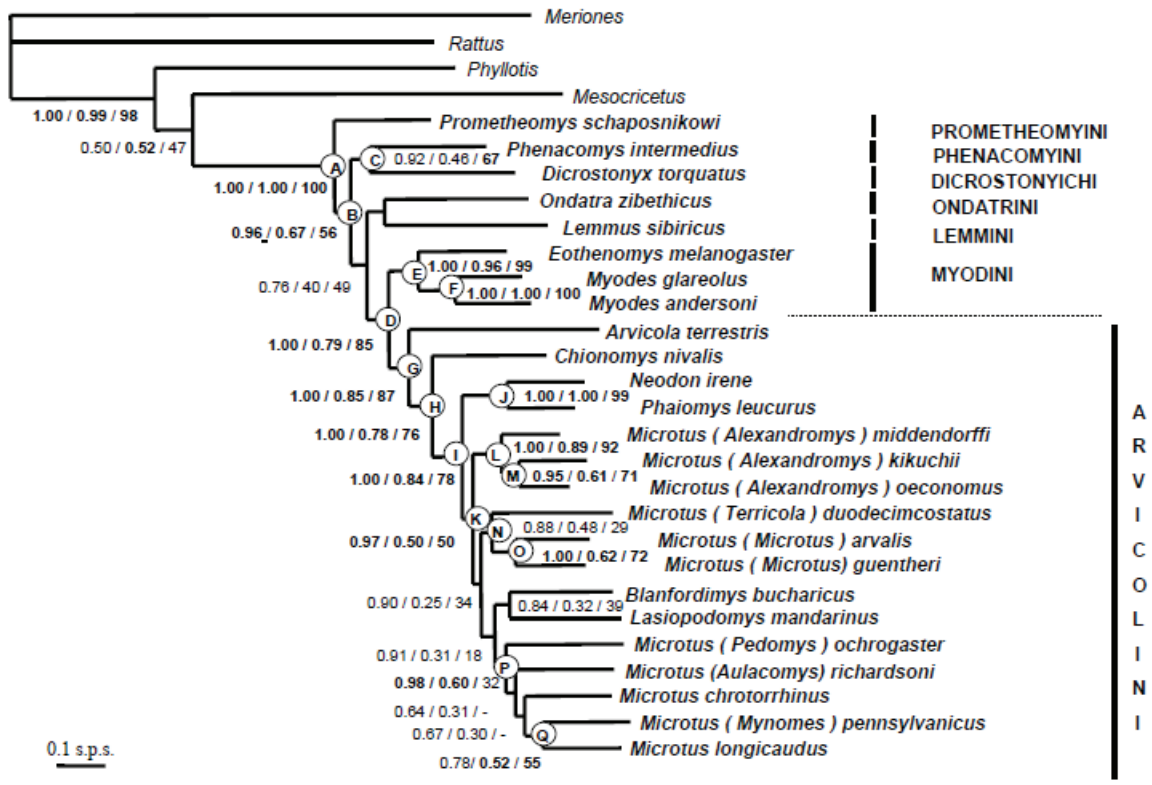


Figure 6 : (Galewski T. and E.,2006) Arbre obtenu en maximum de vraisemblance reconstruit pour la combinaison des séquences mitochondriale (cytochrome b) et nucléaire (GHR). Les 3 indices de fiabilité sont donnés sur les nœuds : les probabilités postérieures bayésiennes / les probabilités postérieures après bootstrap / les pourcentages de vraisemblance. Les lettres (A à Q) illustrent les nœuds retrouvés dans les 2 analyses : Inférence Bayésienne et Maximum de Vraisemblance.

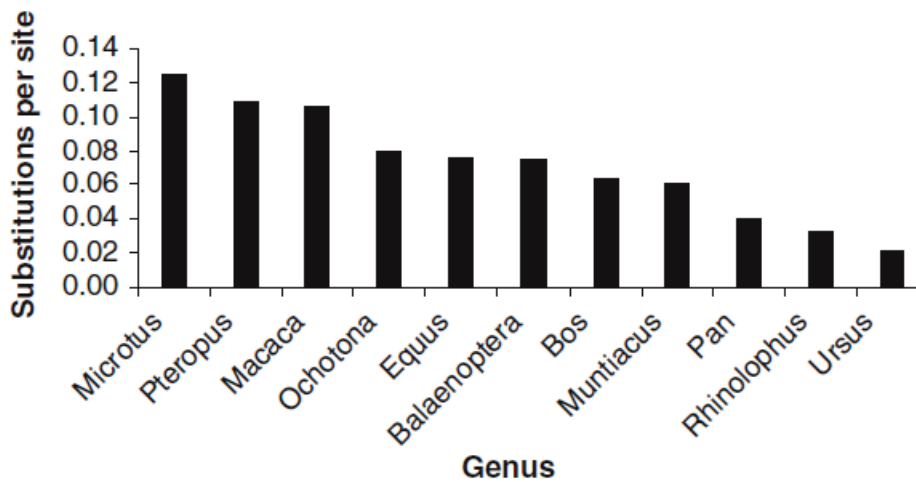


Figure 7 : (Triant D.A. ,2006) Nombre de substitution par site entre 2 séquences du génome mitochondrial de paires d'espèces Mammifères congénériques, sans correction par rapport au temps de divergence.

Hypermutable mitochondriale des Mammifères

Ces dernières années, la forte variabilité du génome mitochondrial à l'échelle des Mammifères a été démontrée, non seulement au travers de diverses études phylogénétiques, faisant apparaître un haut niveau d'homoplasie intra-générique, mais également par diverses études utilisant des méthodes de génomique des populations afin d'établir des causes possibles de cette hypermutabilité (Eyre-Walker A.,1999; Galtier N.,2006).

Cette hypermutabilité généralisée à tous les génomes mitochondriaux de Mammifères a cependant été nuancée par des études récentes, mettant en avant de fortes variations du taux de mutation mitochondriale entre les différents groupes de Mammifères (Nabholz B.,2007; Nabholz B.,2008; Triant D.A. ,2006) et remettant en cause l'utilisation d'horloge moléculaire dans les analyses phylogénétiques. Ces études, s'appuyant sur la diversité nucléotidique ou le taux de substitution mitochondrial, soulignent toutes le haut niveau de variabilité mitochondriale du groupe Rodentia en comparaison à des groupes dont l'évolution apparaît comme plus « lente » : Carnivores, Primates, (Cet-) Artiodactyles (figure 7 et 8). Cette disparité des taux de mutation mitochondriaux explique l'apport important de signal phylogénétique au sein d'études visant à la reconstitution de l'histoire évolutive des groupes « lents » évoqués précédemment en utilisant les génomes mitochondriaux complets : Phocidae (Davis C.S. ,2004; Fulton T.L.,2010), Caniformia (Arnason U.,2007; Delisle I.,2005), Bovidae (Hassanin A.,2009).

Il existe, à ce jour, 3 différentes hypothèses quant aux causes de variation de cette hypermutabilité, toutes liées aux traits d'histoire de vie très hétérogènes des groupes concernés.

L'hypothèse du temps de génération (Bromham L,1996; Li WH,1987; Mooers AO.,1994; Ohta T.,1993) lie un temps de génération court à un grand nombre de réplifications d'ADN, et donc d'erreurs de réplification, par million d'années. Cette hypothèse implique un taux d'erreur de réplification équivalent entre les espèces, et devrait affecter les génomes mitochondrial et nucléaire. Ainsi, les espèces à temps de génération court, répliquent plus fréquemment leur ADN mitochondrial, favorisant ainsi les substitutions (ou insertions) au sein de la séquence répliquée.

L'hypothèse du taux métabolique (Martin AP,1995; Martin AP,1992; Martin AP,1993; Nunn GB.,1998) lie le rôle central de la mitochondrie dans la fonction respiratoire cellulaire et l'hypermutabilité du génome mitochondrial. Les oxydations réalisées au sein de la mitochondrie sont génératrices de ROS (ou radicaux libres mutagènes) pouvant eux-mêmes être la cause de l'hypermutabilité mitochondriale. Une relation linéaire entre le taux métabolique et le taux de mutation est alors attendue.

La quantité de radicaux libres mutagènes produits au cours de la vie d'un organisme est également évoquée comme cause possible de l'hypermutabilité mitochondriale dans l'hypothèse de longévité. L'idée est qu'une pression de sélection négative pourrait contraindre le taux de mutation mitochondrial vers des valeurs faibles chez les organismes longévifs, en raison du coût des mutations somatiques (Nabholz B.,2008). La relation inverse entre concentration en radicaux libres et longévité

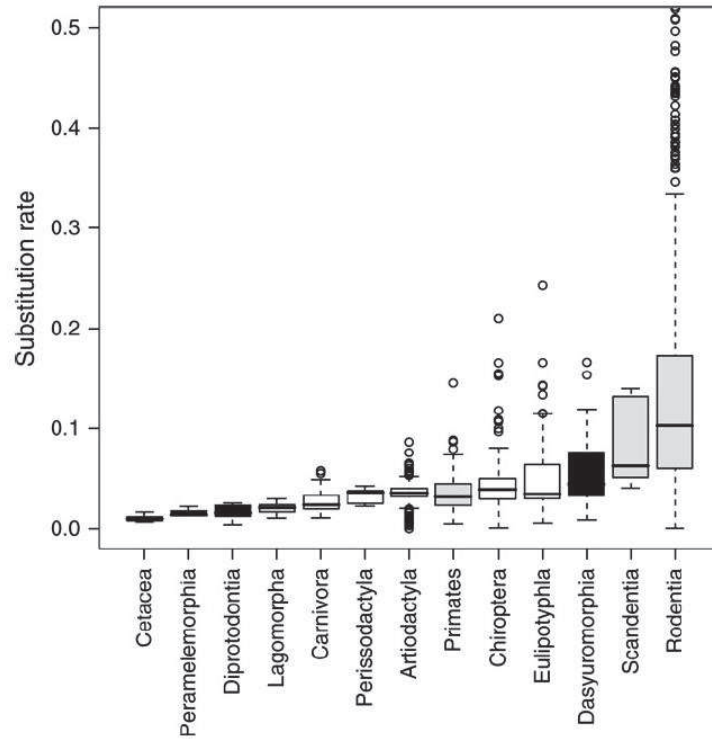


Figure 8 : (Nabholz B.,2007) Distribution intra-ordinale du taux de substitution neutre du gène cytochrome b. Les ordres représentés par plus de 5 espèces sont montrés. Les barres horizontales donnent la médiane des distributions du taux de substitution ; les boîtes donnent les quartiles ; les « moustaches » s'étendent à 1,5 fois la valeur des bornes quartiles. Les cercles représentent les données excentrées. La distribution chez les Rongeurs inclue des valeurs supérieures à 0,5, non montrées. Les taux de substitution sont donnés en substitution / position 3 du codon / Millions d'années

maximale des espèces (Barja G.,2000) vient confirmer les premières théories sur le sujet (Harman D.,1957).

Malheureusement, les prédictions de ces trois hypothèses sont toutes confirmées par les données, en raison de la forte corrélation entre traits d'histoire de vie chez les mammifères: les organismes longévifs ont un taux métabolique (par gramme de tissu) plus faible et un temps de génération plus long, en moyenne, que les organismes à durée de vie courte. La comparaison oiseaux/mammifères, en revanche, apporte un soutien à l'hypothèse "longévité" (Nabholz et al 2009 BMC Evol Biol).

Cette hypermutabilité n'est par ailleurs pas uniformément répartie le long du génome mitochondrial. La présence de sites hypervariables apparaît comme une caractéristique de la diversité nucléotidique des génomes mitochondriaux à l'échelle interspécifique ou intergénérique, notamment avec un nombre de sites polymorphiques dans la moitié ou plus des espèces d'un genre (observées le long du gène cytochrome b de Mammifères) plus élevé que le nombre de cooccurrences attendues sans hypothèse de point chaud mutationnel (Galtier N.,2006). De ce point de vue, fréquence et quantité de sites en situation de hot spots déterminent la variabilité et la diversité des séquences mitochondriales. Cette observation est à confronter avec les théories évoquées précédemment concernant le rôle des radicaux libres dans l'établissement de la diversité génétique mitochondriale.

Les points chauds mutationnels se définissent comme des sites nucléotidiques soumis à de fortes contraintes mutagènes, et donc à des substitutions multiples : ceci explique la présence de sites polymorphiques partagés entre espèces proches. Les points chauds mutationnels, s'ils sont fréquents à l'échelle d'un génome, peuvent expliquer l'excès d'homoplasie intraspécifique détecté dans le génome mitochondrial des Mammifères, ainsi que l'hétérogénéité des taux de substitution mitochondriale entre les différents groupes de Mammifères.

Les travaux existants sur les points chauds mutationnels s'appuient principalement sur l'analyse d'un seul marqueur mitochondrial, cytochrome b, qui représente moins de 10% des données moléculaires accessibles sur le génome. Il est donc nécessaire de généraliser ces observations à l'échelle des génomes mitochondriaux complets, afin d'identifier les différences de taux de mutation entre lignées et entre sites nucléotidiques mitochondriaux.

La constitution de la base de données de génome mitochondriaux d'espèces d'Arvicolinae va permettre : d'étudier la totalité de la diversité génétique du génome mitochondrial complet au sein d'un groupe considéré à évolution rapide ; de décrire précisément l'histoire évolutive de chaque site du génome mitochondrial au sein d'un groupe, et ainsi de détecter des sites mutants plusieurs fois sur une seule branche ; de comparer l'ensemble de ces éventuels hot spots avec ceux pouvant être décrits sur des radiations « lentes » ;

Bilan-Objectifs de la thèse

Le mitogénome des Arvicolinae est donc un marqueur intéressant pour résoudre la phylogénie du groupe, déjà caractérisée comme résultant d'une radiation rapide et récente, mais également en évolution génomique, où l'apport d'une base de données moléculaires constituée de génomes mitochondriaux dits à « évolution rapide », constituerait un support pour la comparaison avec des génomes dits à « évolution lente », et la description de points chauds mutationnels ;

Pourtant, en 2007, seuls 2 génomes étaient disponibles chez les Arvicolinae, et, plus généralement, 13 génomes mitochondriaux de rongeurs (sur 1244 métazoaires représentés, et 233 mammifères), alors que ceux-ci représentent 40% des espèces de mammifères (figure A). Ce biais d'échantillonnage, qui pourrait être dû à des difficultés techniques, elles-mêmes liées à la vitesse d'évolution élevée du génome mitochondrial des petits rongeurs, empêche pour le moment de caractériser l'histoire évolutive de ce groupe, et donc de ce génome. La figure B illustre cette hypothèse en corrélant le taux de substitution neutre détecté en troisième position des codons (Nabholz B., 2008) et la disponibilité des génomes mitochondriaux obtenus en laboratoire. Il apparaît que plus la variabilité détectée du génome mitochondrial est élevée, moins les génomes mitochondriaux de l'ordre ou du sous-ordre sont aisés à obtenir, probablement du fait de la difficulté de définir des amorces appropriées pour la PCR dans les taxons hypervariables.

Les objectifs de cette thèse sont donc :

D'obtenir le mitogénome d'une grande quantité d'Arvicolinae, pour augmenter à la fois le nombre de marqueurs et le nombre d'espèces représentées dans l'échantillonnage taxonomique, tout en dépassant les difficultés techniques liées à la haute variabilité nucléotidique du génome ;

D'obtenir une phylogénie robuste et datée de la totalité de la sous-famille. La restitution fidèle de l'histoire évolutive des Arvicolinae permet, à terme, de reconstituer l'évolution de ce mitogénome au cours du temps et des spéciations ;

De réaliser une étude mitogénomique de taxons à évolution rapide, en comparaison avec des mitogénomes à évolution lente.

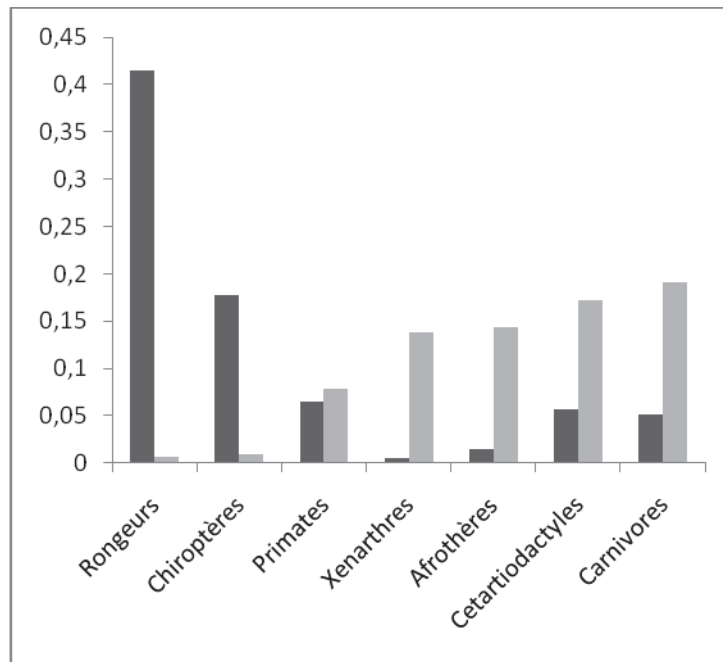


Figure 9 : Taxonomie et génome mitochondrial. Les rectangles gris foncés donnent la proportion d'espèces Mammifères présentées dans chaque ordre ou sous-ordre. Les rectangles gris clairs présentent le ratio entre le nombre d'espèce dont le génome mitochondrial complet est disponible et le nombre total d'espèces de l'ordre ou du sous-ordre.

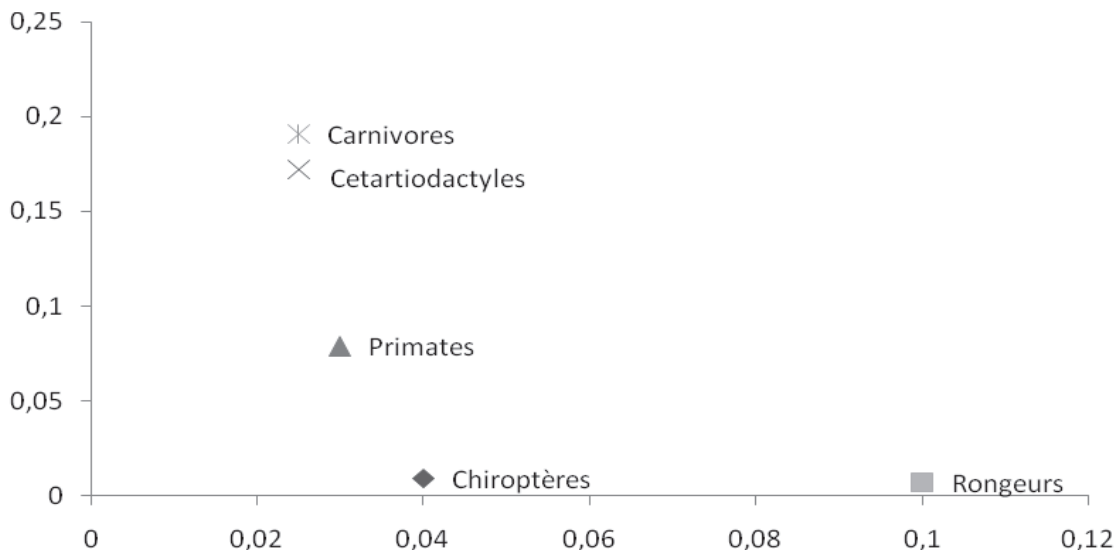


Figure 10 : Variabilité et séquençage. En abscisses, le taux de substitution neutre médian des ordres en substitution/site/millions d'années (Nabholz B.,2008). En ordonnées, le ratio entre nombre d'espèces dont le génome mitochondrial complet est disponible et le nombre total d'espèces de l'ordre ou du sous-ordre.

1^e Partie : Acquisition des mitogénomes complets

La partie expérimentale constitue une part considérable de cette thèse. Il s'agit d'obtenir la totalité de la séquence nucléotidique du génome mitochondrial de l'échantillonnage taxonomique le plus complet possible de campagnols et de rats musqués. A cette fin, les différentes étapes d'extraction d'ADN mitochondrial, d'amplification d'ADN par PCR, et des techniques de séquençage ont été optimisées de manière à compléter la base de données moléculaires mitochondriales des Arvicolinae. Ceci a été réalisé en dépit d'au moins deux inconvénients principaux : (i) la forte variabilité décrite du génome mitochondrial complet des Rongeurs, limitant l'observation de zones conservées à l'échelle des sous-familles et donc la définition d'amorces effectives sur la totalité des génomes mitochondriaux visés ; et (ii) la possibilité d'amplification de pseudogènes ou *numts*, nuclear-mitochondrial translocations, (Lopez JV., 1994), des copies nucléaires de gènes (ou portions de gènes) mitochondriaux, soumises à des pressions sélectives différentes du génome mitochondrial complet, et pouvant biaiser les données visées lors de l'assemblage des séquences.

Echantillonnage taxonomique

L'objectif est, dans un premier temps, de constituer l'échantillonnage taxonomique le plus complet possible à l'échelle des Arvicolinae. Différents tissus (muscles, organes vitaux), en provenance d'individus piégés et identifiés morphologiquement, ont constitué un support pour la mise au point de méthodes d'extraction d'ADN mitochondrial efficaces ainsi que pour la définition d'amorces conservées à un haut niveau taxonomique. Les premiers spécimens échantillonnés, *Arvicola terrestris* (origine : J-F. Cosson, CBGP) et *Microtus duodecimcostatus* (origine : F. Catzeflis, ISEM) ont ainsi permis d'établir la base de la méthodologie expérimentale exploitée au cours de cette thèse.

L'apport progressif de spécimens dans notre échantillonnage taxonomique a été réalisé grâce au piégeage d'un individu identifié comme *Microtus agrestis*, ainsi que par la constitution d'une banque de tissus, grâce à des contributeurs français et internationaux : F. Catzeflis (ISEM), J-F. Cosson (CBGP), J-P. Quere (CBGP), L. Olson (UAM), le Smithsonian Institute (Washington, D.C., Etats-Unis), et le Museum of Vertebrate Zoology (Berkeley, Etats-Unis).

Les informations concernant l'échantillonnage taxonomique complet représenté dans notre banque de tissus sont résumées dans l'Annexe 1. Cet échantillonnage taxonomique représenté dans notre banque de tissus balaie 19 genres et 55 espèces d'Arvicolinae. A terme, l'extraction et le séquençage du génome mitochondrial de 36% des espèces d'Arvicolinae et 68% des genres d'Arvicolinae pourraient être réalisés.

Définitions d'amorces

Afin d'assurer l'efficacité des amorces PCR utilisées pour l'amplification du mitogénome complet à l'échelle des Arvicolinae, les deux génomes complets de *Microtus kikuchii* (NC003041, Institute of Molecular BioSciences, Nouvelle Zélande) et *Microtus rossiaemeridionalis* (NC008064, Purdue University, Etats-Unis) ont été alignés afin de définir des zones conservées, à l'échelle générique, susceptibles d'offrir un minimum de 18 nucléotides consécutifs peu variables d'une espèce à l'autre. Ces zones définies, il s'agit de définir des couples d'amorces permettant d'amplifier des fragments d'une longueur comprise entre 1000 et 2000 paires de bases (pb), chevauchants d'au moins 150 pb, sur l'ensemble des 16000 pb constituant le génome mitochondrial complet (tableau 1).

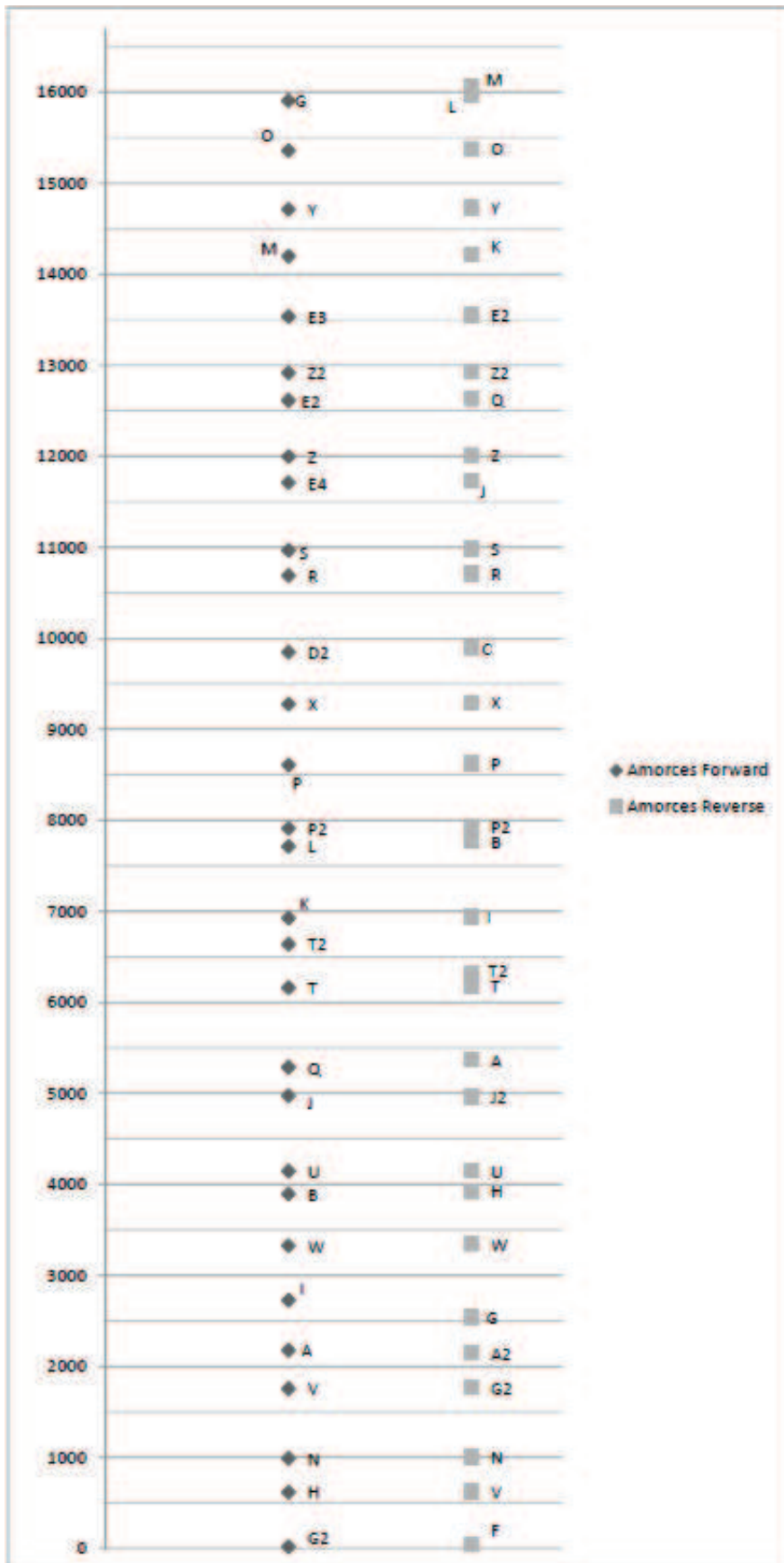


Tableau 1 : Couples d'amorces définis et utilisés lors de cette étude.

Il est important de noter que toutes les parties du génome peuvent être amplifiées par plusieurs couples d'amorces différents, ce qui permet de s'adapter d'une part, à la variabilité élevée des séquences mitochondriales (donc à la possibilité d'échec d'une amplification chez un individu où les zones conservées s'avèrent divergentes), d'autre part, à l'isolement des *numt*, qui peuvent être amplifiés et comparés aux séquences codantes mitochondriales.

Extractions d'ADN mitochondrial

Les extractions d'ADN mitochondrial doivent être redéfinies afin d'obtenir une quantité suffisante d'ADN mitochondrial destinée à subir les amplifications par PCR, à partir d'une quantité limitée de tissus (20 ± 10 mg).

Les kits d'extraction, faisant intervenir plusieurs passages sur colonnes filtrantes, révèlent un faible rendement, avec des quantités d'ADN mitochondrial extraites insuffisantes pour obtenir un résultat satisfaisant après amplification par PCR. Les méthodes d'extraction « chimiques », notamment le protocole classique Phénol-Chloroforme, se montrent beaucoup plus efficaces dans la quantité d'ADN mitochondrial extrait, offrant une concentration en ADN si forte qu'elle peut être saturante au cours d'une réaction PCR. Par la suite, la plupart des échantillons d'ADN sont alors dilués au 50^{ème} ou au 100^{ème}, et testés en amplification au moyen d'un couple d'amorces « universel » (Nf Gr).

Mise au point des conditions PCR

Les PCR ont été conduites sous des conditions standards, excepté pour l'application d'un gradient de température d'hybridation sur toutes les cycles d'amplifications par PCR testées. Ainsi, un couple d'amorces est testé en PCR sous 3 températures d'hybridation différentes (50, 55, et 60°C), permettant une flexibilité dans la fixation des amorces à un niveau sous-familial. La variabilité importante des séquences mitochondriales à l'échelle des Arvicolinae implique que les zones conservées définies par comparaison intra-*Microtus*, peuvent différer de deux nucléotides ou plus entre les différents genres constituant la sous-famille. La baisse de la température d'hybridation augmente les chances de fixation de l'amorce, même si les deux séquences, amorces et cibles, ne sont pas parfaitement complémentaires.

Purification de l'ADN amplifié

La purification de l'ADN amplifié est nécessaire avant d'en réaliser le séquençage. Après avoir testé des méthodes de purification sur gel d'électrophorèse, faisant intervenir des passages sur colonne filtrante, les quantités d'ADN purifié apparaissent comme insuffisantes pour réaliser une lecture des séquences satisfaisante. Ici encore, l'utilisation de colonnes filtrantes semble baisser le rendement en ADN amplifié, et la purification par voie chimique semble être la solution la plus conservatrice en matrice. L'utilisation d'EXOSAP (USB), un mélange enzymatique réalisant la purification des produits de PCR, permet d'obtenir une solution pure et fortement concentrée en séquence cible de la PCR.

Réaction de séquence et séquençage

Les réactions de séquence sont réalisées sur les produits d'amplification purifiés, un amplicon de 2000 pb pouvant subir jusqu'à 8 amplifications au moyen d'amorces internes. Les séquences sont alors précipitées au moyen d'une méthode standard (Ethanol/Sels : EDTA, Acétate de Sodium), et séquencées en fonction de la profondeur de lecture souhaitée : Séquenceur ABI PRISM 310, monocapillaire, profondeur de lecture 500 bp environ, ou Séquenceur 16 capillaires, profondeur de lecture 800 bp environ.

Assemblage des séquences

Les séquences sont ensuite exportées et alignées sous Sequencher. Tous les amplicons assemblés peuvent représenter la séquence mitochondriale complète, mais certains d'entre-eux sont divergents sur les mêmes fragments de séquence. Les deux (ou plus) séquences en contradiction sont alors isolées, et alignées avec les séquences mitochondriales complètes évoquées précédemment (*Microtus kikuchii* et *M. rossiaemeridionalis*) afin d'en réaliser l'alignement protéique, et ainsi d'identifier un fragment abritant éventuellement une rupture du cadre de lecture, alors considéré comme portion de séquence *numt* devant être exclu de la base de données. Cette méthode d'exclusion des *numt* est permise par l'utilisation d'amplicons du génome mitochondrial suffisamment longs qui incluent une partie codante, et par le fait que cette partie codante est rapidement non-fonctionnelle quand elle est copiée dans le génome nucléaire (*numt*) (Triant D.A., 2007). Si les séquences ambiguës ne comportent pas de fonctions codantes, l'alignement évoqué précédemment est utilisé dans une analyse en maximum de vraisemblance, grâce au logiciel PhyML (Guindon S., 2003) avec application d'un modèle GTR+I+ Γ sur les données. Sur le phylogramme obtenu, la séquence la moins divergente des données mitochondriales avérées est alors considérée comme authentique pour le génome mitochondrial complet.

L'intégralité de cette méthode s'appuie sur le constat qu'une séquence fonctionnelle (du mitogénome), conservée, cumule moins de substitution que sa copie (nucléaire) non fonctionnelle. D'un autre côté, la copie de la portion du génome mitochondrial, une fois intégrée au génome nucléaire, est soumise à un taux évolutif moins élevé, à cause de la différence de taux de substitution entre génomes nucléaires et mitochondriaux. Le fragment mitochondrial et sa copie nucléaire sont donc soumis à des forces évolutives différentes, et ces différences sont visibles sur les séquences actuelles, aussi bien sur un alignement nucléotidique, que sur un phylogramme (Triant D.A., 2008).

La totalité de la séquence mitochondriale assemblée est alors alignée avec les séquences de référence ainsi que les séquences mitochondriales obtenues au cours de l'étude. La séquence protéique de chaque gène codant est vérifiée. Les ARNt (23), ARNr (12S et 16S) et la Région de

contrôle (CR) sont alignés par similitude avec les séquences de référence (*Microtus kikuchii* et *M. rossiaemeridionalis*).

Cette thèse aura permis l'obtention de 18 mitogénomes complets, utilisés dans l'analyse phylogénétique et la datation moléculaire de la radiation évolutive des Arvicolinae. Les séquences mitochondriales d'autres espèces sont encore incomplètes ou en cours d'assemblage, soit par extraction de quantité d'ADN mitochondrial insuffisante, soit par échec des amplifications sur des zones importantes du génome. Un bilan des séquences obtenues est présenté sur la figure 11.

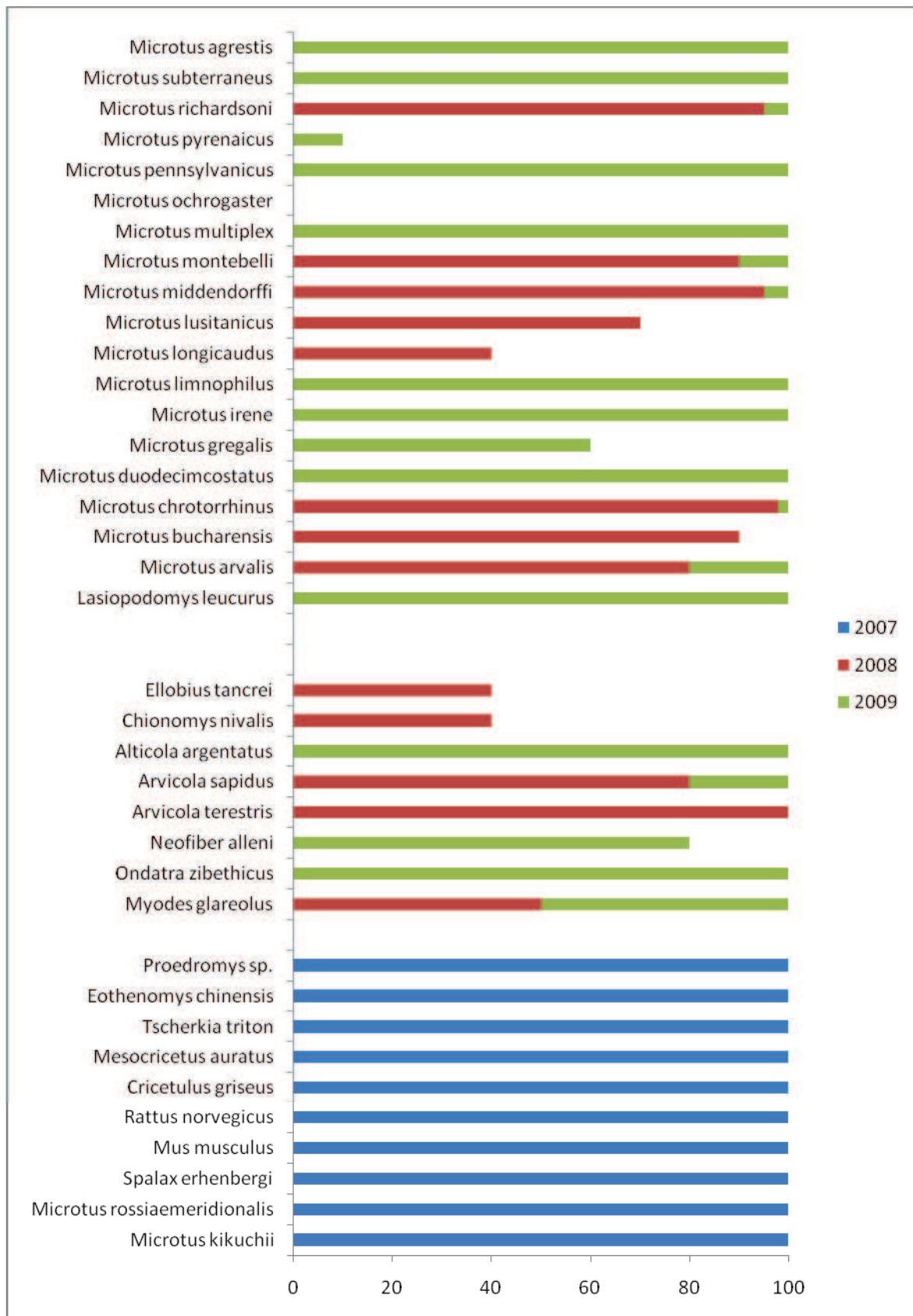


Figure 11 : Etat d'avancement du séquençage du génome mitochondrial complet d'espèces d'Arvicolinae au cours de cette thèse.

Etude de la variabilité des séquences

La méthode utilisant l'algorithme *super distance matrix*, ou SDM (Criscuolo A., 2006), permet, par la combinaison des matrices de distance obtenues d'après l'analyse phylogénétique réalisées sur chaque gène, de restituer des taux d'évolution relatifs, propres à chacun des gènes composant le mitogénome.

Les résultats de cette analyse sont présentés sur le tableau 4. L'hétérogénéité des taux d'évolution est détectable à l'échelle d'un seul mitogénome, ici celui des Arvicolinae. Les gènes ribosomiques 12S et 16S apparaissent comme peu variables, en comparaison avec les gènes ND1 et ND2, évoluant respectivement 9,6 et 8,6 fois plus vite que 12S. Ces résultats soulignent l'hétérogénéité de la variabilité du mitogénome des Arvicolinae à l'échelle de différents marqueurs moléculaires. L'hétérogénéité du taux de substitution mitochondrial des Mammifères peut donc être liée non seulement à la taxonomie, mais également au marqueur moléculaire concerné. Le gène Cytochrome b, sur lequel la plupart des études s'appuient pour affirmer la forte variabilité du génome mitochondrial des Mammifères, apparaît comme 5 fois plus variable que 12S chez les Arvicolinae, et il n'est cependant pas le marqueur le plus variable du mitogénome : les gènes ND1 et ND2 apparaissent comme les plus variables du mitogénome des Arvicolinae et, à ce titre, amplifient le phénomène d'hétérogénéité des taux de substitution entre mitogénomes de Mammifères.

L'intégration de la totalité de ces données hypervariables au sein d'une étude phylogénétique à par la suite été réalisée, au moyen de méthodes phylogénétiques novatrices permettant l'efficacité dans l'extraction d'informations robustes à partir de données moléculaires généralement présentées comme saturées, et donc peu informatives.

Gènes	Position sur le mitogénome	Taux relatifs (SDM)	Taux relatifs par rapport au marqueur le plus lent (12S)
12S	1	0,22427072	1
16S	2	0,53591155	2,4
ND1	3	2,14859237	9,6
ND2	4	1,92911947	8,6
COX1	5	1,20068557	5,4
COX2	6	0,98217656	4,4
ATP8	7	1,47317159	6,6
ATP6	8	1,82000377	8,1
COX3	9	1,35924201	6,1
ND3	10	1,77982943	7,9
ND4L	11	1,85120134	8,3
ND4L	12	1,85829294	8,3
ND5	13	1,74146854	7,8
ND6	14	1,71218831	7,6
CYTB	15	1,26056123	5,6
DLOOP	16	0,77843849	3,5

Tableau 4 : Variabilité des séquences codantes mitochondriales chez les Arvicolinae.

2^e Partie : Phylogénie des Arvicolinae

L'utilisation des données moléculaires mitochondriales acquises au cours de cette thèse permet d'établir un scénario évolutif tangible au sein de la radiation adaptative des Arvicolinae d'après une quantité de caractères conséquente (environ 16000 pb). L'utilisation d'une telle quantité de caractères, couplée à un échantillonnage taxonomique relativement complet au niveau tribal, implique l'utilisation de méthodes d'inférence phylogénétique permettant d'éviter les biais de reconstruction causés par la difficulté des modèles d'évolution et de leurs paramètres à extraire les informations de sites dont les taux de variation sont hétérogènes et variables au cours du temps.

Dans ce cadre, l'utilisation d'une méthode novatrice d'inférence bayésienne, permet d'appliquer différents regroupements de sites en fonction de leur pattern évolutif, ainsi que différents modèles d'évolution aux partitions du génome ainsi définies.

L'article présenté par la suite expose les méthodes utilisées, ainsi que les résultats importants et nouveaux obtenus dans l'étude :

- Monophylies des principaux clades d'Arvicolinae, au niveau générique et tribal ;
- Définition robuste des nœuds profonds et récents au sein de cette radiation rapide ;
- Monophylies des tribus géographiques de *Microtus* ;
- Description temporelle de la radiation adaptative des Arvicolinae ;
- Apport du génome mitochondrial complet dans l'étude de la radiation.

Last revision by ED — 11/03/2011 13:03:54

Manuscript intended for *Molecular Phylogenetics and Evolution* — Nombre de mots = 4990.

Phylogeny, molecular dating and biogeography of voles (Arvicolinae): a mitogenomics approach

Cédric DELSOL, Benjamin MARTEAUX, Nicolas GALTIER & Emmanuel J. P. DOUZERY ^(1, 2)

⁽¹⁾ Université Montpellier 2

⁽²⁾ CNRS, Institut des Sciences de l'Evolution (UMR 5554)

CC064 - Place Eugène Bataillon

34 095 Montpellier Cedex 05 — France

Running title: Voles mitogenomics

Keywords:

^(*) Corresponding author: Tel = 33 4 67 14 48 63 / Fax = 33 4 67 14 36 10 / e-mail = emmanuel.douzery@univ-montp2.fr

Abstract

Background

Mitochondrial genes are usually employed to resolve relationships within recently evolved groups, whereas nuclear genes are used to resolve phylogenies at deeper level. In cases of rapid and recent evolutionary radiations, the two most used mitochondrial markers, cytochrome b (CYTB) and 12S rRNA subunit (12S), still lead to poorly resolved trees, especially on deep nodes. The rapid and recent radiation of *Microtus* genera, representing ground voles of the Arvicolinae subfamily (Rodentia, Muroidea, Cricetidae), took place in the Northern Hemisphere between 2 and 5 Mya, and led to 60 extant species. Recent studies on the mitochondrial genome of *Microtus* species show high variability of sequences on an interspecific level, in comparison with other congeneric mammalian species. Here, we show how the whole mitochondrial genome can be employed in molecular systematics, in order to describe fast-evolving radiation coupled with high diversity on a taxonomic level.

Results

We provide 19 new sequences of the whole mitochondrial genome for species belonging to Arvicolinae. The 13 protein coding and the two rRNA genes were aligned to infer phylogenetic relationships using maximum likelihood and Bayesian inference. Thus, this allows evaluating the relative contribution of each marker for describing the phylogenetic relationships among the geographical lineages of *Microtus* species, and also among the main genera of the subfamily. This relative contribution of each marker is also taken into account in molecular dating of main splits in Arvicolinae radiation.

Conclusion

Introduction

Rodentia, which represents 42% of the ~5000 species of mammals, is by far the most diverse mammalian order. This high level of species richness resulted from a number of explosive radiations, which occurred independently in distinct lineages, so that several families or subfamilies of rodents include more than 100 species, although they originated during the last 10 or 15 million years (Wilson and Reeder, 2005).

Arvicolinae (voles, muskrats, and lemmings; 28 genera, 151 species) are an example of such an impressive, recent placental radiation. High specific diversity in this group has been acquired in a short period of time, as the product of a complex evolutionary history, often described as a “star-shaped” radiation. To date, paleontological data and phylogenetic reconstructions suggest that two successive speciation pulses have promoted the diversification of extant arvicolines. The first one, which might have occurred roughly 10 million years ago (Mya), made the split between current, and extinct, arvicolinae genera (Chaline *et al.*, 1999; Stepan *et al.*, 2004). The relationships among genera are still largely unresolved on both morphological and molecular grounds. The second radiation presumably occurred between 5 and 2 Mya, among species of the *Microtus* genus (Buzan E.V., 2008; Jaarola M. *et al.*, 2004; Stepan S.J., 2004; Triant D.A., 2006). This genus includes 60 extant species, i.e., nearly half the total number of Arvicolinae species. The second pulse of speciations is impressive because it means that 30 speciations per Myr could have occurred in this genus.

Regarding biogeography, Arvicolinae are restricted to the Northern Hemisphere. Some current genera (e.g. *Myodes*) have a distribution widespread through the Holarctic area, while some other are restricted to Palearctic regions (*Arvicola*, *Alticola* or *Eothenomys*) or Nearctic regions (*Neofiber*). *Microtus*, or “meadows voles”, are generally found in grassy areas of North America, Europe and Asia. The current biogeographic distribution of this genus is considered as the result of an extremely fast, very recent Holarctic radiation (Conroy C.J., 1999).

So far, the molecular phylogeny of Arvicolinae has been approached using a small number of nuclear or mitochondrial genes. The mitochondrial cytochrome b (cytb) and has proved useful for species delineation and study of recent diversification (Bannikova A., 2010; Conroy C.J., 2000; Brunet-Lecompte P. Tougard C., Fabre M., Montuire S., 2008), while the nuclear GHR, LCAT have helped clarifying to some extent the earliest divergences in the subfamily (Lebedev V.S. Abramson N.I., Bannikova A.A., Tesakov A.S., 2009; Lebedev V.S. Abramson N.I., Tesakov A.S., Bannikova A.A., 2009). The combination of the two kinds of markers yielded a plausible evolutionary scenario for Arvicolinae

(Galewski et al 2006), but a number of issues remain unresolved, among which (i) the phylogenetic relationships between Arvicolinae genera, (ii) the origins, monophyly and radiation of geographic groups within *Microtus*, and (iii) the timing of generic and specific divergence in this group. Regarding the dating issue, it is noteworthy that nuclear and mitochondrial genes have been reported to be in strong conflict (Conroy C.J.,1999; Steppan S.J.,2005), the latter supporting an Arvicolinae / Murinae split twice as old as what the former suggests.

Obtaining a robust phylogenetic and chronologic framework for Arvicolinae would be of interest for several reasons. First, it would help quantifying and qualifying the rate and pattern of diversification of this remarkably species-rich group. Secondly, a robust time frame would be useful for the study of molecular evolutionary rates. Arvicolinae have been identified as one of the most fast-evolving mammalian groups as far as *cyt b* sequences are concerned (Nabholz et al 2008), in relationship with the small size and short lifespan of these animals. A deeper investigation of these intriguing effects would require a proper dating of divergences within the group.

In this study, we approach the diversification of Arvicolinae using whole-genome mitochondrial data. Mitogenomic data have proven efficient for resolving recent evolutionary radiations, in mammals as well as in other taxonomic groups (Davis C.S. ,2004; Delisle I.,2005; Fulton T.L. ,2006; Hassanin A.,2009; Xu X.,1996). Our hope is to substantially improve the level of resolution of both branching orders and divergence times in Arvicolinae, thanks to a ten-fold increase in length of the mitochondrial sequence data set.

Material & Methods

Taxonomic sampling and molecular biological methods

Our study is based on a taxonomic sample constituted of 27 species for 12 genera of Arvicolinae. In particular, the *Microtus* genera radiation is represented by 16 species; including *Lasiopodomys (Microtus) leucurus* (see Table 1). Several species for which mitogenome data were available were incorporated as outgroup: *Spalax erhenbergi*, *Mus musculus musculus*, *Rattus norvegicus*, *Cricetulus griseus*, *Eothenomys chinensis*, *Tscherkia triton* and *Proedromys sp.*.

DNA extractions were performed on a minimum amount of 10 mg ethanol-preserved tissues, using a standard Phenol/Chloroform protocol.

Reference sequences of *Microtus rossiaemeridionalis* (NC008064) and *Microtus kikuchii* (NC003041) were aligned using MEGA4 in order to design PCR primers on well conserved and at least 18 nucleotides long, part of mitochondrial genome. This work led to a primer bank, constituted of nearly 25 couples of primers, potentially amplifying fragments between 1000 and 2000 bp, with a minimum overlapping of 250 bp (data not shown). All possible combinations of primers (forward and reverse) were tested on one specimen of *Microtus* and one specimen of *Arvicola*. Couples of primers effective on both DNA extractions were then selected to be used on other species of the sample.

PCR conditions were 95°C 5 min, followed by 95°C 30 sec (35 times), then 55°C ± 5°C 30 sec (35 times), then 72°C 1 min (35 times), and a final extension at 72°C 5 min. Hybridization temperatures are tested on each fragment, by gradient-PCR, in order to manage variable specificity of primers.

PCR products were purified using EXOSAP-IT enzyme mix (USB) and sequenced on both strands using automatic sequencing (Big Dye Terminator cycle kit) on an ABI 310 (PE Applied Biosystems) or on a 16 capillary sequencer (Modèle ?).

In total, 19 taxa have their mitochondrial genome sequenced in our study. These Arvicolinae sequences, which double the database of Muroid mitochondrial whole mitochondrial genome sequences, have been deposited in the EMBL databank (see table 1). For *Neofiber alleni*, the sequencing is still in progress, and its mitogenome is represented by a partial sequence.

Genus	Specie	Accession
<i>Spalax</i>	<i>erhenbergi</i>	NC005315
<i>Mus</i>	<i>musculus</i>	FJ803909
<i>Rattus</i>	<i>norvegicus</i>	GU997608
<i>Mesocricetus</i>	<i>auratus</i>	NC013276
<i>Cricetulus</i>	<i>griseus</i>	NC007936
<i>Tscherkia</i>	<i>triton</i>	
<i>Eothenomys</i>	<i>chinensis</i>	NC013571
<i>Proedromys</i>	<i>sp</i>	NC013563
<i>Microtus</i>	<i>kikuchii</i>	NC003041
<i>Microtus</i>	<i>rossiaemeridionalis</i>	NC008064
<i>Neofiber</i>	<i>alleni</i>	this study
<i>Ondatra</i>	<i>zibethicus</i>	this study
<i>Alticola</i>	<i>argentatus</i>	this study
<i>Myodes</i>	<i>glareolus</i>	this study
<i>Arvicola</i>	<i>sapidus</i>	this study
<i>Arvicola</i>	<i>terrestris</i>	this study
<i>Lasiopodomys</i>	<i>leucurus</i>	this study
<i>Microtus</i>	<i>irene</i>	this study
<i>Microtus</i>	<i>kikuchii</i>	this study
<i>Microtus</i>	<i>montebelli</i>	this study
<i>Microtus</i>	<i>limnophilus</i>	this study
<i>Microtus</i>	<i>middendorffi</i>	this study
<i>Microtus</i>	<i>richardsoni</i>	this study
<i>Microtus</i>	<i>chrotorrhinus</i>	this study
<i>Microtus</i>	<i>pennsylvanicus</i>	this study
<i>Microtus</i>	<i>agrestis</i>	this study
<i>Microtus</i>	<i>arvalis</i>	this study
<i>Microtus</i>	<i>subterraneus</i>	this study
<i>Microtus</i>	<i>duodecimcostatus</i>	this study
<i>Microtus</i>	<i>multiplex</i>	this study

Table 1: Taxonomic sampling and accession numbers.

Sequence alignment and phylogenetic analyses

A preliminary sequence alignment of the complete mitogenomes was obtained by MAFFT (Katoh K., Press, Humana, 2009). Minor refinements were then introduced with the ED editor of the MUST package (Philippe H., 1993), including (i) the removal of the end of the control region (599 sites) owing to its too high variability among arvicolines, cricetines, and murines ; and (ii) the removal of 808 sites which were gapped in more than 5% of the species. The final alignment is 15,860 sites long. Protein-coding and rRNA sequences were identified through their similarity with the known mitochondrial map of *Microtus kikuchii* and *Microtus rossiaemeridionalis*. All protein-coding sequences respected open reading frames.

The best-fitting model identified for the complete mitogenomic alignment under the AIC criterion with ModelTest (Posada D., 1998) was GTR, with a Gamma (Γ) distribution plus a fraction (I) of invariable sites for the among-sites substitution rate heterogeneities. Maximum likelihood analyses were conducted with PAUP* (Swofford D.L., Associates, Sinauer, 2003), version 4b10, with a neighbor joining starting tree refined by tree-bisection reconnection (TBR) branch swapping. All base frequency, nucleotide exchangeability, and rate heterogeneity parameters were estimated from the data. Node support was estimated by 100 replicates of bootstrap, conducted in the same conditions as previously described.

Bayesian Inference was realized using PhyloBayes 3.2d (Lartillot N., 2009), allowing the use of non-parametric methods for modeling site-specific features of sequence evolution. Rather than grouping each specific position of codons in a single mixture models category, the program allows using two combined, and independent, Dirichlet process mixtures, where the number of categories in a mixture model is not directly specified, but integrated over a prior distribution. The first defines rates for each site of the alignment, the second one aimed at describing the site-specific profiles. By the definition, for each site, of a rate and a frequency vector profile over the 4 bases, we assume the fact that each position of a mitochondrial gene is under very specific selective constraint, and realize the Bayesian Inference applying specific evolution models on each site.

Variable length bootstrap

To evaluate the number of mitochondrial nucleotide sites required to resolve with high support the various node of the Arvicoline tree, we used a resampling method called Variable Length Bootstrap (VLB). With this method, a variable number of character are resampled, and bootstrap support are estimated on each node of the reference tree. On the whole mitochondrial alignment,

nucleotides sites were resampled to generate pseudomatrices of 500, 1000, and until 16000 (15500) characters with increasing steps of 500 sites. All bootstrap searches were then performed using ML analyses with PAUP.

Molecular dating

Molecular divergence time estimates implies the definition of at least one calibration point, giving lower and upper bounds to one speciation, represented by a node in the phylogenetic tree. These calibration points are then taken into account by the model (with or without molecular clock), to infer dates of speciation represented by each node in the tree.

One well defined speciation in our study is the split between *Ondatra* and *Neofiber* genera, for which both molecular and paleontological data are available. As first real Arvicolinae forms are found from 5 Mya, *Pliopotamys* is considered as the oldest ancestor of *Ondatra* modern forms, and is found in Pliocene sediments (5,3-2,5 Mya) (Chaline J.,1999). Several paleontological and morphological studies underlines that the *Neofiber/Ondatra* split is older than 4 Mya (Repenning C.A.,2001; Repenning C.A.,Fefjar O., Heinrich W.-D., 1990).

Another important split represented in our analyses is *Mus / Rattus* dichotomy. Stepan, 2004a, re-evaluate this divergence time, challenging the basal position of *Progonomys* (11.8-12.1 Mya) and *Antemus* (13.75-12.5 Mya) as MRCA of *Mus* and *Rattus*. In fact, these two forms are now considered as predecessor or MRCA of all extant Murinae, and the calibration bounds for *Mus/Rattus* split are 8.8-10.3 Mya.

Results

1. Phylogenetic and biogeographic analyses

The ML and Bayesian analyses of the mitogenomic dataset provides trees that are congruent about the major lineages of Arvicolinae. Four major clades are defined by strongly supported monophylies: Ondatrini (*Ondatra*, *Neofiber*), 100 BP/0.99 PP, Myodini (*Myodes*, *Alticola*, *Eothenomys*), 100 BP/ 0.99 PP, Arvicolini (*Arvicola*), 100 BP/0.99 PP and Microtini (*Microtus* including *Lasiopodomys*), 97 BP/0.99 PP.

For the first time, the use of two species in *Arvicola* genera and of a complete representation of the *Neofiber/Ondatra* digeneric tribe, define separated Arvicolini and Ondatrini tribes with strong support.

If Arvicolini is sister-group of Microtini in both analyses, definition of deeper relationships is different, depending on the analysis: MP detailed tree defines Myodini as sister-group of Arvicolini+Microtini with a low bootstrap score (50%), while BI defines Ondatrini as the tribe sharing the MRCA with Arvicolini+Microtini (0.83 PP).

Microtini tribe is monophyletic, but phylogenetic position of *Proedromys sp.* is not clearly resolved. *Proedromys sp.* is included in the tribe as an early derivated forms of microtini MRCA in MP analysis (71 BP). In BI analysis, *Proedromys sp.* is sister group of “Asian” and “Himalyan” *Microtus* forms (0.62 PP).

In *Microtus* genera radiation, represented on both detailed trees (Figure 1 and 2), the topology appears as strongly linked to the worldwide distribution of *Microtus* species. By the way, four biogeographic clades are well supported.

The Himalayan clade groups together *Microtus irene* and *Lasiopodomys leucurus* (100 BP/1.00 PP). This clade is the sister-group of the Asiatic clade (98 BP/0.99 PP), grouping *Microtus middendorffi*, *Microtus limnophilus*, *Microtus montebelli*, and *Microtus kikuchii* (100 BP/0.99 PP).

The North American clade groups together *Microtus chrotorrhinus*, *Microtus pennsylvanicus*, and, as first species appeared in this clade, *Microtus richardsoni* (100 BP/ 0.99 PP). Its sister group is the European clade (83 BP/ 0.99 PP), composed by *Microtus duodecimcostatus*, *Microtus multiplex*, *Microtus subterraneus*, *Microtus arvalis*, *Microtus rossiaemeridionalis* and, in the most basal node, *Microtus agrestis* (82 BP/0.99 PP).

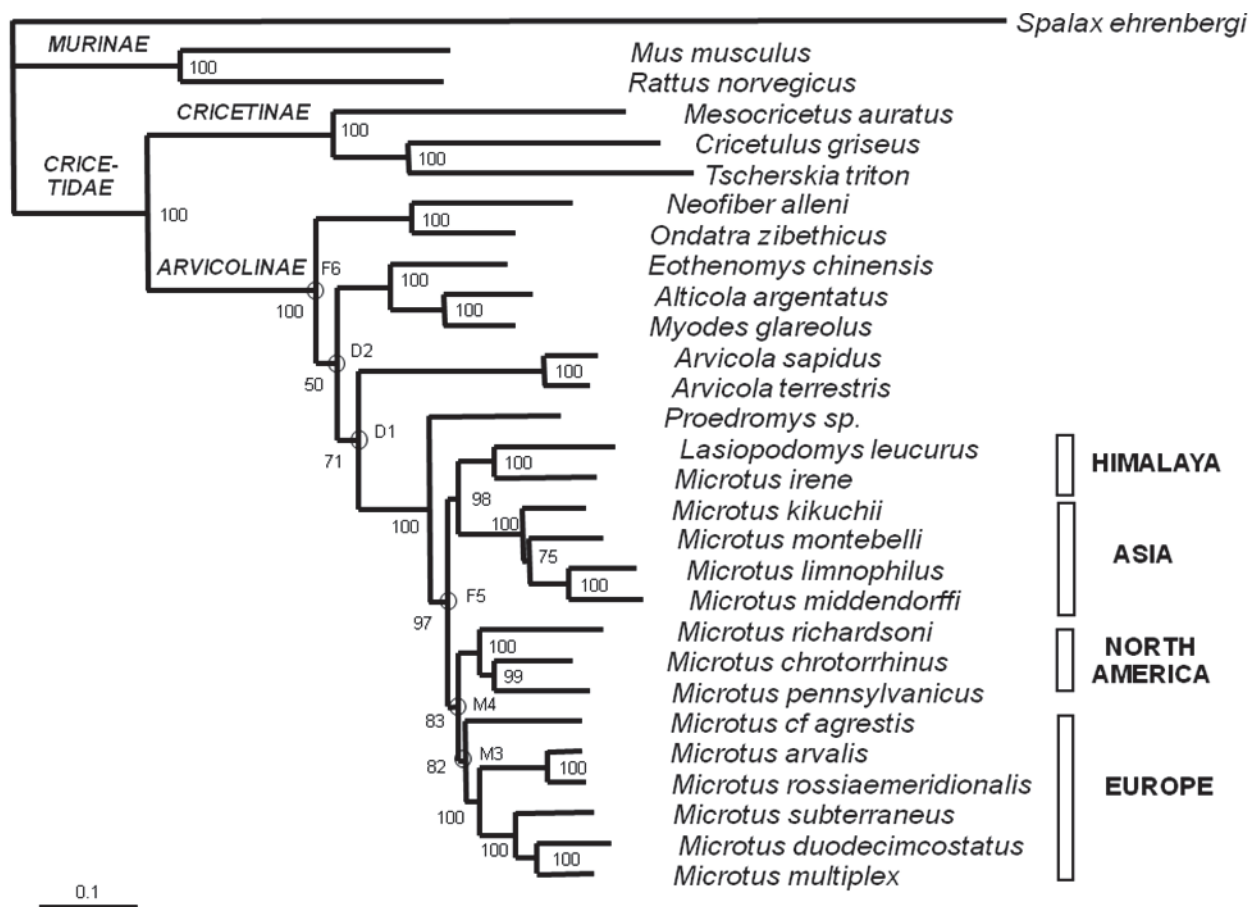


Figure 1: Maximum likelihood (ML) tree depicting the phylogenetic relationships among Arvicolinae as inferred from complete mitochondrial genome comparisons. Robustness of nodes is given with ML bootstrap percentages (BP). Branch lengths are proportional to the estimated number of substitutions per site. Circled nodes are tested through VLB analysis.

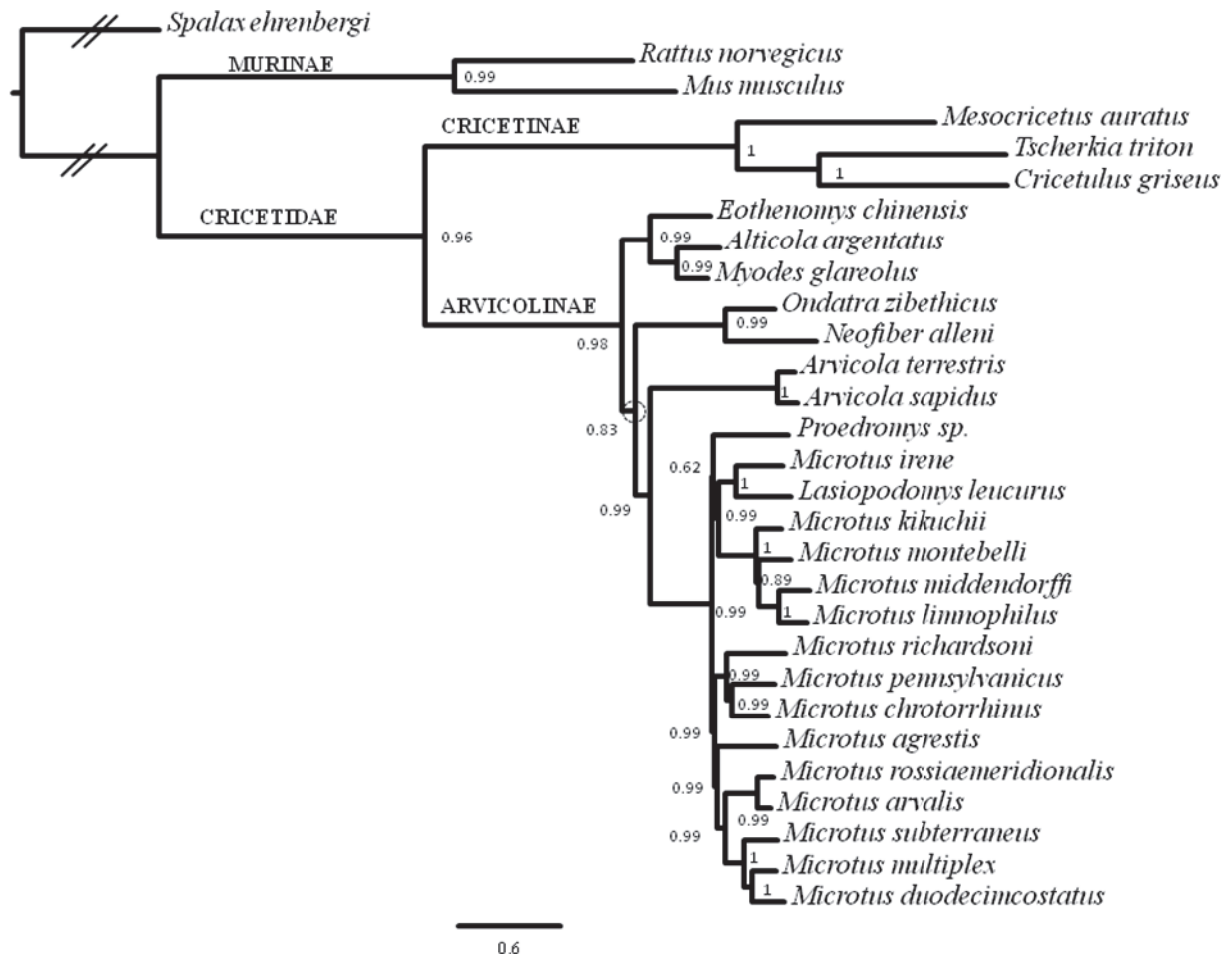


Figure 2: Bayesian Inference (BI) tree. Robustness of nodes is given with Posterior Probabilities (PP). Branch lengths are proportional to the estimated number of substitution per site. The circle indicates the node tested in VLB analysis.

2) Single-gene versus complete mitogenome resolving power

The 6 first examples of VLB curves are provided on fig. 4: two represent weakly supported nodes (D1, D2), two represent moderately supported nodes (M3,M4) and two represent strongly supported nodes (F5,F6) in ML tree. The seventh curve represents VLB results of a tested node, supported in BI tree, grouping together Ondatrini/Arvicolini+Microtini.

VLB replications allow defining, for a considered node, quantity of nucleotides markers from mitochondrial genome needed to provide enough informative signal, and support the node with strong bootstrap values.

Strongly supported nodes in the MP tree results underline the quality of mitochondrial data resolving power in deep nodes (Arvicolinae monophyly, F5, bootstrap saturation: 2000bp), as well as in recent radiation (*Microtus* s.l. excluding *Proedromys* sp, F6, bootstrap saturation: 8000bp). Moderately supported nodes, both found in recent *Microtus* biogeographic clades, show that mitochondrial dataset provides maximum informative signal when 6000 to 9000 nt are replicated. The signal saturation can be explained by model bias during MP analysis, and show how the variability of each position in the mitochondrial genome of Arvicolinae has to be fitted by a more complete model. This observation is completed by VLB results of weakly supported nodes in MP tree. D1 node defines a deep relationship between sister-group Microtini and Arvicolini. As saturation values are reached with a high number of nucleotides (14000 bp), quantity of informative signal in MP analyses is strongly linked to the number of nucleotides analyzed and how genome variability is fitted by MP models.

D2 is the more questioning node of our analysis. Weakly supported in MP tree, another hypothesis is evocated in BI analysis, and both have been tested through VLB analysis. By this way, we can observe if MP models can fit BI hypothesis. While curve D2 shows an increase of bootstrap values with number of nucleotides resampled, and so supports Myodini as sister-group of Arvicolini/Microtini, curve "TEST" demonstrate that MP models do not support BI hypothesis, as the node tested consider Ondatrini sister-group of Arvicolini/Microtini.

Providing a greater number of markers in our phylogenetic analysis allows description of strong relationships that were not identified in previous studies. The use of mitochondrial genome as phylogenetic marker implies comprehensive methods in order to extract maximum information despite high variability heterogeneity among sites. For these reasons, tree used in divergence time estimates is BI tree (fig. 2), best fitted to mitochondrial dataset, thanks to model mixture methods.

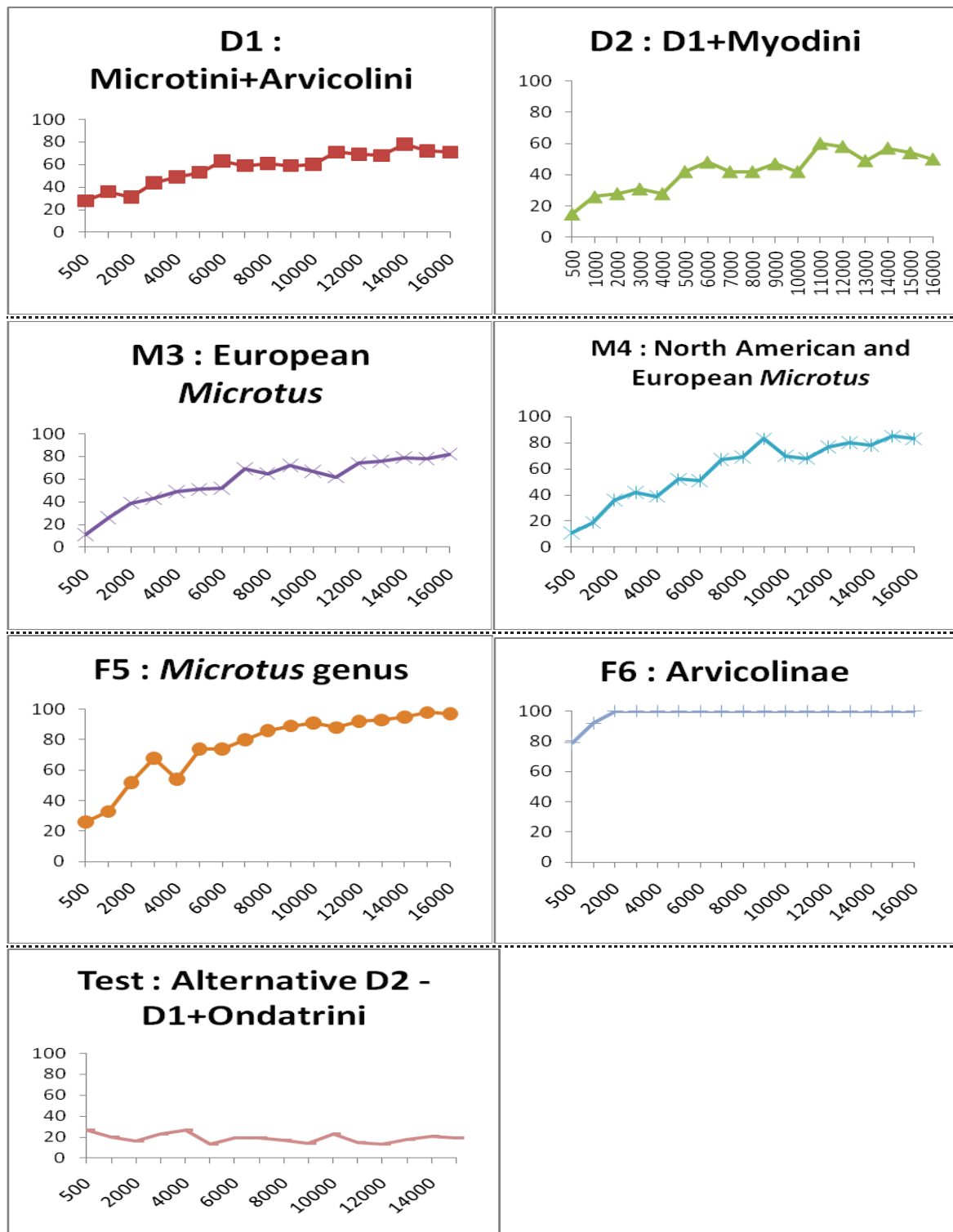


Figure 3: Plots of the variable-length bootstrap percentage 1)for six nodes recovered in Bayesian and ML analyses 2)for the node linking Ondatrini/(Arvicolini+Microtini) recovered only in BI analyses. The X-axis is the number of sites resampled (in 500 increments, first, than in 1000 increments from 500 to 16000-bp), and the Y-axis is the maximum likelihood BP percentage (out of 100 pseudoreplicates).

3. Divergence Time Estimates within and among Arvicolinae lineages

Estimated divergence times are indicated on the chronogram (Figure 3). The Arvicolinae radiation began in late Miocene, about 11 Mya. A first pulse of speciation took place between 10.95 and 9.32 Mya. This pulse generated the four major lineages described above: Ondatrini (10.22 Mya), Myodini (10.95 Mya), Arvicolini and Microtini (9.32 Mya).

A second pulse occurred between 8.39 and 4.64 Mya (late Miocene to early Pliocene), leading to the emergence of several genera in the four lineages. Ondatrini splitted between *Ondatra* and *Neofiber* genera 4.64 Mya, *Eothenomys* genera diverged in Myodini tribe 8.39 Mya, *Proedromys* is isolated from *Microtus* radiation more than 5 Mya.

The *Microtus* genera diversification began 5.25 Mya. It is characterized by several fast and contemporaneous radiations, taking place in several locations in Northern Hemisphere.

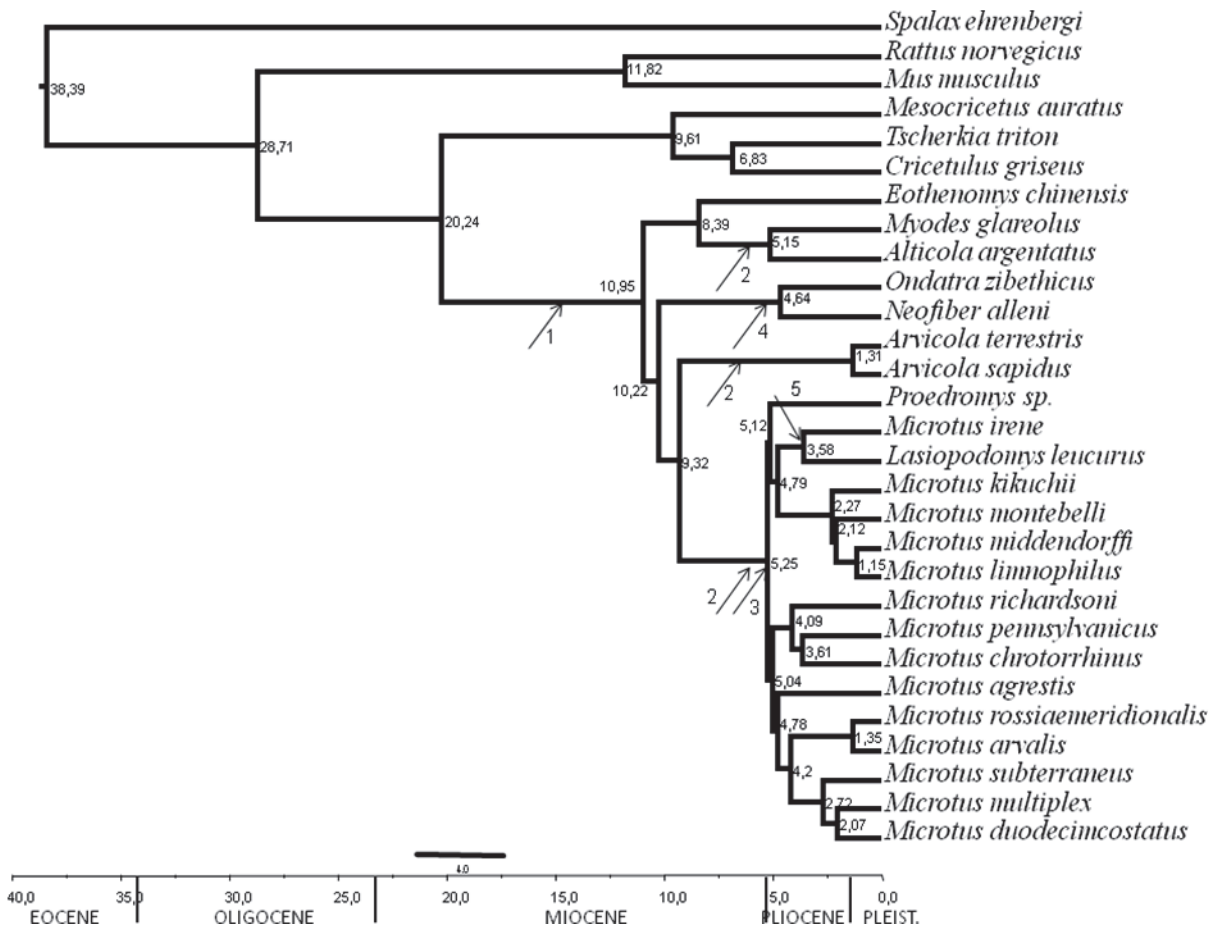


Figure 4: Chronogram with main Arvicolinae groups. For each group the oldest fossil is indicated by an arrow; Calibrations. Rectangles: 95% credibility intervals on divergence time estimates.

Discussion

This high variability between each mitochondrial genome in a same taxa, lead to a difficulty in obtaining the whole alignment: before this study, only 2 Arvicolinae and 13 rodents mitochondrial genomes were available, 5.5% of the 233 Mammals mitochondrial genomes available, while Rodents represent 40% of mammalian diversity.

In this study, we substantially improved the resolution of the phylogeny of Arvicolinae by sequencing and analyzing 18 complete mitochondrial genomes, thus more than doubling the total mitogenomic database in rodents. The paucity of full mitochondrial genomes in this group is probably explained by their fast evolutionary rate, which complicates primer design for PCR. This was overcome in voles thanks to the use of a large battery of primers, iteratively designed during the project as new sequence data became available.

1. Contribution of the whole mitochondrial genome

Improvements in the quantity of phylogenetic markers, and in the way to extract phylogenetic information from mitochondrial data, allows the definition of strong relationships between each tribes and to distinct clearly each *Microtus* biogeographic clade.

Complete mitochondrial data in phylogenetic analysis improve node support at different levels of Arvicolinae sub-family, detailing chronology of speciations and cladogenesis.

In comparison with other phylogenetic studies, the number of markers has been increased, while taxonomic sampling is representative of most voles and muskrats genera and biogeographic clades. The whole mitochondrial genome alignment carries more phylogenetic information than the mitochondrial gene Cytochrome b. By comparing VLB results obtained by (Galewski T. and E.,2006) on a node represented in both studies, we can say that Arvicolini/Microtini monophyly is poorly supported by cytochrome b dataset, while it is one of the most supported node with mitochondrial dataset.

MP analysis based on the whole dataset clearly supports most nodes of the tree, even if Ondatrini/Myodini position relative to Arvicolini/Microtini has to be clarified. The use of several variable partitions and models in Bayesian Inference allows extracting a less biased informative

signal. This analysis also underlines the questioning position of Ondatrini or Myodini as sister group of Arvicolini/Microtini.

Heterogeneity of the nucleotide variability on each site implies the use of non-parametric methods, which avoid model bias introduction in phylogenetic analyses.

2. Inter-genera relationships of Arvicolinae subfamily

The definition of deep relationships between each tribe constituting the subfamily allows describing a scenario of the different specific radiations, also observed on a paleontological point of view. The evolutionary history of Arvicolinae subfamily is segmented in three radiation phases. These radiations can be corroborated with the description of fossils record in (Chaline J.,1999).

The first Nearctic radiation of Arvicolinae, represented by *Ogmodontomys sawrockensis*, (1) in Hemphillian deposits of North America, possibly led to the Ondatrini tribe formation. *Neofiber alleni* has now a distribution limited to Florida, while *Ondatra zibethicus*, due to several anthropic introductions, can also be found in North Europe or Argentina. Observation of *Pliopotamys minor* (4) in ancient Pliocene, and high morphological similarities with *Ondatra* genera, confirm that the digeneric tribe Ondatrini is one of the first cladogenesis which has occurred in the sub-family.

The fossil genera *Prosomys* (2), found in ancient Pliocene, seem to be the founder genera of the first Palearctic radiation. *Prosomys insuliferus* (2) lineage contains *Mimomys*, but also derived contemporaneous genera: *Arvicola*, *Clethrionomys* (*Myodes*), and *Eothenomys*. At the present time, Myodini tribe still has a Holarctic distribution.while Arvicolini tribe is limited to Palearctic regions.

On morphological and paleontological grounds, these 2 non-Microtini radiations can be considered as simultaneous in Late Miocene, ancient Pliocene, in agreement with MP tree topology, and deep relationships between defined tribes. If we consider BI tree in our analysis, 2 different Palearctic radiations occurred in early Pliocene: Myodini, which distribution is currently Holartic, and Arvicolini. This hypothesis implies more ancestral population migrations, as founder species can move from Nearctic and Palearctic through Beringia.

3. *Microtus* Genera Biogeography

Allophaiomys (3) genera, derived from *Mimomys*, are represented in Eburian glacial phase deposits (2 Mya), through the Holarctic area. As these fossil forms are widespread through North Hemisphere, independent evolutions of different *Allophaiomys* populations are evocated to explain the second Palearctic radiation, also qualified as “Modern Voles radiation”. In fact, we can observe 4 different simultaneous radiations occurring in the same time in North America, Europe and Asia. Furthermore, this observation leads to morphological convergences evidences, especially between each biogeographical lineage. The definition of four biogeographic clades in *Microtus* genera, with mitochondrial markers, allows the reconstitution of an evolutionary scenario without the morphological bias generated by such an explosive radiation.

a) Himalayan *Microtus*

Microtus irene and *Lasiopodomys leucurus* share morphological characters with *Allophaiomys pliocaenicus* (5). These species are isolated from other Asiatic species, since their population is fixed in or around Himalaya, between 4.79 and 3.58 mya.

b) Asiatic *Microtus*

Our analysis separate Asiatic *Microtus* in two groups: continental mainland *Microtus* (*M. middendorffi*, Central Russia; *M. limnophilus*, North China) and insular *Microtus* species (*M. montebelli*, Japan endemism; *M. kikuchii*, Taipei endemism). This group shares a common history with Himalayan *Microtus* (*Allophaiomys pliocaenicus*), but fossil register is still poorly provided. *Microtus montebelli*, with an arvaloid dental morphology, shows convergence with *Microtus arvalis* and *Microtus rossiaemeridionalis*, two species contained in the European biogeographic clade.

c) North American *Microtus*

This clade groups together *Microtus pennsylvanicus* and *Microtus chrotorrhinus*, sharing a derived dental morphology (agrestoid) from *Phenacomys deeringensis* (Alaska, Middle Pleistocene). This morphology is convergent with the agrestoid form found in *M. agrestis* (European *Microtus*)

d) European *Microtus*

With morphological characters defining the *Terricola* subgenus, *M. subterraneus*, *M. multiplex* and *M. duodecimcostatus*, are also grouped together in our molecular analysis. Both morphological and molecular data agree on the arvaloid forms,

grouping together *M. arvalis* and *M. rossiaemeridionalis*. As *Microtus agrestis* is the only European specie showing agrestoid dental morphology, population fragmentation and migration during quaternary climatic changes can explain an eventual morphological homology with *M. chrotorrhinus* and *M. pennsylvanicus*.

CONCLUSION

The use of complete mitochondrial genome in phylogenetic analyses represents potential improvements of previously defined relationships. The quantity of supplementary markers, in comparison with previous phylogenetic studies based on nuclear or mitochondrial markers, provides much more informative sites than Cytochrome b gene. However, variability heterogeneity among sites has to be taken in count in the analysis, and several models and partitions must be tested in order to extract maximum information from the dataset. From this point of view, Phylobayes inference is the most fitted to the dataset, as it explore several possibilities in one analysis, while PAUP applies one model (here, GTR) to all positions in the alignment.

Improvements in Arvicolinae relationships description have been brought by our analysis. Previously found Arvicolinae tribes are defined by strong monophylies in our results, even if sister-group of Arvivolini/Microtini still has to be described. The information provided by the dataset allows dating the first pulse of Arvicolinae radiation in late Miocene.

The high variability of mitochondrial genome detected in *Microtus* genera allows detecting recently described biogeographic clades, even if this radiation is considered as explosive.

In general, complete mitochondrial data can be useful in phylogenetic analyses when pulses of speciation, and consecutive explosive radiations, are detected. This genome high variability allows breaking basal polytomies, such as the ones evocated when molecular or morphological markers are saturated.

In the future, addition of an improved taxonomic sampling and sequencing complete mitochondrial genome could help investigating the whole Arvicolinae radiation, by including lemmings (*Dicrostonyx*, *Lemmus*, *Lemmiscus*,...), more rare taxa in Arvicolinae tribes (*Prometheomys*, *Alticola*, *Myopus*,...) or in *Microtus* genera radiation, for which only ¼ of current taxa are represented in our analysis.

Acknowledgments

This work has been supported by the bioinformatics cluster of the ISE-M, the Agence Nationale de la Recherche (ANR) Biologie Systémique (BIOSYS06_136906), "Génomique Animale" (ANR-08-GENM-036-01 "CoGeBi"), and "Domaines Emergents" (ANR-08-EMER-011 "PhylAriane"). This publication is contribution N° 2011-YYY of the Institut des Sciences de l'Evolution de Montpellier (UMR 5554 - CNRS).

References

- Abramson N.I., L.V.S., Bannikova A.A., Tesakov A.S., 2009a. Radiation events in the subfamily Arvicolinae (Rodentia) : Evidence from nuclear genes. *Doklady Biological Sciences* 428, 713-717.
- Abramson N.I., L.V.S., Tesakov A.S., Bannikova A.A., 2009b. Supraspecies relationships in the subfamily Arvicolinae (Rodentia, Cricetidae) : an unexpected result of nuclear gene analysis. *Molecular Biology* 43, 897-909.
- Chaline, J., Brunet-Lecomte, P., Montuire, S., Viriot, L., Courant, F., 1999. Anatomy of the arvicoline radiation (Rodentia): palaeogeographical, palaeoecological history and evolutionary data. *Ann. Zool. Fennici* 36, 239–267.
- Chaline J., B.-L.P., Montuire S., Viriot L., Courant F., 1999. Anatomy of the Arvicolinae radiation (Rodentia) : palaeogeography, palaeoecological history and evolutionary data. *Annales Zoologici Fennici* 36, 239-267.
- Conroy C.J., C.J.A., 1999. MtDNA evidence for repeated pulses of speciation within Arvicolinae and Murid rodents. *Journal of Mammalian Evolution* 6, 221-245.
- Galewski T., T.M.-K., Sanchez S., Chevret P., E., P.E.a.D., 2006. The evolutionary radiation of Arvicolinae rodents (voles and lemmings): relative contribution of nuclear and mitochondrial DNA phylogenies. *BMC Evolutionary Biology* 6, 80.
- Katoh K., A.G., Toh H., 2009. Multiple Alignment of DNA sequences with MAFFT. In: Press, H. (Ed.), *Bioinformatics for DNA sequence analysis*, pp. 39-64.
- Lartillot N., L.T., Blanquart S., 2009. Phylobayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25, 2286-2288.
- Philippe H., 1993. MUST, a computer package of Management Utilities for Sequences and Trees. *Nucleic Acids Residuals* 21, 5264-5272.

Posada D., C.K.A., 1998. Modeltest : testing the model of DNA substitution. *Bioinformatics* 14, 817-818.

Repenning C.A., 2001. Beringian climate during intercontinental dispersal: a mouse eye view. *Quaternary Science Reviews* 20, 25-40.

Repenning C.A., F.O., Heinrich W.D., 1990. Arvicolid rodent biochronology of the Northern Hemisphere. In: Fefjar O., H.W.-D. (Ed.), *International Symposium on Evolution, Phylogeny, and Biostratigraphy of Arvicolids (Rodentia, Mammalia)*, Pfeil-Verlag, Prague, pp. 385-418.

Steppan S.J., A.R.M., Spinks P.Q., Hale C., 2005. Multigene phylogeny of the Old World mice, Murinae, reveals distinct geographic lineages and the declining utility of mitochondrial genes compared to nuclear genes. *Molecular Phylogenetics and Evolution* 37, 370-388.

Steppan, S.J., Adkins, R.M., Anderson, J., 2004. Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. *Syst. Biol.* 53, 533–553.

Swofford D.L., 2003. PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4. In: Associates, S. (Ed.), Sunderland, Massachusetts.

Wilson, D.E., Reeder, D.M., 2005. *Mammal species of the world*. Johns Hopkins University Press, Baltimore.

Figures

Figure 1: Maximum likelihood (ML) tree depicting the phylogenetic relationships among Arvicolinae as inferred from complete mitochondrial genome comparisons. Robustness of nodes is given with ML bootstrap percentages (BP) . Branch lengths are proportional to the estimated number of substitutions per site. Circled nodes are tested through VLB analysis.

Figure 2: Bayesian Inference (BI) tree. Robustness of nodes is given with Posterior Probabilities (PP). Branch lengths are proportional to the estimated number of substitution per site. The circle indicates the node tested in VLB analysis.

Figure 3: Chronogram with main Arvicolinae groups. For each group the oldest fossil is indicated by an arrow; Calibrations. Rectangles: 95% credibility intervals on divergence time estimates.

Figure 4: Plots of the variable-length bootstrap percentage 1) for six nodes recovered in Bayesian and ML analyses 2) for the node linking Ondatrini/(Arvicolini+Microtini) recovered only in BI analyses. The X-axis is the number of sites resampled (in 500 increments, first, than in 1000 increments from 500 to 16000-bp), and the Y-axis is the maximum likelihood BP percentage (out of 100 pseudoreplicates).

Tables

Table 1: Taxonomic sampling and accession numbers.

3^e Partie : Séquençage de masse

Avec la méthode classique de séquençage de génomes mitochondriaux, évoquée en 1^{ère} partie, 18 génomes mitochondriaux ont été obtenus en 18 mois, soit $18 \times 16 \text{ kb} = 288000 \text{ pb}$. L'échantillonnage taxonomique représente un total de 55 espèces d'Arvicolinae, soit 37 génomes mitochondriaux en attente d'amplification et de séquençage. L'exploration des NMS (Nouvelles Méthodes de Séquençage) afin d'accélérer le rythme de lecture des séquences et de constitution des génomes mitochondriaux a été réalisée avec de nombreuses adaptations méthodologiques.

Le pyroséquençage (par approche 454) implique le mélange de la totalité des amplicons (entre 500 et 1000 pb de longueur) obtenus sur le génome mitochondrial d'un individu. Ce mélange doit être équimolaire et comprendre un minimum de 500ng de matériel génétique pur à une concentration de 20 ng/ μl . Une amplification par émulsion PCR au moyen de méga-amorces (couples d'amorces originales + tag A + tag B) permet d'identifier chaque amplicons au sein du pool séquencé, et d'exposer à la lecture un brin de la séquence visée.

Le pyroséquençage entrepris avec GATC-Biotech implique donc différents impératifs : (i) l'amplification de fragments de génomes mitochondriaux plus courts (entre 500 et 1000 pb) et donc plus nombreux, afin d'obtenir la totalité de la séquence mitochondriale avec un pool de 30 amplicons, (ii) la purification des produits d'amplification avec une méthode permettant l'obtention d'une solution pure d'ADN amplifié, et (iii) le dosage précis de chacun des échantillons.

Redéfinition des fragments d'amplification

Grâce au tableau 2 (présenté ici partiellement), représentant la totalité des amorces définies, leur position et la taille attendue des fragments d'amplification, il est possible de définir la totalité des couples d'amorces compris entre 500 et 1000 pb de longueur sur l'ensemble du génome mitochondrial.

Les amplifications ainsi définies sont testées par PCR sur les extractions d'ADN d'un individu dont le mitogénome à été séquencé de manière classique, *Myodes glareolus*, et d'un individu non séquencé, *Microtus townsendi*, et permettent de sélectionner 30 couples d'amorces fonctionnant à l'échelle sous-familiale et couvrant la quasi-totalité du génome mitochondrial. Ces PCR sont réalisées en gradient de température d'hybridation afin d'améliorer le taux de réussite des amplifications à large échelle taxonomique. Sur un même fragment visé, 3 températures d'hybridations sont pratiquées, et les amplifications les plus concentrées en séquences cibles sont sélectionnées et purifiées.

Purification des produits d'amplification

Les produits d'amplification doivent être purifiés de manière à retirer les sous-produits de la réaction de PCR de la solution, afin d'en réaliser le dosage et le pyroséquençage.

La meilleure des solutions adoptée ici est la purification par billes poreuses magnétiques, qui fixent l'ADN et permettent le lavage de la solution, réalisé sur la plate-forme génotypage séquençage de l'IFR 119. L'automatisation de la méthode permet de purifier 96 puits en une demi-journée, soit potentiellement 4 génomes mitochondriaux complets.

Amorces/Primers	Reverse>>	F_rv		N_rv		G2_rv		G_rv		H_rv	
Forward	POSITION	39		993		1770		2490		3892	
			TA		TA		TA		TA		TA
G2_fw	26		13		967		1744		2464		3866
H_fw	602		15858		391		1168		1888		3290
N_fw	974		15486		19		796		1516		2918
V_fw	1752		14708		15662		18		738		2140
A_fw	2133		14327		15281		16058		357		1759
I_fw	2703		13757		14711		15488		16208		1189
W_fw	3324		13136		14090		14867		15587		568
B_fw	3837		12623		13577		14354		15074		55
U_fw	4144		12316		13270		14047		14767		16169
J_fw	4943		11517		12471		13248		13968		15370
Q_fw	5285		11175		12129		12906		13626		15028
T_fw	6160		10300		11254		12031		12751		14153
K_fw	6890		9570		10524		11301		12021		13423
L_fw	7677		8783		9737		10514		11234		12636
P_fw	8608		7852		8806		9583		10303		11705
X_fw	9276		7184		8138		8915		9635		11037
D2_fw	9806		6654		7608		8385		9105		10507
R_fw	10689		5771		6725		7502		8222		9624
S_fw	10966		5494		6448		7225		7945		9347
E4_fw	11664		4796		5750		6527		7247		8649
Z_fw	11997		4463		5417		6194		6914		8316
E2_fw	12615		3845		4799		5576		6296		7698
Z2_fw	12915		3545		4499		5276		5996		7398
E3_fw	13536		2924		3878		4655		5375		6777
M_fw	14084		2376		3330		4107		4827		6229
Y_fw	14712		1748		2702		3479		4199		5601
O_fw	15360		1100		2054		2831		3551		4953
G_fw	15810		650		1604		2381		3101		4503

Tableau 2 (partiel) : Exemples de couples d'amorces et tailles attendues des amplifications.

Dosage des produits d'amplification

Les dosages des fragments amplifiés et purifiés sont réalisés au moyen de l'instrumentation NanoDrop (UV-Vis).

Les résultats obtenus sur les deux échantillons testés regroupent des amplicons en solution peu concentrée (environ 3 ng/μl) à très concentrée (plus de 20 ng/μl), ce qui reste loin des 150 ng/μl requis.

Comparaison des deux méthodes de séquençage

Le tableau 3 présente un bilan des exigences et des résultats obtenus au travers des deux méthodes utilisées au cours de cette thèse.

Dans la mesure où les quantités d'ADN nécessaires sont très importantes et difficiles à obtenir par des méthodes classiques et peu consommatrices de temps, l'échec des premières tentatives de pyroséquençage est très fortement lié à ces faibles quantités d'ADN amplifiés et mis à disposition pour les réactions sous-traitées en PCR émulsion.

Plusieurs solutions peuvent être envisagées afin d'obtenir les amplicons d'intérêt dans des concentrations plus élevées : le clonage, après amplification, en est une, et permettrait d'obtenir en environ une semaine, la quantité d'ADN souhaitée. Cependant, l'utilisation du temps qui en résulterait pourrait être mise à profit dans l'obtention des génomes mitochondriaux par la méthode classique (1^{ère} partie), avec amplification de séquences plus longues et lecture de celle-ci en méthode de séquençage Sanger, technique « classique ».

L'enrichissement de la base de données en génomes mitochondriaux d'Arvicolinae est encore envisageable au moyen des deux techniques évoquées, chacune présentant des avantages et des inconvénients d'ordre principalement temporel. De ce point de vue, la méthode « classique » reste à ce jour, la stratégie la plus rentable pour l'obtention rapide de génomes mitochondriaux complets et cartographiés.

	BAS DEBIT - SANGER	HAUT DEBIT - 454
TECHNIQUE DE SEQUENCAGE	Séquençage laser	Pyroséquençage
CAHIER DES CHARGES		
PCR		
Longueur des fragments amplifiés	2500 pb maximum	1000pb maximum
Nombre de fragments par génomes	environ 15	environ 30
Concentration en amplicons nécessaire	3 ng/μl minimum	150 ng/μl minimum
PURIFICATION	EXOSAP-It (USB) Purification chimique	Billes magnétiques poreuses Purification physique
DOSAGE	Intensité lumineuse sur électrophorèse post-PCR	Dosage NanoDrop
VOLUME DISPONIBLE	15 μl/amplicon	8 μl/amplicon
CONCENTRATION MOYENNE APRES PCR	15 ng/μl	
TEMPS THEORIQUE D'OBTENTION D'UN MITOGENOME COMPLET POUR UN INDIVIDU	15 à 20 jours	20 à 30 jours
Semaine 1	Extraction - Amplification - Dosage - Purification	Extraction - Amplification - Purification
Semaine 2	Réactions de Séquence - Précipitations Séquençages - Corrections des séquences - Réamplifications	Purification - Dosage NanoDrop - Expédition
Semaine 3	Assemblage des Séquences - Obtention des génomes complets	Sous-traitance : Pooling, Coligation, Nebullisation, Taging, Pyroséquençage
Semaine 4	Eventuellement amplification avec amorces spécifiques pour combler les zones non séquencées	Fin sous-traitance, Acquisition des données
RESULTATS OPTIMAUX	Génome mitochondrial complet	Génome mitochondrial quasi-complet
COUT	environ 10€ par amplicons séquencés 30 à 60 séquençage par individu 300 à 600€ par individu	210€ préparation préalable sous-traitance : environ 350€ pour 30 amplicons du même individus 550 à 650€ par individu
CAS D'ECHEC	Intervention possible à tous les niveaux du protocole expérimental : réduction des coûts	Protocole expérimental à recommencer : coût multipliés par 2 OU Possibilité de clonage des amplicons pour augmenter la concentration en séquence cible
COUT TOTAL	350-650€ par génome mitochondrial	>1000€ par génome mitochondrial

Tableau 3 : Comparaison des deux méthodes de séquençage utilisées au cours de cette thèse.

Perspectives : Génomique comparative

Les mitogénomes d'Arvicolinae générés au cours de cette thèse ont été utilisés pour clarifier l'histoire évolutive de ce groupe. Ces données fournissent par ailleurs l'opportunité de mieux comprendre la dynamique évolutive du génome mitochondrial chez les mammifères, et notamment les variations entre sites et entre lignées du taux de mutation. Il s'agit en effet du premier jeu de données raisonnablement complet de mitogénomes dans un groupe de mammifères à évolution rapide. Les études comparatives du processus mutationnel menées à ce jour chez les mammifères (Galtier et al 2006, Nabholz et al 2008 Genetics, Nabholz et al 2008 MBE) se basent sur un seul gène, le cytochrome b. Il semble pertinent de passer à l'échelle génomique, et d'obtenir une image globale de la distribution, de l'intensité, et de la conservation évolutive des points chauds mutationnels dans l'ADN mitochondrial des mammifères.

A titre d'illustration, nous présentons une analyse préliminaire des points chauds mutationnels dans le génome mitochondrial des Arvicolinae, que nous comparons avec des analyses similaires réalisées en parallèle chez un groupe à évolution lente, les pinnipèdes, qui présente une profondeur phylogénétique comparable en terme de divergence mitochondriale, quoiqu'étant d'origine bien plus ancienne.

Constitution des bases de données : mitogénomes à évolution lente

Le site internet OGRE (Organellar Genome Retrieval system <http://drake.physics.mcmaster.ca/ogre/>), permet de retrouver les lignées de Mammifères les plus représentées par leur génome mitochondrial complet et cartographié. Ainsi les clades Carnivora (54 mitogénomes disponibles), Cetartiodactyla (53) et Primates (28) apparaissent comme les plus riches en génomes mitochondriaux complets, par opposition au clade Rongeurs, regroupant seulement 16 mitogénomes complets. De plus, dans le cadre d'une étude en génomique comparative, ces clades fortement représentés sont également décrits comme à évolution mitochondriale lente, et présentent de fortes variabilités de traits d'histoire de vie.

Les génomes de Pinnipèdes ont tous été alignés, et découpés en fonction des gènes représentés sur les séquences, afin d'établir des alignements de toutes les troisièmes positions des codons, qui seront utilisés par la suite.

Une analyse phylogénétique par maximum de vraisemblance (modèle GTR+I+ Γ) est réalisé sur l'ensemble des données de chaque clade, afin d'établir une topologie et des longueurs de branches propres à chacun d'eux.

Comparaison des distances génétiques - Observation Polymorphisme/Divergence

L'utilisation du logiciel compdist (Unix) dans la mesure de la distance génétique maximale et médiane d'un arbre donnée, implique l'inférence préalable des topologies et longueurs de branches par maximum de vraisemblance, réalisée au moyen du logiciel phyML (Guindon S.,2003). L'hétérogénéité dans la variabilité des sites mitochondriaux de mammifères implique l'observation de distances génétiques équivalentes à petite échelle chez les groupes à fort taux de substitution mitochondriale

et à plus large échelle chez les groupes présentant un polymorphisme moléculaire mitochondrial plus réduit.

A grande échelle, la comparaison des distances génétiques entre le genre *Microtus* et les deux sous-familles Phocidae-Otariidae, implique la réalisation des arbres de la figure 12, sur la totalité des gènes codants concaténés.

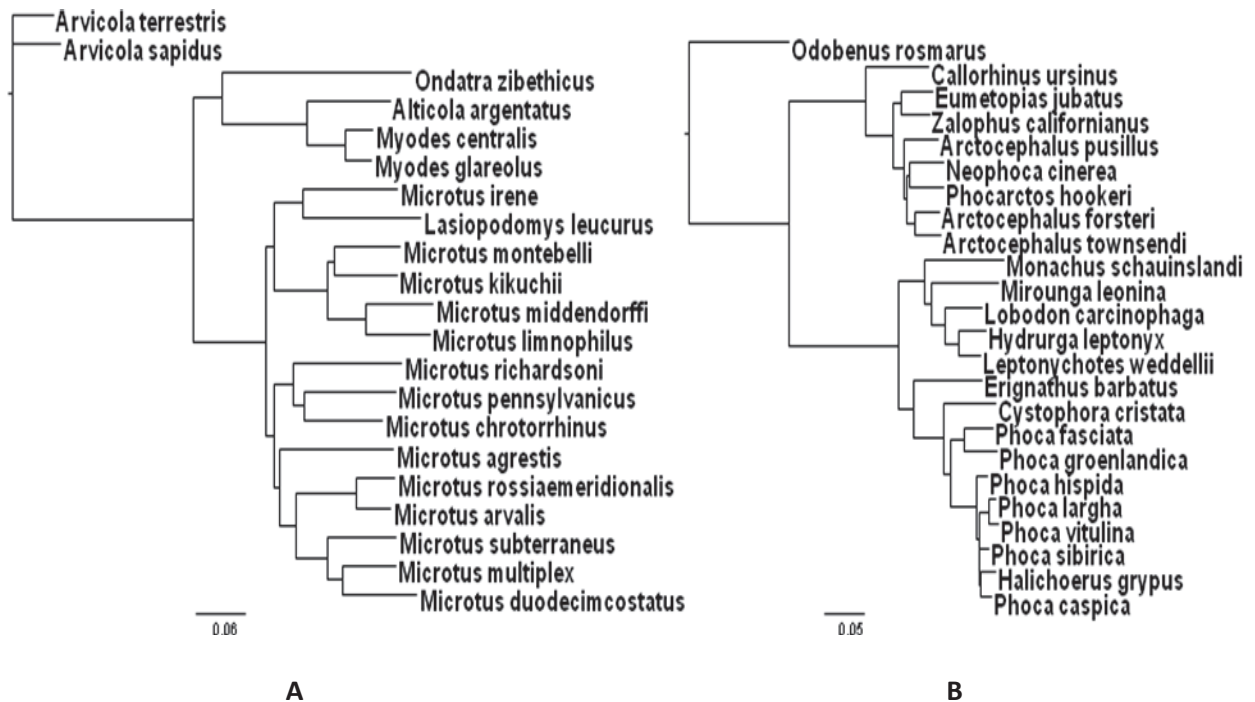


Figure 12 : Arbres ML représentant le genre *Microtus* (A), 21 taxons et 11399 caractères et les deux sous-familles Phocidae-Otariidae (B), 24 taxons et 11410 caractères. Le modèle utilisé est GTR+I+Γ.

Les différences des longueurs de branches soulignent la vitesse d'évolution mitochondriale accrue chez le genre *Microtus*, relativement aux sous-familles Phocidae-Otariidae.

La distance génétique médiane de 0,46 chez *Microtus* et de 0,42 pour le clade Otariidae-Phocidae, et des distances maximales de 0,79 et 0,6 respectivement, montre à quel point la divergence génétique atteint les mêmes proportions, mais à des échelles totalement différentes. Ainsi la distance génétique accumulée au cours du temps entre deux espèces proches de *Microtus*, séparées entre 3 et 2 millions d'années, peut être plus grande que la distance accumulée au cours d'un temps plus long séparant une espèce Otariidae et une espèce Phocidae, la séparation des deux sous-familles étant datée à environ 30Ma (Arnason U.,2007). Cette comparaison permet de réaffirmer le caractère exceptionnel de la forte variabilité du génome mitochondrial au sein du genre *Microtus*.

Détection des points chauds mutationnels

Il s'agit d'identifier et de dénombrer, au sein de plusieurs lignées de Mammifères, les sites soumis à de forts taux de substitution et donc fortement polymorphiques à l'intérieur d'une ou plusieurs lignées.

La méthode à utiliser est différente de celle utilisée en phylogénie moléculaire. Ici, la topologie seule est gardée, mais le taux de substitution de chaque site doit être recalculé sans effectuer de regroupement de sites en fonction de leur pattern de substitution modélisé.

Connaissant topologie et longueurs de branches de l'arbre, le logiciel Rate4 Site (Mayrose I., 2004) permet d'obtenir des taux de substitution pour chaque site mitochondrial d'un clade donné. Le taux de substitution obtenu est ici relatif, c'est-à-dire qu'il définit le taux de substitution d'un site en fonction du taux de substitution moyen inféré à partir des données. Un site pourra donc être décrit comme évoluant, par exemple, deux fois plus vite que les autres sites de la séquence, ce qui a une importance réelle dans cette étude comparative. Il est cependant nécessaire de comparer des groupes dont les distances génétiques maximales sont équivalentes, afin d'établir des taux relatifs de substitution entre des lignées hétérogènes au niveau de la variabilité des séquences mitochondriales. L'établissement précédent d'un cadre temporel pour la radiation de la sous-famille Arvicolinae, permettra d'obtenir les taux absolus de substitution par site.

Afin de n'établir que des taux de substitutions neutres, et dans le but d'exclure les phénomènes sélectifs, seules les troisièmes positions des codons doivent être analysées par la suite.

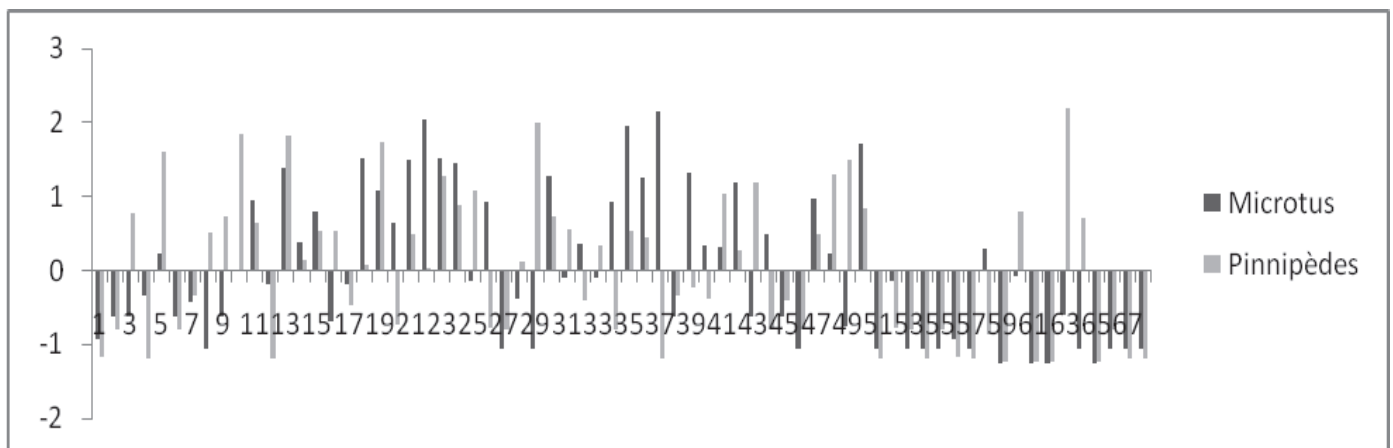


Figure 13 : Résultats obtenus avec R4S sur les troisièmes positions des codons du gène ATP8 chez le genre *Microtus* (gris foncé) et les Pinnipèdes (gris clair). En abscisses, les positions des codons sur la séquence codante, en ordonnées, le taux relatif de substitution.

Chez ces groupes à peu près équivalents en distance génétique, on retrouve des sites en position 3 fortement variables, par contre il y a également des sites fortement variables chez un seul clade (Pinnipèdes ou *Microtus*). Le site extrêmement variable en position 63 de l'alignement ATP8 chez les Pinnipèdes, et décrit en revanche comme peu variable chez les *Microtus*. Dans ce cas, il s'agit d'un point chaud mutationnel détecté chez les Pinnipèdes, et non chez *Microtus*.

Ces observations posent la question des hot spots partagés : à l'intérieur des clades évoqués, il est pour l'instant impossible de savoir si un site détecté comme très variable est effectivement variable à l'échelle du clade entier, auquel cas le hot spot est partagé entre les membres du clade, ou fortement variable uniquement chez un individu ou un sous-clade au sein du clade en question.

Il est donc nécessaire d'observer les substitutions multiples à une échelle plus réduite. En conséquence, 5 sous-clades ont été définis d'après les relations phylogénétiques mise en place précédemment. La figure 14 représente ces groupes, dont la variabilité des troisièmes positions des codons ont été analysées à travers le logiciel Rate4Site.

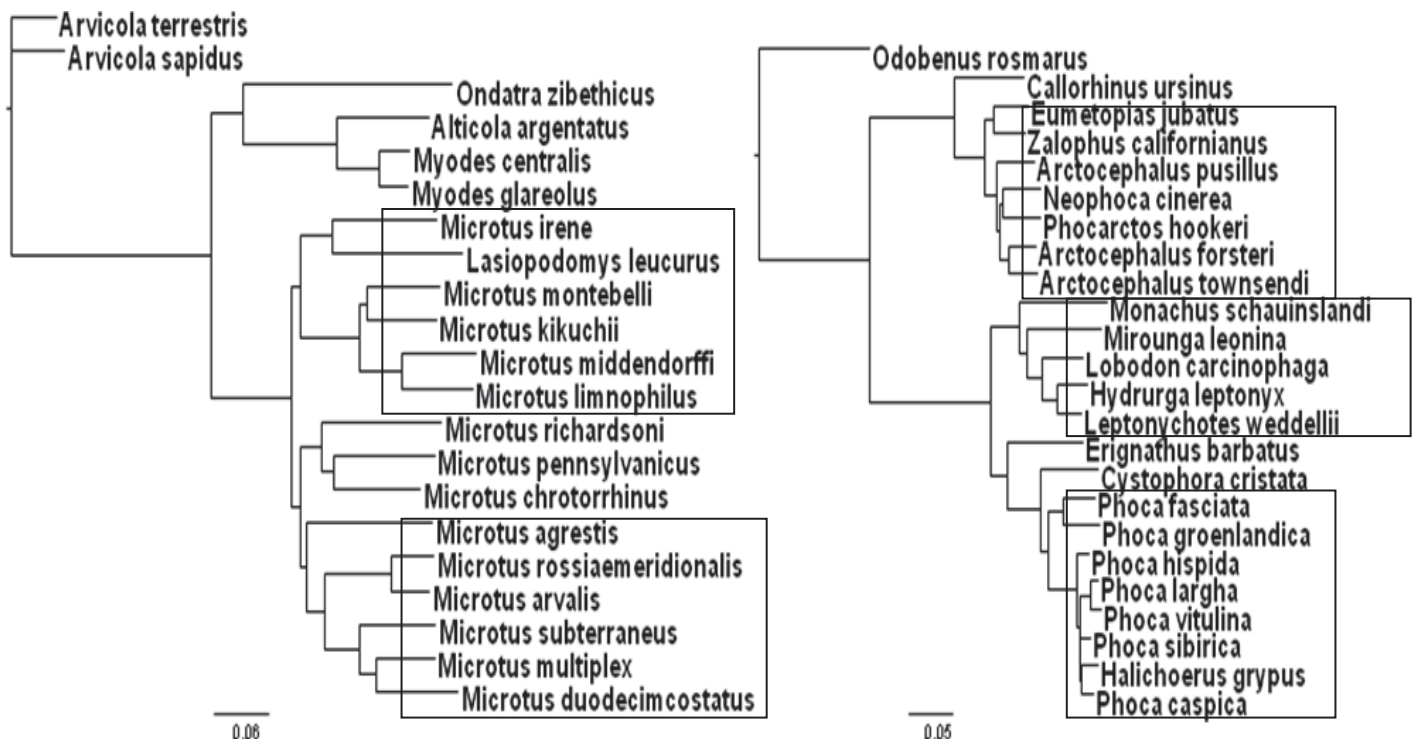


Figure 14 : Sous groupes définis pour l'étude des hot spots partagés. De haut en bas, de gauche à droite : MIC2 (Microtus Asiatiques), MIC1 (Microtus Européens), CAR1, CAR2, et CAR3.

Détection des points chauds partagés et non-partagés

La figure 15 expose les résultats obtenus chez les 5 sous-clades, au niveau des troisièmes positions des codons du gène ATP8.

Le site en position 63 sur l'alignement, détecté comme fortement variable à l'échelle des Pinnipèdes, apparaît également comme fortement variable à l'échelle des 3 sous-clades définis chez les Carnivores. Il s'agit d'un point chaud partagé par tous les membres du clade, et ce site apparaît comme faiblement variable chez l'ensemble des *Microtus*.

Il existe donc un phénomène cumulatif des hots spots, où la forte variabilité des sites à l'échelle d'un, ou plusieurs, clade restreint, est répercutée aux observations sur des clades plus larges. Le site 64 montre une forte variabilité chez les Pinnipèdes en général, et cette tendance est soulignée par la variabilité plus forte du site chez le groupe CAR3. La forte variabilité du site dans un seul des clades étudiés est porteuse de l'information sur la variabilité du site à l'échelle du clade Pinnipèdes entier. A l'opposé, le site 37 apparaît comme fortement variable chez l'ensemble des clades *Microtus*, alors qu'il ne l'est pas chez les Pinnipèdes.

Des observations précédentes, il est possible de déduire la position sur les arbres où le point chaud mutationnel, c'est-à-dire la forte variabilité d'un site, se met en place. La figure 16 résume les observations précédentes.

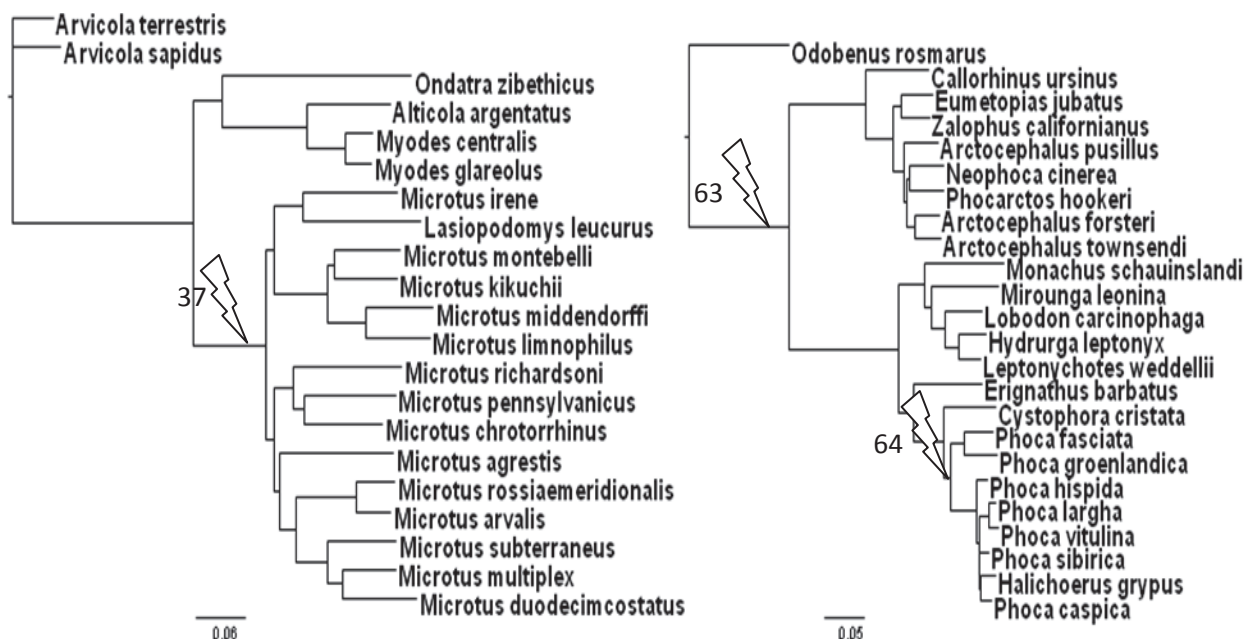


Figure 16 : Arbres ML avec les sites repérés en situation de point chaud mutationnel (« éclair »). Le chiffre donne la position du site concerné.

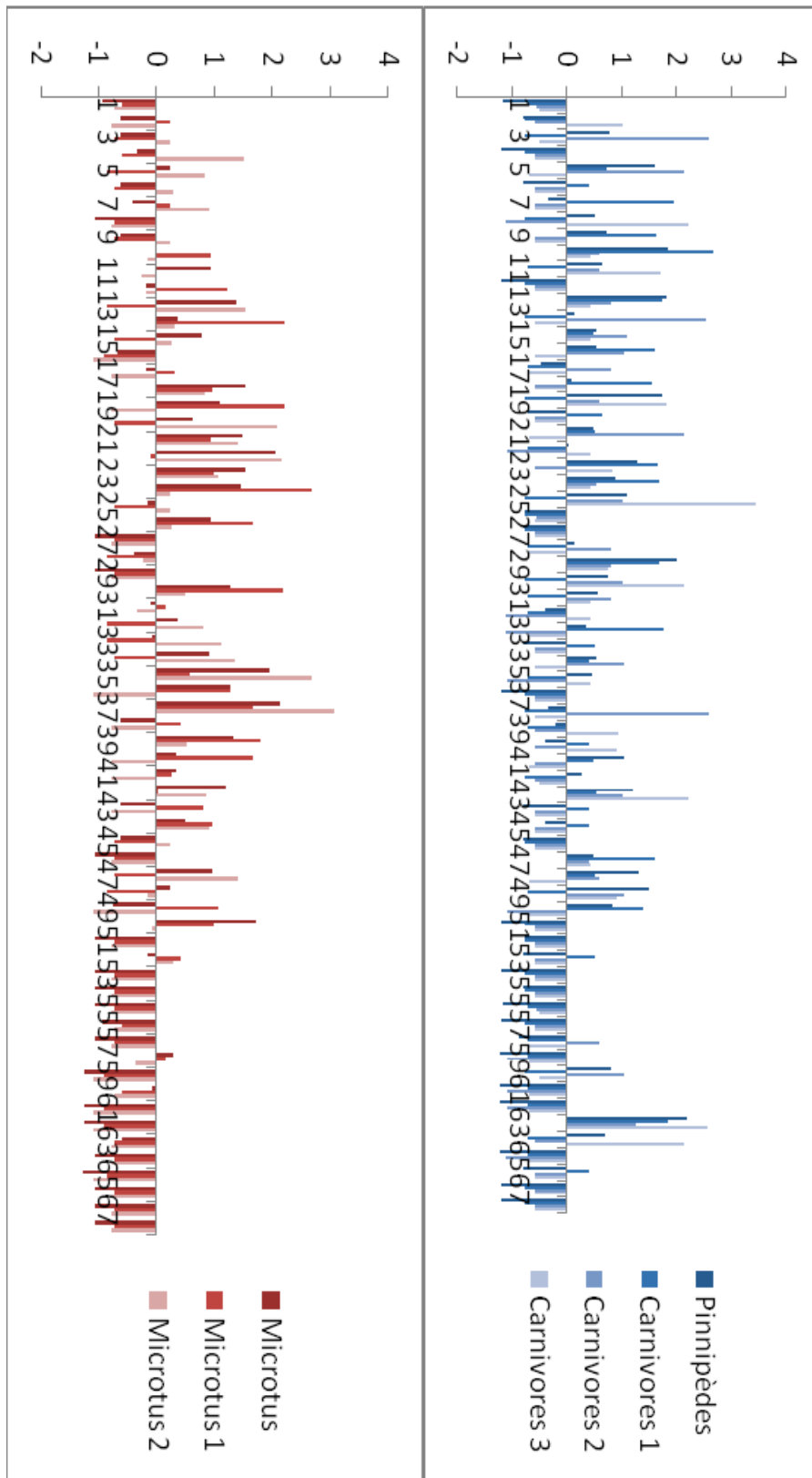


Figure 15 : Résultats R4S obtenus chez les sous-clades définis. Le taux de variation est représenté dans la couleur la plus foncée pour les clades complets (Figure E) et dans des couleurs plus claires pour les sous-groupes.

Conclusion et perspectives générales

Des 19 genres et 55 spécimens collectés dans notre échantillonnage taxonomique, il a été obtenu les génomes mitochondriaux complets de 18 individus de la sous-famille des Arvicolinae. La mise en place de méthodes visant à extraire, amplifier et séquencer un maximum de génome mitochondrial hypervariables en un minimum de temps s'avère efficace. La description de la radiation adaptative des Arvicolinae, déjà enrichie par la grande quantité de marqueurs moléculaires et la manière de les prendre en compte dans nos analyses, peut encore être améliorée par la quantité de taxons pris en compte.

L'obtention rapide des séquences complètes de mitogénomes hypervariables permet d'entreprendre des démarches identiques appliquées à des ordres, des familles, ou des sous-familles de Mammifères différentes, et ainsi de généraliser l'observation de la variabilité du génome mitochondrial à des échelles très diversifiées, d'autant qu'il a été constaté que les génomes mitochondriaux de groupes à évolution lente sont plus faciles à obtenir que ceux des groupes à évolution rapide.

En génomique comparative, il devient aujourd'hui possible, à partir de la description précise des relations phylogénétiques d'un groupe, d'identifier les sites du génome mitochondrial soumis à de fortes contraintes mutationnelles, et donc de vérifier si la théorie des hot spots mutationnels est la source de l'hétérogénéité des taux de substitution mitochondriaux des Mammifères. Il devient alors primordial d'établir des bases de données comparatives, afin de définir l'état de partage des sites en situation de hot spot (partagés ou non partagés). De telles études, encore inenvisageables sans l'existence d'une base de données en mitogénomes rapides, permettront de définir les points chauds mutationnels en fonction de leur quantité et de leur position sur le génome, de leur durée d'application sur le site concerné, ou encore de l'héritabilité de ces caractères hautement variable au sein d'un clade. Dans ce cadre, l'obtention des taux absolus de substitution doit être réalisée à partir des datations moléculaires des événements de spéciation, afin d'établir éventuellement un seuil de variabilité au-delà duquel un site est considéré en situation de hot spot.

BIBLIOGRAPHIE

- Abramson N.I., Lebedev V.S., Bannikova A.A., Tesakov A.S. , 2009. Radiation events in the subfamily Arvicolinae (Rodentia) : Evidence from nuclear genes. Doklady Biological Sciences 428, 713-717.
- Abramson N.I., Lebedev V.S., Tesakov A.S., Bannikova A.A., 2009. Supraspecies relationships in the subfamily Arvicolinae (Rodentia, Cricetidae) : an unexpected result of nuclear gene analysis. Molecular Biology 43, 897-909.
- Arnason U., Gullberg A., Janke A., Kullberg M., 2007. Mitogenomic analyses of caniform relationships. Molecular Phylogenetics and Evolution 45, 863-874.
- Bannikova A., Lebedev V., Lissovsky A., Matrosova V., Abramson N., Obolenskaya E., Tesakov A., 2010. Molecular phylogeny and evolution of the Asian lineage of vole genus *Microtus* (Rodentia: Arvicolinae) inferred from mitochondrial cytochrome b sequence. Biological Journal of the Linnean Society 99, 595-613.
- Barja G., Herrero A., 2000. Oxidative damage to mitochondrial DNA is inversely related to maximum life span in the heart and brain of mammals. The Federation of American Societies for Experimental Biology Journal 14, 312-318.
- Bromham L., Rambaut A., Harvey PH, 1996. Determinants of rate variation in mammalian DNA sequence evolution. Journal of Molecular Evolution 43, 610-621.
- Buzan E.V., Krystufek B., Hänfling B., Hutchinson W.F., 2008. Mitochondrial phylogeny of Arvicolinae using comprehensive taxonomic sampling yields new insights. Biological Journal of the Linnean Society 94, 825-835.
- Chaline J., Brunet-Lecomte P., Montuire S., Viriot L., Courant F. , 1999. Anatomy of the arvicoline radiation (Rodentia): palaeogeographical, palaeoecological history and evolutionary data. Ann. Zool. Fennici 36, 239–267.
- Chaline J., Brunet-Lecomte P., Montuire S., Viriot L., Courant F., 1999. Anatomy of the Arvicolinae radiation (Rodentia) : palaeogeography, palaeoecological history and evolutionary data. Annales Zoologici Fennici 36, 239-267.
- Conroy C.J., Cook J.A., 1999. MtDNA evidence for repeated pulses of speciation within Arvicolinae and Murid rodents. Journal of Mammalian Evolution 6, 221-245.
- Conroy C.J., Cook J.A., 2000. Phylogeography of a post-glacial colonizer : *Microtus longicaudus* (Rodentia : Muridae). Molecular Ecology 9, 22-33.
- Courant F., David B., Laurin B., Chaline J., , 1997. Quantification of cranial convergences in arvicolines. Biological Journal of the Linnean Society 62, 505-517.
- Criscuolo A., Berry V., Douzery E.J.P., Gascuel O., 2006. A fast distance-based approach for (super) tree building in phylogenomics. Systematic Biology 55, 740-755.
- Davis C.S., Delisle I., Stirling I., Siniff D.B., Strobeck C. , 2004. A phylogeny of the extant Phocidae inferred from complete mitochondrial DNA coding regions. Molecular Phylogenetics and Evolution 33, 363-377.
- Delisle I., Strobeck C., 2005. A phylogeny of the Caniformia (order Carnivora) based on 12 complete protein-coding mitochondrial genes. Molecular Phylogenetics and Evolution 37, 192-201.
- Eyre-Walker A., Smith N.H., Smith J., 1999. How clonal are human mitochondria. Proceedings of the Royal Society of London 266, 477-483.
- Fulton T.L., Strobeck C., 2010. Multiple fossil calibrations, nuclear loci and mitochondrial genomes provide new insight into biogeography and divergence timing for true seals (Phocidae, Pinnipedia). Journal of Biogeography 37, 814-829.
- Fulton T.L. , Strobeck C. , 2006. Molecular phylogeny of the Arctoidea (Carnivora): Effect of missing data on supertree and supermatrix analyses of multiple gene data sets. Molecular Phylogenetics and Evolution 41, 165-181.
- Galewski T., Tilak M.-K. , Sanchez S., Chevret P., E., Paradis E. and Douzery, 2006. The evolutionary radiation of Arvicolinae rodents (voles and

lemmings): relative contribution of nuclear and mitochondrial DNA phylogenies. BMC Evolutionary Biology 6, 80.

Galtier N., Enard D., Redondy Y., Bazin E., and Belkhir K., 2006. Mutation hot spots in mammalian mitochondrial DNA. Genome Research 16, 215-222.

Guindon S., Gascuel O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52, 696-704.

Harman D., 1957. Aging: a theory based on free radical and radiation chemistry. Journal of Gerontology 2, 298-300.

Hassanin A., Ropiquet A., Couloux A., Cruaud C., 2009. Evolution of the mitochondrial genome in Mammals living at high altitude : new insights from a study of the tribe Caprini (Bovidae, Antilopinae). Journal of Molecular Evolution 68.

Jaarola M., Martinkova N., Gündüz I., Brunhoffa C., Zimab J., Nadachowskie A., Amorif G., Bulatovag N.S., Chondropoulosh B., Fragedakis-Tsolish S., Gonzalez-Estebani J., Lopez-Fusterj M.J., Kandaurovk A.S., Kefelioglud H., da Luz Mathiasl M., Villatei I., Searlem J.B., 2004. Molecular phylogeny of the speciose vole genus *Microtus* (Arvicolinae, Rodentia) inferred from mitochondrial DNA sequences. Molecular Phylogenetics and Evolution 33, 647-663.

Katoh K., Asimenos G., Toh H., 2009. Multiple Alignment of DNA sequences with MAFFT. In: Press, Humana (Ed.), Bioinformatics for DNA sequence analysis, pp. 39-64.

Lartillot N., Lepage T., Blanquart S., 2009. Phylobayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25, 2286-2288.

Lecompte E., Aplin K., Denys C., Catzeflis F., Chades M., Chevret P., 2008. phylogeny and biogeography of African Murinae based on mitochondrial and nuclear gene sequences, with a new tribal classification of the subfamily. BMC Evolutionary Biology 8, 199.

Li WH, Tanimura M., Sharp P., 1987. An evaluation of the molecular hypothesis using mammalian DNA sequences. Journal of Molecular Evolution 25, 330-342.

Lopez JV., Yuhki N., Masuda R., Modi W., O'Brien SJ., 1994. *Numt*, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. Journal of Molecular Evolution 39, 174-190.

Martin AP, 1995. Metabolic rate and directional nucleotide substitution in animal mitochondrial DNA. Molecular Biology and Evolution 12, 1124-1131.

Martin AP, Naylor G., Palumbi S., 1992. Rates of mitochondrial DNA evolution in sharks are slow compared with Mammals. Nature 357, 153-155.

Martin AP, Palumbi S., 1993. Body size, metabolic-rate, generation time, and the molecular clock. Proceedings of the National Academy of Sciences, USA. 90, 4087-4091.

Mayrose I., Graur D., Ben-Tal N., Pupko T., 2004. Comparison of site-specific rate-inference methods for protein sequences : empirical bayesian methods are superior. Molecular Biology and Evolution 21, 1781-1791.

Mooers AO., Harvey P., 1994. Metabolic rate, generation time, and the rate of molecular evolution in birds. Molecular Phylogenetic and Evolution 3, 344-350.

Nabholz B., Glémin S. and Galtier N., 2007. Strong variations of mitochondrial mutation rate across Mammal - the Longevity Hypothesis. Molecular Biology and Evolution 25, 120-130.

Nabholz B., Mauffrey J.-F., Bazin E., Galtier N., and Glemin S., 2008. Determination of mitochondrial genetic diversity in Mammals. Genetics 178, 351-361.

Nunn GB., Stanley S., 1998. Body size effects and rates of cytochrome b evolution in tube-nosed seabirds. Molecular Biology and Evolution 15, 1360-1371.

Ohta T., 1993. An examination of the generation-time effect on molecular evolution. Proceedings of the National Academy of Sciences, USA 90, 10676-10680.

Philippe H., 1993. MUST, a computer package of Management Utilities for Sequences and Trees. Nucleic Acids Residuals 21, 5264-5272.

Posada D., Crandall K.A., 1998. Modeltest : testing the model of DNA substitution. *Bioinformatics* 14, 817-818.

Repenning C.A., 2001. Beringian climate during intercontinental dispersal: a mouse eye view. *Quaternary Science Reviews* 20, 25-40.

Repenning C.A., Fejfar O., Heinrich W.D., 1990. Arvicolid rodent biochronology of the Northern Hemisphere. In: Fefjar O., Heinrich W.-D. (Ed.), *International Symposium on Evolution, Phylogeny, and Biostratigraphy of Arvicolids (Rodentia, Mammalia)*, Pfeil-Verlag, Prague, pp. 385-418.

Robovsky J., Ricankova V., Zrzavy J., 2008. Phylogeny of Arvicolinae (Mammalia, Cricetidae): utility of morphological and molecular data sets in a recently radiating clade. *Zoologica scripta* 37, 571-590.

Steppan, S.J., Adkins, R.M., Anderson, J., 2004. Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. *Syst. Biol.* 53, 533–553.

Steppan S.J., Adkins R.M., Anderson J., 2004. Phylogeny and divergence-date estimates of rapid radiations in Muroid rodents based on multiple nuclear genes. *Systematic Biology* 53, 533-553.

Steppan S.J., Adkins R.M., Spinks P.Q., Hale C., 2005. Multigene phylogeny of the Old World mice, Murinae, reveals distinct geographic lineages and the declining utility of mitochondrial genes compared to nuclear genes. *Molecular Phylogenetics and Evolution* 37, 370-388.

Swofford D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. In: Associates, Sinauer (Ed.), Sunderland, Massachussets.

Tougaard C., Brunet-Lecompte P., Fabre M., Montuire S., 2008. Evolutionary history of two allopatric *Terricola* species (Arvicolinae, Rodentia) from molecular, morphological, and palaeontological data. *Biological Journal of the Linnean Society* 93, 309-323.

Tougaard C., Ranvoisé E., Petitjean A., Quéré J.P., 2008. New Insight into the Colonization Processes of Common Voles: Inferences from Molecular and Fossil Evidence. *PLoS ONE* 3, e3532.

Triant D.A. , DeWoody J.A., 2007. Extensive mitochondrial DNA transfer in a rapidly evolving rodent has been mediated by independent insertion events and by duplications. *Gene* 401, 61-70.

Triant D.A. , DeWoody J.A., 2008. Molecular analyses of mitochondrial pseudogenes within the nuclear genome of arvicoline rodents. *Genetica* 132, 21-33.

Triant D.A. , DeWoody J.A. , 2006. Accelerated molecular evolution in *Microtus* (Rodentia) as assessed via complete mitochondrial genome sequences. *Genetica* 128, 95-108.

Wilson, D.E., Reeder, D.M., 2005. *Mammal species of the world*. Johns Hopkins University Press, Baltimore.

Xu X., Janke A., and Arnason U., 1996. The Complete Mitochondrial DNA Sequence of the Greater Indian Rhinoceros, *Rhinoceros unicornis*, and the Phylogenetic Relationship Among Carnivora, *Perissodactyla*, and *Artiodactyla* (+ *Cetacea*). *Molecular Biology and Evolution* 13, 1167-1173.

ANNEXE 1

Systématique des Arvicolinae, Echantillonnage taxonomique et numéro d'accèsion

GENRES	ESPECES	N° ID TISSU	N° ID ADN	DATE EXT	INIT. PREL.
Alticola	albicaudus				
	argentatus	NT0609TA01	Aar1		
	barakshin				
	lemminus	UAM Mamm 23201	Ale1		UAM
	macrotis				
	montosa				
	olchonensis				
	roylei				
	semicanus				
	stolickanus				
	strelzowi				
Arborimus	albipes	UAM Mamm 42255	Aal1		UAM
	longicaudus	UAM Mamm 42294	Alo1		UAM
	pomo				
Arvicola	amphibius				
	sapidus	T - 2165	Asa	09/09/08	FC
	scherman				
	terrestris	LS0310_I339	Ate	11/03/08	JFC
Blanfordimys	afghanus				
	bucharensis	T - 1060	Mbu Mbu-2	09/10/08 11/12/08	FC
Caryomys	eva				
	inez				
Chionomys	gud				
	nivalis	4389	Cni-1	27/10/08	JPQ
			Cni-2	11/12/08	
		5055	Cni-3	05/04/09	JPQ
	roberti				
Dicrostonyx	groenlandicus	UAM Mamm 82411			UAM
	hudsonius				
	nelsoni				
	nunatakensis				
	richardsoni				
	torquatus	UAM Mamm 89075			UAM
	unalascensis				
vinogradovi					
Dinaromys	bogdanovi				
Ellobius	alaicus				
	fuscocapillus				
	lutescens				

	talpinus				
	tancrei	NT0609SA1	Eta-1	29/01/09	JPQ
Eolagurus	luteus				
Eothenomys	custos				
	melanogaster				
	miletus				
	olitor				
	proditor				
	wardi				
Hyperacrius	fertilis				
	wynnei				
Lagurus	lagurus	T-1536	Lla1	06/07/09	FC
Lasiopodomys	brandtii				
	fuscus				
	mandarinus	T - 1066	Mma, Mma-2	16/06/08	FC
Lemmiscus	curtatus	UAM Mamm 70560 MVZ 218374	Lcu1		UAM MVZ
Lemmus	amurensis	UAM Mamm 101992			UAM
	lemmus	UAM Mamm 43047	Llm1		UAM
	portenkoi				
	sibiricus	T-1336	Lsi1	06/07/09	FC
	trimucronatus	UAM Mamm 98349			UAM
Microtus	abbreviatus				
	agrestis	4821	Mag-1	20/10/08	JPQ
	cf agrestis	T-5335	Mxx	09/10/08	FC+CD
	anatolicus				
	arvalis	3872 LS0504-1755	Mar	20/06/08	JPQ JFC
	bavaricus				
	brachycercus				
	breweri				
	cabreræ				
	californicus	USNM569290 MVZ 223492			SI/SP MVZ
	canicaudus	UAM Mamm 67599	Mcn-1		UAM
	chrotorrhinus	T - 603	Mch	09/09/08	FC
	clarkei				
	daghestanicus				
	dogramacii				
	duodecimcostatus	3249 3250	Mdu-1, Mdu-2	20/10/08	JPQ JPQ
	evoronensis				
	felteni				
	fortis	UAM Mamm 73462	Mfo-1		UAM

gerbei	AR0211 0112	Mpy-3 Mpy-4	04/03/09 JPQ 25/03/09
	LA00110305		
gregalis	NT0609AN1	Mgr-1	27/10/08 JPQ
	NT0609AN2	Mgr-3	04/03/09 JPQ
	UAM Mamm 34178	Mgr-4	UAM
guatemalensis			
guentheri			
hyperboreus	UAM Mamm 80280	Mhy-1	UAM
ilaeus			
irani			
irene	HL0709AQ01	Mir-1	
kikuchii	dispo		FC
leucurus	XQ0202AS23	Lle-1	23/01/09 FC
	XQ0207AS01	Lle-2	
	TT0607SA01	Lle-3	25/03/09
	TT0607SA01	Lle-4	31/03/09
levis	dispo		FC
liechtensteini			
limnophilus	RT0406SU1	Mli-1	20/10/08 JPQ
longicaudus	T - 136	Mlo Mlo-2 Mlo-3	20/06/08 FC 11/12/08 31/03/09
	MVZ 222629		MVZ
lusitanicus	SP0211XA06	Mlu-1	23/01/09
majori			
maximowiczii			
mexicanus			
middendorffii	T - 3509	Mmi, Mmi-2	09/09/08 FC
miurus	UAM Mamm 85562	Mmr-1	UAM
mongolicus			
montanus	UAM Mamm 69036	Mmn-1	UAM
	MVZ 222741		MVZ
montebelli	T - 144	Mmo	20/06/08 FC
mujanensis			
multiplex	4264	Mmu-1	27/10/08 JPQ
oaxacensis			
ochrogaster	T - 130	Moc	09/10/08 FC
	UAM Mamm 36773	Moc-2	UAM
oeconomus	UAM Mamm 85850	Moe-1	UAM
oregoni	UAM Mamm 50891	Mor-1	UAM
paradoxus			
pennsylvanicus	T - 140	Mpe	09/10/08 FC
	USNM570201		SI/SP
pinetorum	USNM568644	Mpi1	SI/SP
qazvinensis			
quasiater			

	richardsoni	T - 598	Mri	09/09/08	FC
	rossiaemeridionalis	dispo			FC
	sachalinensis				
	savii				
	schidlovskii				
	socialis				
	subterraneus	3993 4079	Msu-1, Msu-2	20/10/08	JPQ JPQ
	tatricus				
	thomasi				
	townsendii	USNM568058	Mto1		SI/SP
	transcaspiscus				
	umbrosus				
	xanthognathus	UAM Mamm 87433	Mxa-1		UAM
Myodes	andersoni				
	californicus				
	centralis	NT0609AB2	Mce-1	23/01/09	JPQ
	gapperi				
	glareolus	Cg0507Y35	Mgl, Cgl Cgl-2	20/06/08 11/12/08	JFC FC
		T-907 5009	Cgl3	04/03/09	JPQ
	imaizumii				
	rex				
	rufocanus				FC
	rutilus				FC
	shanseius				
	smithii				
Myopus	schisticolor	UAM Mamm 80762	Msh-1		UAM
Neodon	forresti				
	irene	HL0709AQ01	Mir-1		
	juldaschi				
	sikimensis				
Neofiber	alleni	USNM 568085 UAM Mamm 51583	Nal1 Nal-4		SI_SP UAM
Ondatra	zibethicus	3498	Ozi-1	23/01/09	JPQ
Phaiomys	leucurus				
Phenacomys	intermedius	T-672 MVZ 207658	Pin1	06/07/09	FC MVZ
	ungava				
Proedromys	bedfordi				
Prometheomys	schaposchnikowi				FC
Synaptomys	borealis	UAM Mamm 101983			UAM
	cooperi				
Volemys	millicens				

RESUME

L'acquisition, en laboratoire, des génomes mitochondriaux complets de plusieurs espèces de rongeurs Arvicolinae permet la définition de l'histoire évolutive du groupe dans un cadre évolutif et temporel. Ces relations phylogénétiques sont inférées sur un grand nombre de caractères (16000 pb), grâce à des méthodes d'analyses statistiques novatrices.

La base de données moléculaires constituée représente un échantillon de génomes mitochondriaux à évolution dite "rapide", ouvrant des perspectives quant à la description des points chauds mutationnels dans des études de génomique comparative, et permettant d'identifier plus globalement la source de l'hétérogénéité des taux de substitution mitochondriale au sein des génomes mitochondriaux des Mammifères.

MOTS-CLEFS

Taux d'évolution, phylogénomique, mitochondrie, Arvicolinae, *Microtus*

SUMMARY

Experimental acquisition of complete mitochondrial genome for several Arvicolinae species allows defining evolutionary history of the group in evolutionary and temporal frames. These phylogenetic relationships are inferred on a great number of characters (16000 bp), thanks to innovator statistical analysis methods.

Constituted molecular data set represents a sample of "fast evolving" mitochondrial genomes, and allows describing mutational hot spots in comparative genomic studies, and identifying globally the origin of mitochondrial substitution rate heterogeneity detected in Mammalian mitochondrial genomes.

KEY WORDS

Evolutionary rates, Phylogenomic, Mitochondria, Arvicolinae, *Microtus*